Proceedings of Virtual Reality International Conference (VRIC 2011), 6-8 April 2011, Laval, France.
RICHIR Simon, SHIRAI Akihiko Editors. International conference organized by Laval Virtual.

Invoked computing

Spatial audio and video AR invoked through miming

ZERROUG, Alexis, CASSINELLI, Alvaro, ISHIKAWA, Masatoshi

Ishikawa Komuro Laboratory, the University of Tokyo
azerroug@gmail.com

Figure 1. "Invoked Computing" at home; mail on a pizza box (right) and virtual birds on the wall (left).

Abstract— **Direct interaction with everyday objects augmented with artificial affordances is clearly a very efficient approach leveraging natural human interaction capabilities. Hence the idea of conceiving ubiquitous computing as an invisible world which can be "condensed" on real objects. Rich Gold's [1] described ubiquitous computing field as an "enchanted village" in which people discover hidden affordances in everyday objects. With this project, we explore the reverse scenario: a ubiquitous intelligence capable of discovering affordances *suggested or represented symbolically by human beings* (as actions and scenarios involving objects and drawings). An example: taking a banana and bringing it closer to the ear. The gesture is clear enough: directional microphones and parametric speakers hidden in the room would make the banana function as a real handset on the spot.**

Keywords - Augmented Reality (AR); Ubiquitous Computing; New Human Interfaces and Displays; affordances [2] detection;

I. Introduction

The aim of the "invoked computing" project is to develop a multi-modal system able to turn everyday objects into computer interfaces or communication devices on the spot. To "invoke" an application, the user has just to mimic a specific scenario or represent a tool (graphically for instance). The system will try to recognize the suggested affordance, and actually instantiate the represented object and its function, through common AR techniques. For instance, to invoke a laptop computer, the user could take a pizza box, open it and "tape" on its surface; to invoke a sound system, the user could perhaps draw a vinyl, and do as if he were putting earphones.

We are interested here on developing a multi-modal AR system able to augment objects with video as well as sound using this interaction paradigm. Rather than a Human-to-Machine Interface, this could be called a Machine-to-Human Interface.

II. Related Works

A calibrated camera/projector pair is a common set-up for spatial augmented reality. On a previous work, one of the authors used such configuration to develop a "Volume Slicing Display" [3], a device capable of projecting a slice of a 3d volume on a passive screen. The area was somehow limited; works such as [4] introduce a way to increase the interaction area by using a steerable base for the camera/projector pair. This is the approach taken in the present project. In addition, we propose to improve the possibilities of interaction by using a parametric speaker giving the impression that sound comes from the augmented object itself.

The discovery of affordances by an artificial intelligence has been studied in the field of robotic in order to give to machine the ability to recognize objects as well as their specific uses. Interestingly, in [5], the authors describe a specific learning process relying on imitation.

III. Technical Details

As explained above, a particularity of the proposed AR system is the inclusion of *parametric speakers*, such as the AudioSpotlight; this technology is capable of creating audible sound from an array of ultrasound sources. Due to the high directivity and low attenuation of ultrasound, the sound can be heard from very far inside a specific area. Outside of this area, the sound is almost inexistent. For short distance applications, the sound bounces on obstacles and create an impressive effect of spatialized sound. By analogy to light, usual speakers would be incandescent bulbs when parametric speakers are flashlights. The combination of a projector and a parametric speaker on a steerable base create a really strong association of the sound with the image and increase the effect of spatialization even with a single source.

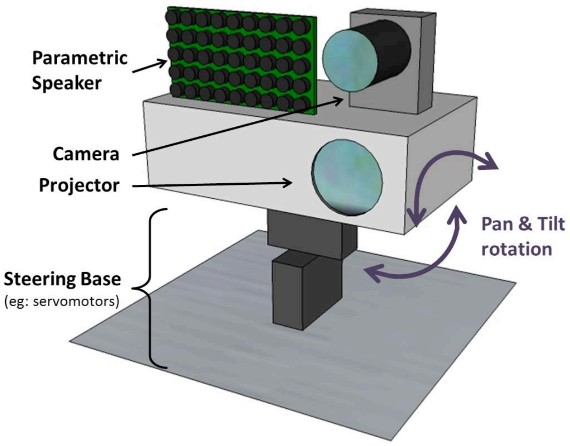

Figure 2. System detail

The most challenging part of this proposal is the automatic detection of suggested affordances. Object discovery and classification using perceived affordances have been studied in previous research [6]. In our proof of principle experiment, we have implemented a system with a very reduced capacity: the system is capable of detecting when something resembling a flat box is opened or closed, - and then infer that it may be a laptop computer. We are now exploring several tracking techniques both for detecting interesting objects in the scene as well as augmenting them with light and sound:

- High-speed vision systems [7] give a good stability for the tracking of moving objects.
- Marker based tracking such as ARToolkit (much appreciated for its stability and the precision of the 3d pose estimation required for the projection of the virtual display onto a surface.)
- In the case of marker-less interaction, we are planning to use a time-of-flight camera (e.g. SwissRanger or Kinect) to detect flat surfaces for projection or other pattern for tracking.
- Edge detection on the background might also be used to detect possible projection surfaces on the wall.

IV. Demonstration

We would like to create a living room setup with a sofa, a table and everyday objects that could be used as interfaces. If possible, we prefer to have white walls in order to use them as projection surfaces. Our system has to be fixed to the ceiling, so we need a support on the top of the booth. We also need a quiet sound environment to be able to hear the sound from the parametric speakers

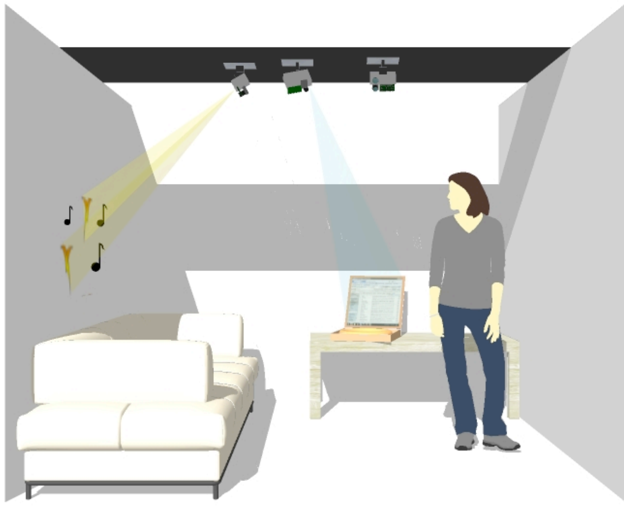

Figure 3. Desired set-up for Laval Virtual Revolution 2011

References

[1] R. Gold, “This is not a pipe.” *Commun. ACM* 36, July 1993.

[2] J.J. Gibson, “The theory of affordance. In: Perceiving, Acting, and Knowing”, Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

[3] A. Cassinelli and M. Ishikawa, “Volume Slicing Display”, SIGGRAPH ASIA 2009, Emerging Technologies, Yokohama (2009). Emerging Technologies Catalog, p.88.

[4] C. Pinhanez, “Using a steerable projector and a camera to transform surfaces into interactive displays”, In CHI '01 extended abstracts on Human factors in computing systems (CHI '01). ACM, New York, NY, USA, 2001.

[5] L. Montesano, M. Lopes, A. Bernardino, J. Santos-Victor, "Learning Object Affordances: From Sensory--Motor Coordination to Imitation," Robotics, IEEE Transactions on , vol.24, no.1, pp.15-26, Feb. 2008.

[6] M. Stark, P. Lies, M. Zillich, J. Wyatt and B. Schiele, “Functional Object Class Detection Based on Learned Affordance Cues”, Lecture Notes in Computer Science, Volume 5008/2008, 2008.

[7] T. Komuro, T. Tabata, M. Ishikawa, “A Reconfigurable Embedded System for 1000 f/s Real-Time Vision”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 20, No. 4, pp. 496-504, 2010.