INNOVATION PARTNERSHIPS
UNIVERSITY OF MICHIGAN

WorldScribe: Towards Context-Aware Live Visual Descriptions

TECHNOLOGY NUMBER: 2024-453

Technology ID
2024-453

Category
Software
Software & Content
Accessible Technologies/Blind
Accessibility

Inventor
Anhong Guo

Further information
Ashwathi Iyer
ashwathi@umich.edu

View online

OVERVIEW

WorldScribe is a mobile system that generates live, context-aware visual descriptions to help blind people understand their surroundings with independence, adapting to user intent and environment in real time.

- **Core Features:** Offers personalized, just-in-time visual narration by dynamically using AI models for object, scene, and sound recognition.
- **Market Opportunity:** Addresses the growing demand for scalable, affordable, and customizable assistive technology for visually impaired individuals, where "on-demand, context-sensitive accessibility" remains unsolved.

BACKGROUND

Millions globally experience visual impairment, creating a vast market for technologies that enable independent mobility and environmental awareness. Traditional accessibility tools, like static captions for images or remote human assistance via video calls, fall short in real-world scenarios: they lack timely, rich detail and can be costly, unavailable, or privacy-intrusive.

Recent advances in artificial intelligence—particularly vision-language models—have enabled automated descriptions for digital media, but these are largely asynchronous and fail to keep pace with dynamic, everyday scenes. There's a clear gap: users need fast, accurate, and adaptable descriptions in changing environments, for everyday tasks from navigation to social interaction. Demographic trends (aging populations), legal mandates, and growing smartphone

adoption all point toward rising demand for smarter, standalone accessibility solutions.

INNOVATION

WorldScribe solves this problem with an integrated approach:

Users specify their goals (e.g., “find my silver laptop”). The system then identifies relevant objects and visual attributes using speech and AI prompts. As the user moves or the scene changes, WorldScribe extracts key frames based on camera orientation and live video analysis, then generates a succession of descriptions with varying depth and speed—fast, simple summaries for dynamic moments, detailed narratives for stationary scenes.

Descriptions are prioritized by semantic relevance to the user’s intent and physical proximity, ensuring what matters most gets described first. The system adapts audio delivery automatically, pausing or changing volume in response to environmental sound, for better usability.

Unlike previous solutions, WorldScribe’s pipeline balances accuracy, customization, and real-time feedback. It uses multiple AI models simultaneously to offer adaptive granularity, and is one of the first platforms to fully integrate sound awareness and user-driven customization into live environmental description—making it not just more informative, but practical for everyday autonomy, at scale.

ADDITIONAL INFORMATION

REFERNCES:

"WorldScribe: Towards Context-Aware Live Visual Descriptions"