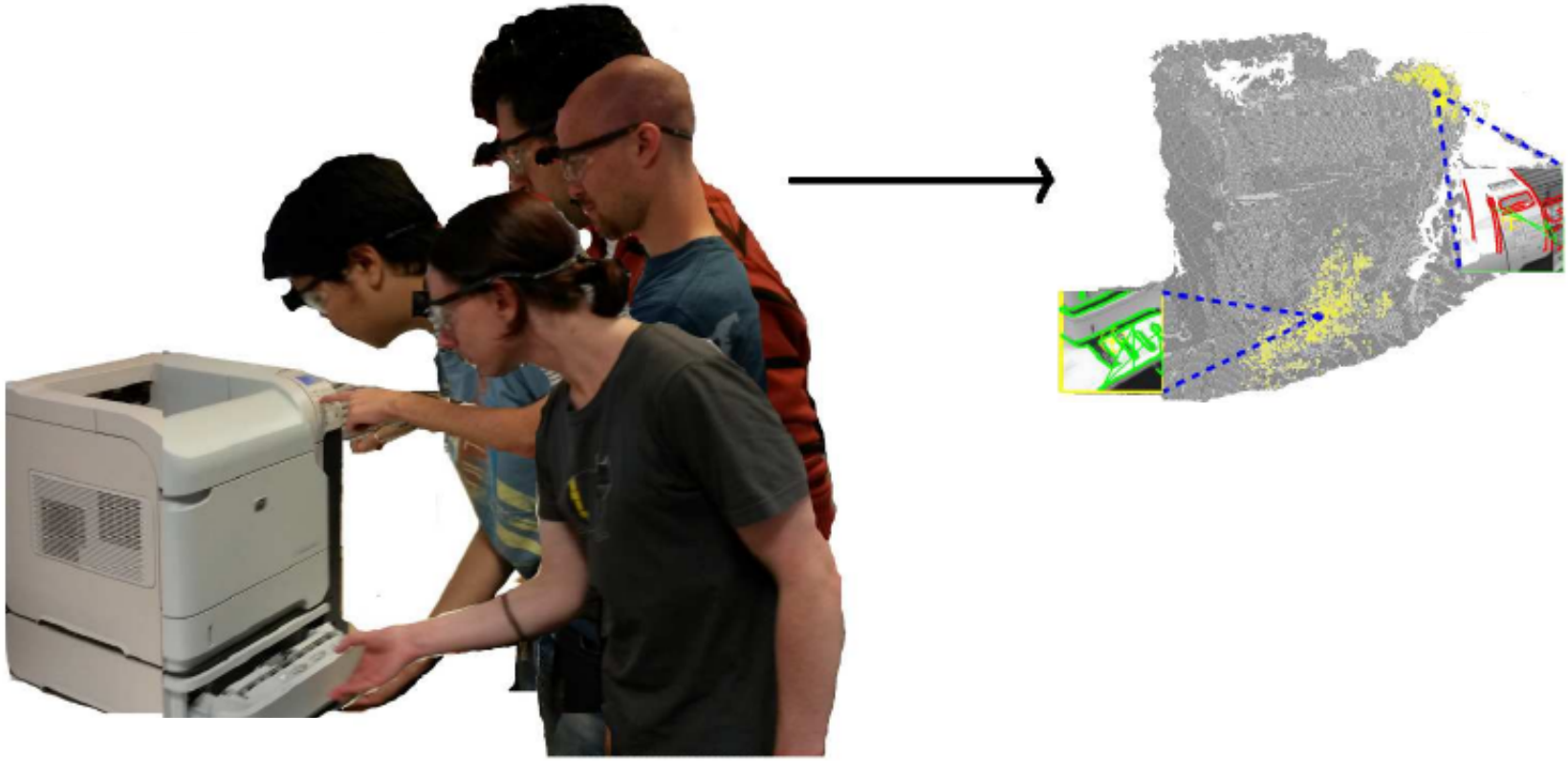
You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video

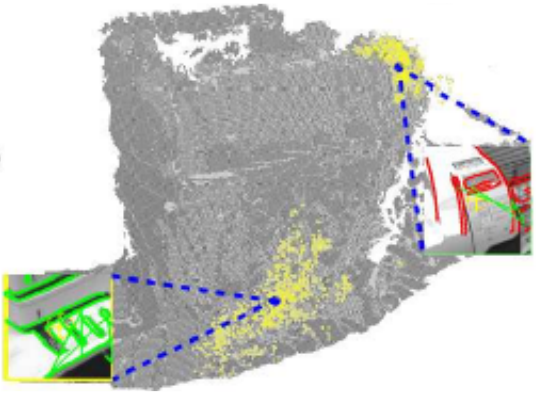
Dima Damen, Teesid Leelasawassuk, Osian Haines
Andrew Calway and Walterio Mayol-Cuevas

University of Bristol, United Kingdom

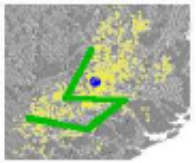
University of BRISTOL

Nov 3rd 2013

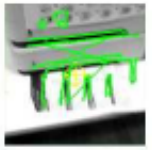




Three sources of information harvested

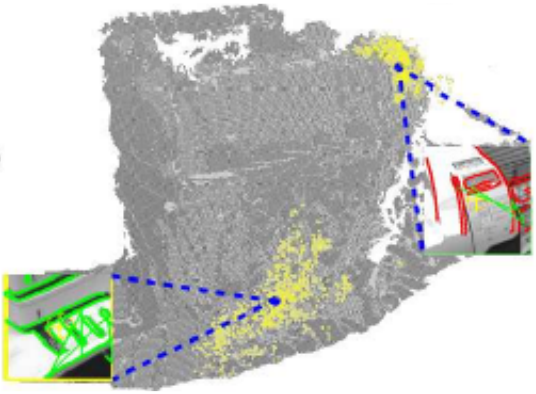


3D Gaze paths

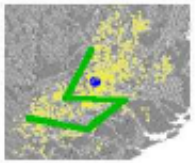


Appearance

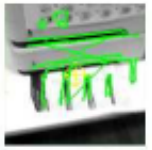
Action



Three sources of information harvested



3D Gaze paths



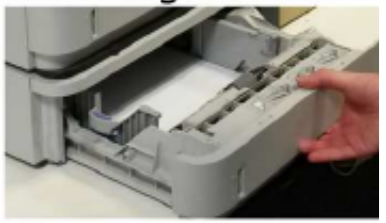
Appearance

Action

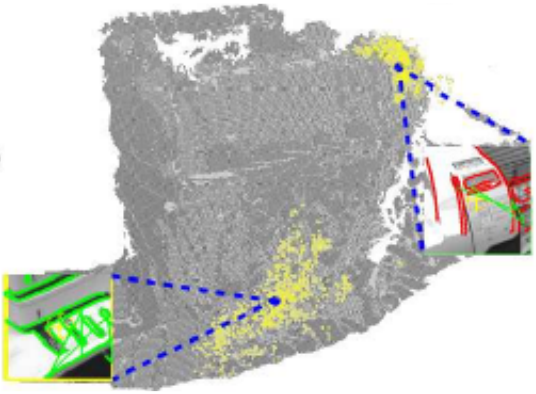
Automatically extracted object Usage Modes

"Mode 1"

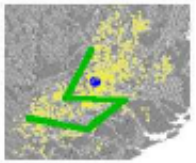
"Mode 2"



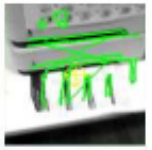
"Mode 3"



Three sources of information harvested



3D Gaze paths



Appearance

Action

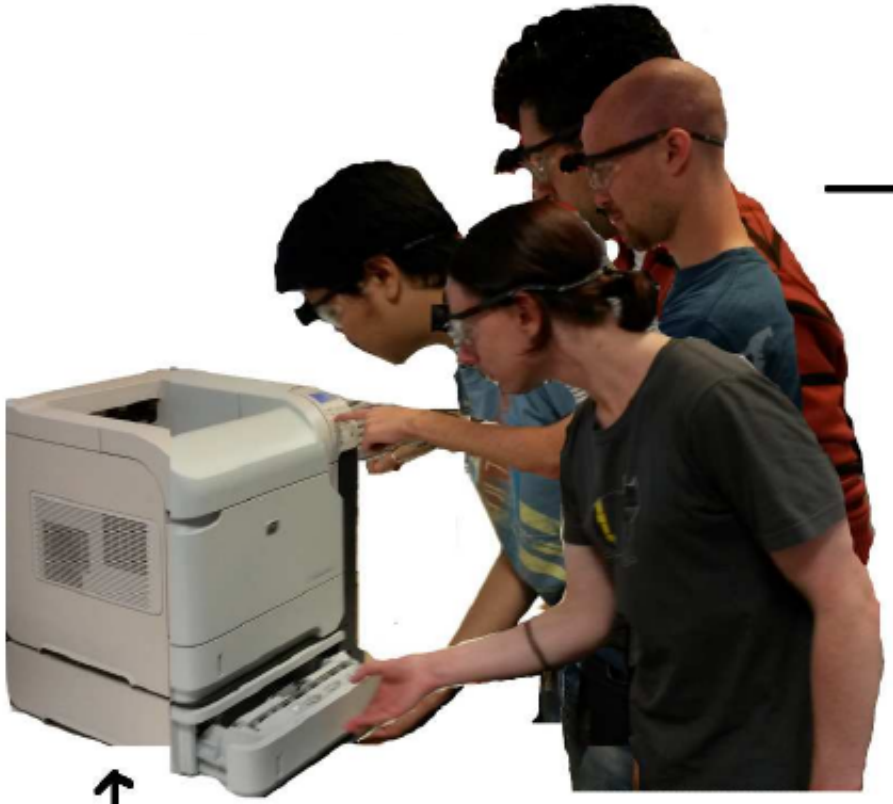
Automatically extracted object Usage Modes

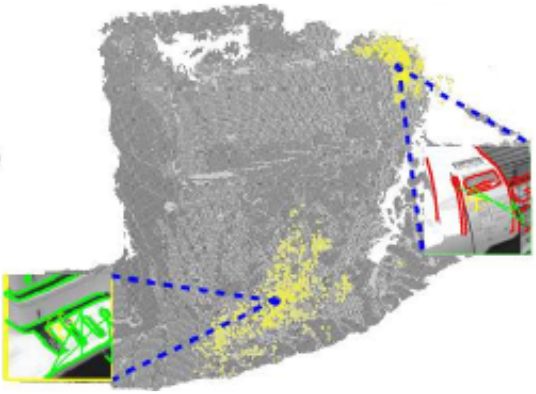
"Mode 1"

"Mode 2"

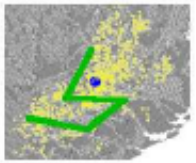
"Mode 3"

Automatically generated
How-To video guides

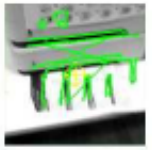




Three sources of information harvested



3D Gaze paths



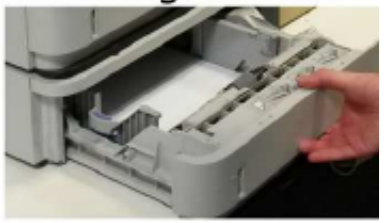
Appearance

Action

Automatically extracted object Usage Modes

"Mode 1"

"Mode 2"



"Mode 3"

Triggered by looking

Automatically generated
How-To video guides

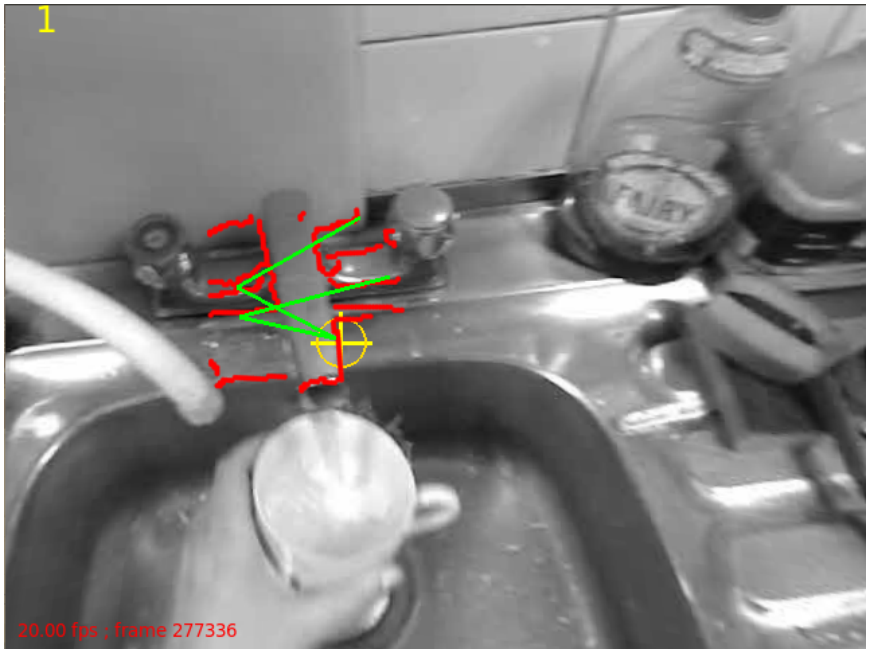
University of BRISTOL

The Proposed Method

- ***fully unsupervised***
- 3D Gaze Fixations
- Online clustering to discover objects
- Fixed objects like a coffee machine
- Moveable objects like a cup
- For each object, from multiple operators, video snippets are extracted.
- Used as help guides – triggered by gaze

Real-time Learning and Detection

- Multiple Texture-minimal Objects [1]



[1] Damen, Dima and Bunnun, Pished and Calway, Andrew and Mayol-Cuevas, Walterio (2012). Real-time Learning and Detection of 3D Texture-less Objects: A Scalable Approach. British Machine Vision Conference (BMVC) [Best Poster Paper]

Results