

WorldScribe: Towards Context-Aware Live Visual Descriptions



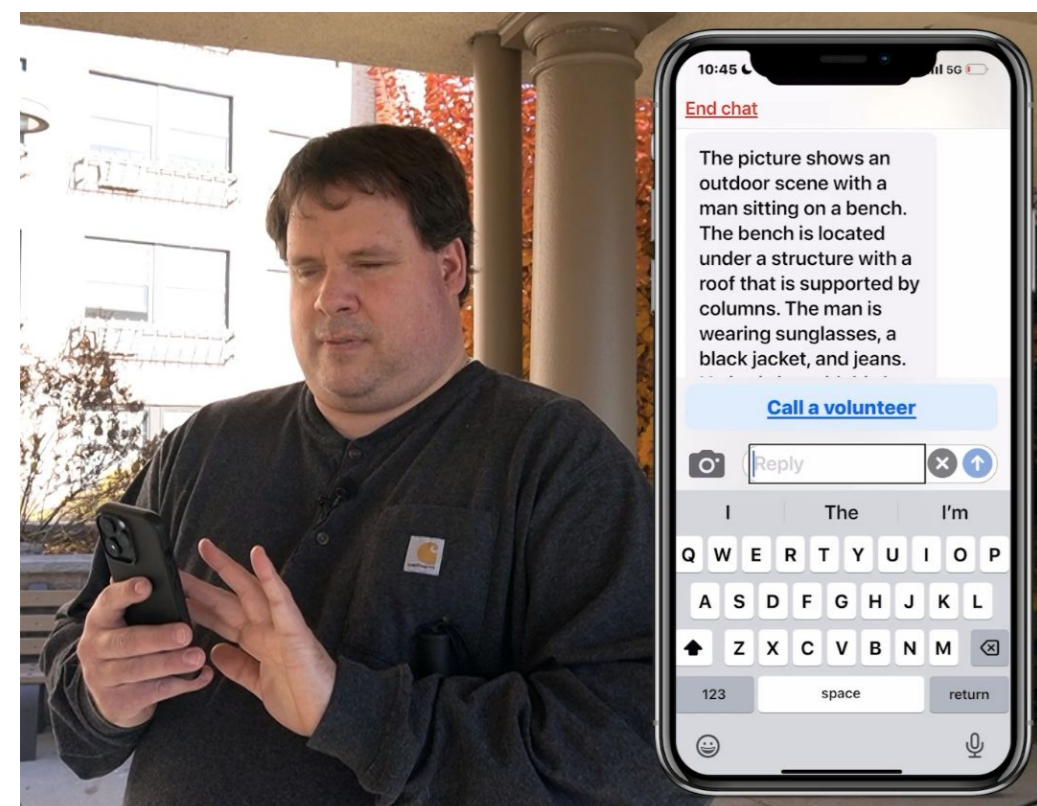
Wed, 16 Oct | 3:50 PM - 4:05 PM. Location: Allegheny 2.

Ruei-Che Chang (<https://rueiche.me/>) Yuxuan Liu Anhong Guo



Motivation

AI models



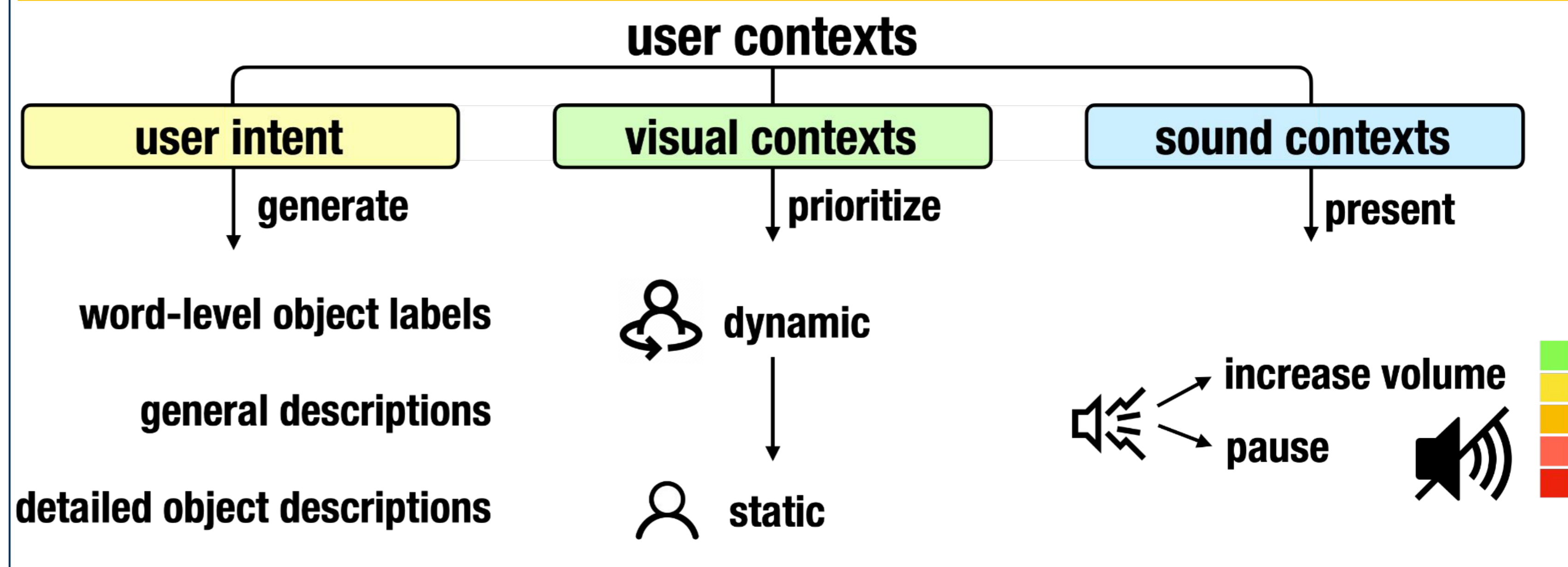
asynchronous
one-size-fits-all

Sighted assistance

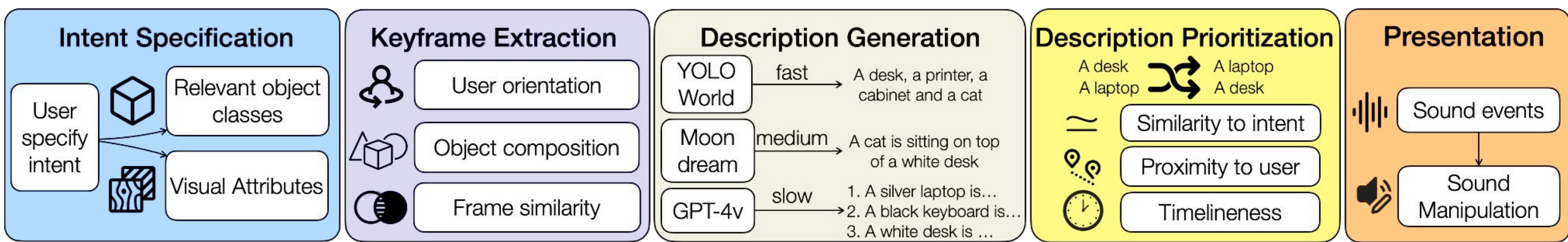


not always available
expensive

System Contribution



WorldScribe Components



Intent Specification

User-specified Intent	Object classes	Visual attributes
(a) Not Specified	'person', 'bicycle', 'car', 'motorcycle', 'airplane', ...	Color Texture Shape Spatial Normal Normal Normal Normal
(b) I want to explore anything. Describe with more color and texture information.	General GPT-4: 'person', 'bicycle', 'car', 'motorcycle', 'airplane', ...	Color Texture Shape Spatial Verbose Verbose Normal Normal
(c) Find a silver laptop on the desk, and other office objects might be around it.	Specific GPT-4: 'laptop', 'desk', 'mouse', 'monitor', 'keyboard', ...	Color Texture Shape Spatial Verbose Normal Normal Verbose

YOLO World → Object classes; Prompt for VLMs → Visual attributes

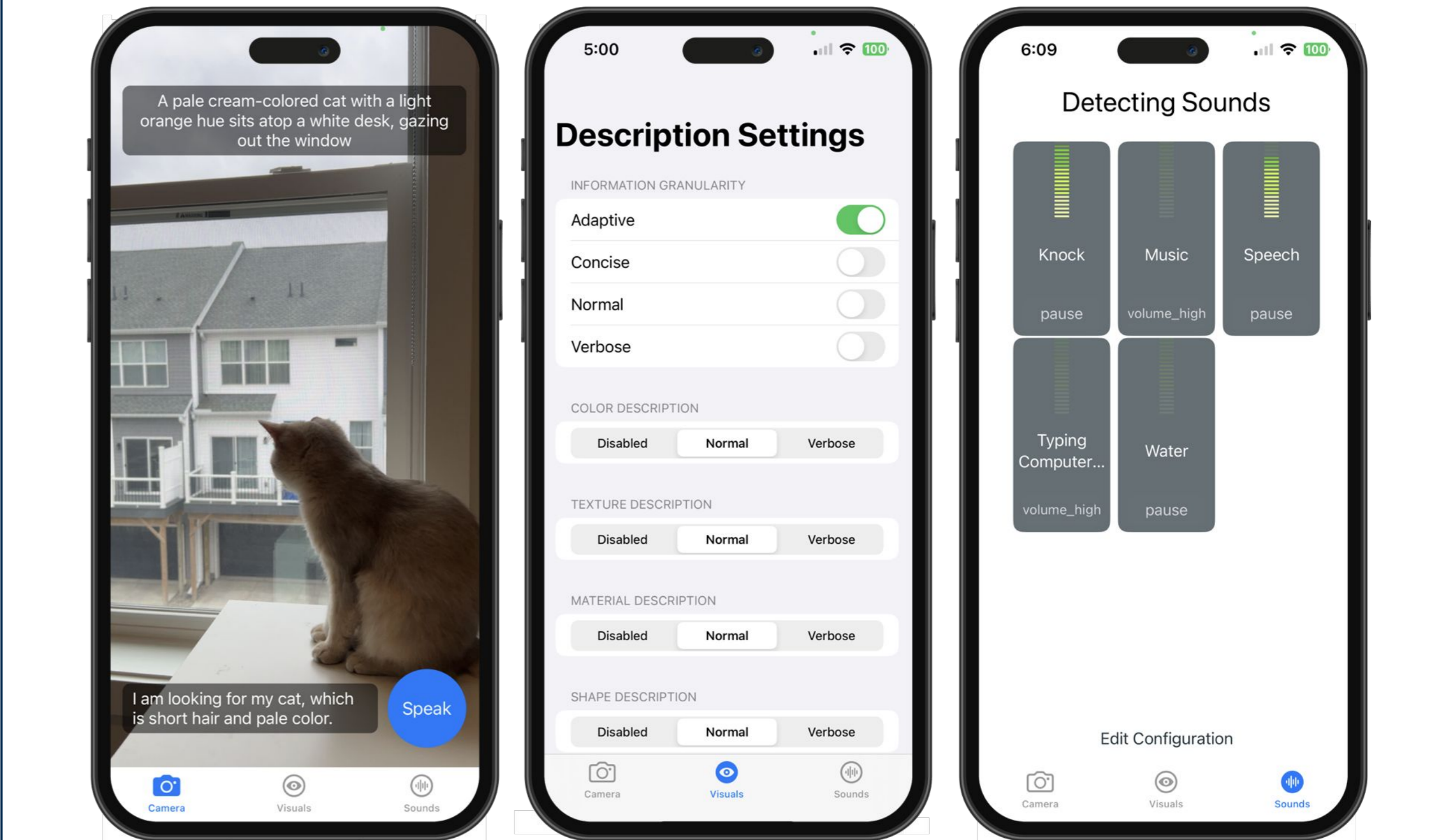
Description Generation

	(a) User turns quickly	(b) User remains static
YOLO World	A laptop, a desk, a monitor, and a lamp	A laptop, a desk, and a monitor
Moondream	A laptop is sitting on a desk next to a monitor and a silver desk lamp. (Outdated)	A cat sitting on top of a white desk, looking out the window.
GPT-4v	The desk lamp has a sleek silver finish with a curved arm and a cylindrical head... (Outdated)	A light cream-colored cat with a slender body and long tail is sitting and looking out of a window. Three closed drawers underneath the open one appear to have ...
WorldScribe (this work)	A laptop, a desk, a monitor, and a lamp	A cat sitting on top of a white desk, looking out the window.

Description Prioritization

Input Image	GPT4 Descriptions	CLIPSeg	Depth	Similarity scores		
				Describe paintings in detail.	Describe any person.	Describe building decoration and appliances.
	A person wearing a black beanie and a dark green coat sits facing a large abstract painting.		220	0.45	0.24	0.25
	The painting on the wall is a mix of various colors and textures with apparent urban elements.		33	0.66	0.18	0.35
	A wooden bench with metal supports gives a person a place to sit and the artwork.		200	0.33	0.22	0.29
	Two cylindrical trash receptacles, one silver and one black, are labeled for recycling and trash disposal.		39	0.04	-0.02	0.27

Customization Interface



User Evaluation and Future Work

- Participants perceived WorldScribe descriptions to be accurate based on the **contextual clues they ascertained**.
- Participants found WorldScribe descriptions **useful with adaptive and customized visual information**.
- Adding **long-term memory** and **learning user intent** on the fly (e.g., revisited place, visual changes).
- Evaluating **the utility of descriptions** based on different contexts (e.g., responsiveness, timeliness, detailedness).
- Directions with rapid evolution of future large models (e.g., dynamic prompting based on user context).



<https://worldscribe.org/>