Assembling LEGO set with augmented reality instructions

Project report
Object Recognition and Computer vision - MVA

Othman Sbai, Pierre-Alain Langlois

December 20, 2016

1 Introduction

The advent of augmented reality devices such as Microsoft Hololens, Sony SmartEyeglass or Google Glass and others have made possible many interesting applications that augment the visual experience of the user with 3D holograms that are blended on his reality. Applications range from interior decoration and design, gaming but also increasing productivity in businesses by enhancing the real world and giving birth to broader imagination.

Among the applications of this new promising technology still in development is providing instructions for people to help them accomplish tasks either with human supervision or with annotating reality. In fact, as presented in [1], we can extract from the tutorial videos available online instructions for performing many tasks such as changing car tires, assemble furniture and also performing Cardiopulmonary rescucitation. These instructions can be efficiently provided to user with a 3D augmented reality device in the form of holograms and world annotations that are much comprehensible than paper instructions. An example of this world annotation is illustrated in the following figure 1.

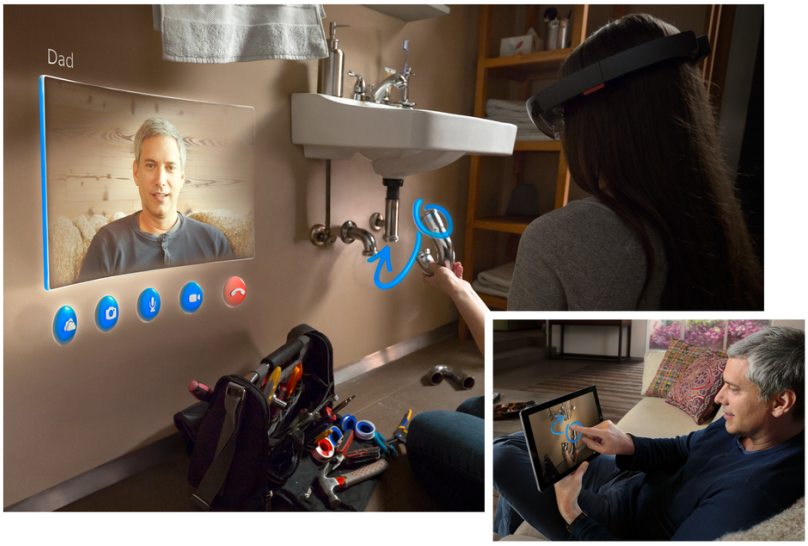

Figure 1: Annotating the world with Microsoft Hololens

We have chosen the topic : *A.3 Instructions for assembling simple lego objects.*

The advent of augmented reality tools such as Microsoft Hololens have made possible many interesting applications that augment the visual experience of the user providing relevant informations and distractions. Across the web, one can find lots of tutorial videos for performing a certain task be it assemble a furniture, prepare a meal, change tire in addition to DIY videos. These explanations can really be enhanced through augmented reality technology by overlaying instructions in the view of Hololens for example in order to adapt the tutorial to the real world's configuration.

We propose to tackle the problem of providing instructions for assembling simple LEGO set. This simple game-related problem is a good start in manipulating and recognizing 3D objects from a head-mounted camera. Our goal is to create an assistant that is able to recognize the state of the LEGO mounting problem and suggest the next step by blending virtual movement on the reality perceived by the player thanks to Hololens.

We will first implement a system that allows to recognize and locate the pieces in the frame in real time [3]. This system will provide simple informations to help the user assembling the pieces.

If the system works well enough, we will try to add the pose estimation in the process in order to give more precise informations about the pieces orientation [4].

Some researchers already worked on the problem of LEGO brick identification and retrieval in realtime from 2D images[2]. In our case, we will generate data directly from the hololens (both RGB and geometrical data), since we have only a few lego pieces to recognize. If needed, we will also use data augmentation techniques.

2 Plan of work

- GOAL: Implement an interactive assistant that helps solving/assembling a LEGO set. Providing relevant statistics to evaluate the performance of the method.
- We assume the knowledge of the set of sequences required for the assembly, this can be either supplied by LEGO from the manual, or can be deduced from tutorial videos, as was done in the paper [1].
- From a head mounted camera, recognize LEGO part to be moved and its destination and display a hint of the movement overlaid on the hololens.
- If the previous work is good enough, we will try to add support for pose estimation and more advanced instructions.

3 Operational organization

3.1 Group members

- Othman Sbai (MVA & École des Ponts ParisTech)
- Pierre-Alain Langlois (MVA & École des Ponts ParisTech)

3.2 Plans for work sharing

Pierre-Alain will be focused on the work regarding the object detection (including segmentation) and the pose estimation task. Othman will be focused on the tracking constraints, and the implementation and experiments on the hololens device. Both of us will also perform testing on each other implementations in order to make the produced code more robust.

References

[1] Jean-Baptiste Alayrac, Piotr Bojanowski, Nishant Agrawal, Josef Sivic, Ivan Laptev, and Simon Lacoste-Julien. Unsupervised Learning from Narrated Instruction Videos. *arXiv:1506.09215 [cs]*, June 2015. arXiv: 1506.09215.

[2] Leendert Botha, Lynette Van Zijl, and Mcelory Hoffmann. *Realtime LEGO Brick Image Retrieval with Cellular Automata*. July 2009.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. *arXiv:1506.02640 [cs]*, June 2015. arXiv: 1506.02640.

[4] Paul Wohlhart and Vincent Lepetit. Learning Descriptors for Object Recognition and 3d Pose Estimation. *arXiv:1502.05908 [cs]*, February 2015. arXiv: 1502.05908.