WROCŁAW UNIVERSITY OF SCIENCE AND TECHNOLOGY
FACULTY OF ELECTRONICS

FIELD: Control Engineering and Robotics (AIR)
SPECIALIZATION: Embedded Robotics (AER)

MASTER OF SCIENCE THESIS

Augmented reality goggles in robotic applications

Zastosowanie okularów rozszerzonej rzeczywistości
w aplikacjach robotycznych

AUTHOR:
Dawid Śliwa

SUPERVISOR:
dr inż. Janusz Jakubiak, I-6

GRADE:

WROCŁAW 2019

Contents

Chapter 1

Introduction

Robots become more an more often seen in our environment. Starting from nowadays standard industrial applications and ending on home appliances robots. They all have more or less user friendly interface created to programm or control them. In factories can be seen most often stationary or handheld controllers and in consumer appliances, smartphone almost every time is used. Problem is that, this kind of interaction is not natural for humans. For comparison, communication between two employes working together is mostly done by voice, gestures and sometime touch. That is why modern controllers should been using these. This could improve a way of interaction on human-machine level.

A few years ago, a revolution called Industry 4.0 began which most important statement was to not replace peoples in factories by machines, but allow them to cooperate at production line. From that time companies are trying to simplyfy teaching process of robots and give them ability to sense the changing environment. Also enhancements are done on the other side. Employees are equipped with many solutions with are extending they perception. This is allowing to get better undersatnding what machines are doing or even see what they are "thinking".

1.1 Purpose and scope of work

This thesis will focus on Augmented Reality and they usage in modern factories and research facilities. At the beginning, different types of AR technologies will be compared to give overall view on how this is working. Then industrial or commercial products which are available right now on the market will be presented. The last part of topic studies will try to present selected solutions with are already used in real world applications.

Research part of this thesis will try to present simple examples of implementation AR in robotic applications. The topic will cover the issue of planing movement of robotic arm and also controlling and presenting data from mobile robot. This should give more or less understanding what this technology is capable and whats are its current limitations.

Chapter 2

Introduction to Augmented Reality

The perception of our surroundings is made to a large extent by the organ of sight. Thanks to that we are able to navigate and operate in our real environment. But what if we will try to trick him by placing displays in front of our eyes? Depending on content generated by computer it could simply show some additional information or create ilusion of being completely somewhere else. To distinguish types of immersion the concept of a "Reality-Virtuality Continuum" was created. It graphical representation is shown on figure 2.1.
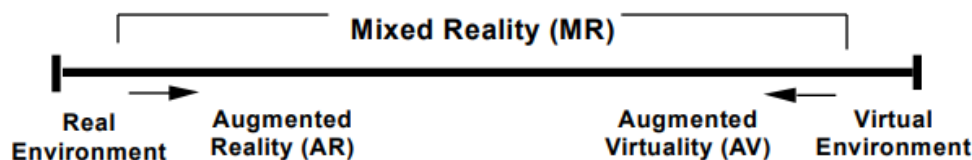

Figure 2.1: Reality-Virtuality (RV) Continuum

On the most left side of this line there is our real environment with real objects in no way disturbed by computer graphics. On the other hand on the most right side there is Virtual Reality with fully 3D generated world that could even in example not holding known by us laws of physics. Between two of those is everything with is mixing one part with another. Depending on what balance is, then we will talk about Augmented Reality or Augmented Virtuality. For example when there operator is getting some simulated cues to augment his natural feedback then it is AR. When in virtual world appears some real life objects or persons then it is called AV. Good example for that are modern news where reporters are working at green screen and in television they appear in 3D generated studios. As this paper will only discuss topic of using Augmented Reality in robotics applications that is why technical aspects of Augmented Virtuality or Virtual Reality will be not considered.

2.1 Technology overview

AR devices can take many different forms but way of processing data is almost always the same. In the simplest way it could be explained as following process. First device need to capture an image and localize itself or the user in environment then mix those data with CG objects and at the end display them on a screen. Figure 2.2 presents this pipeline in graphical form with few additional steps.

Figure 2.2: Simplified AR pipeline

This section will show different types of technologies with where used at implementation of individual elements of this process.

2.1.1   Types of image projection

There is several way of displaying virtual objects in our environment. Depending on its type they could be more or less immersive for the user.

The most popular and the simplest types of projection is video-mixing. It involves the usage of devices equipped with a camera and standard display to present AR content. In most cases smartphones or tablets are used because they have built-in every needed component and they are very portable. Figure 2.3 present example of such projection. The great advantage of this approach compared to others is that you can use one device in cooperation with other people, which could result in a significant reduction of operating costs. The disadvantages should be mentioned that the employee is not able to observe the environment and use both hands to do his job. Therefore, this type of projection is most often used in devices used to supervise the operation of machines.

Figure 2.3: Worker using tablet as video-mixing device

Second type of projection is slightly different than first one. It is called spatial display. In this case virtual objects are shown directly on real environment surfaces by usage of digital projectors. It could be realised on two way. We could have handheld device with will work similar to flashlight or stationary mounted projector. In first case we could

for example inspect virtual paths of transportation robots in warehouses or factory by highlighting floors. Also autonomous cars can use their headlights to display information on the road and for instance inform pedestrian that they could safely cross the passage. Stationary projectors are most often used with collaborative robots. They can display for example some cues for worker which item it will pick-up or where it will place it on shared workspace. Presented types of projection could be seen on Figure 2.4.

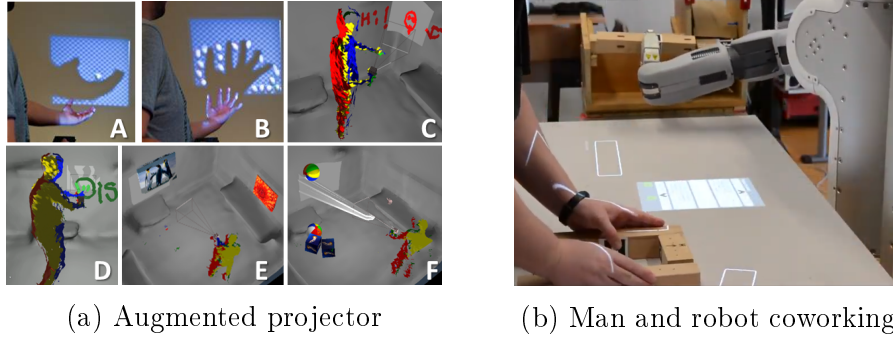(a) Augmented projector

(b) Man and robot coworking

Figure 2.4: Examples of spatial projection

Next AR display technology which will be presented is optical see-through. Its principle of operation boils down to projecting images on partially reflecting surface. This allow to combine real world view with generated graphics which will appear as floating in space holograms. There are two ways to achieve this goal. Device could use a head-mounted or head-up displays. Figure 2.5 shows how the projection is carried out with particular case. First type presented at image 2.5a is commonly used in situations where worker need to have both hands free to do his work. At HMD could be shown some step by step instructions or in case operating with robots, live sensors data. Example from image 2.5b show the most common scenario of usage of HUD. Nowdays they are showing informations about speed and navigation guidance for a driver but in autonomous vehicles they could display also some cues what car see and what it intends to do.
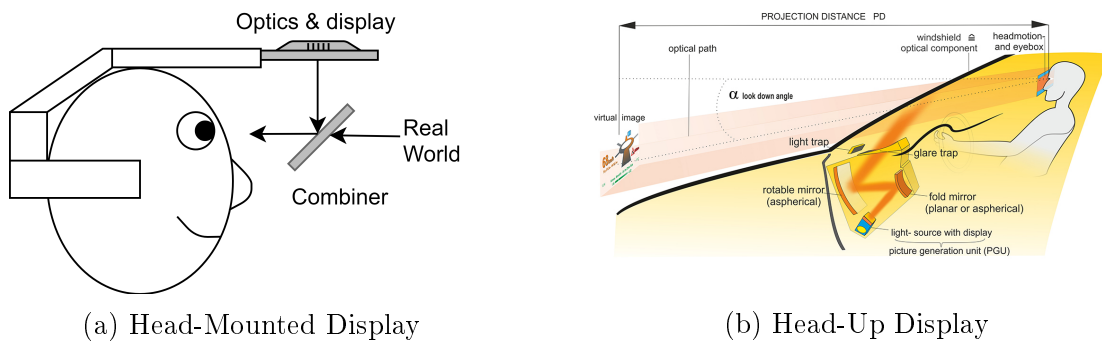

(a) Head-Mounted Display

(b) Head-Up Display

Figure 2.5: Examples of optical see-through projections

2.1.2 Positioning and location

Only small part of AR devices is presenting data that is not depending on real environment obstacles. These include basic versions of HUD and smartglasses. Rest of them need somehow localize itself or detect position of the user. In this section various types of positioning techniques will be presented.

Marker-based localization is simplest one. The principle of operation consists in detection of predefined shape or image by the device camera. To achieve this first of all data need to be preprocesed.

Most trivial example is when our marker have square shape and high contrast. In that case algorithm need only know what is length of side. With this information it is able to calculate position and orientation of the device basing on detected corrners. As the most important are only the edges the center can have any shape. This allow to store some information inside and also help to determinate from with side of square camera is looking. Figure 2.6a show detected marker with read associated to it ID number. Pros of this approach is quite low requirement for computational power and easy way to generate many unique ID's. The disadvantage is that the user must print these images and arrange them in their chosen locations.

There is also posibility to track normal images or even 3D objects. In this case there are used algorithms with are extracting some particular features from provided data (Figure 2.6b). Basing on them device could be able to localize itself in space only by looking on for example machine logo. Pros of this approach is better tracking of object even when whole image is not in camera view. Disadvantage is more complex computation.
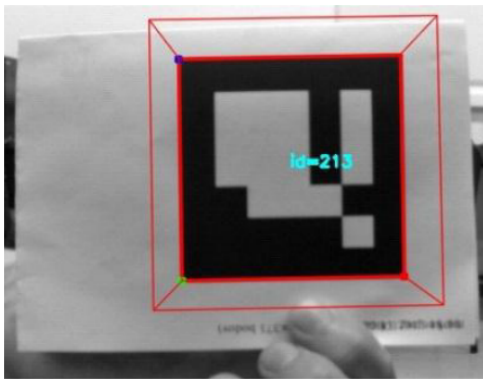
(a) Detected ArUco marker with ID
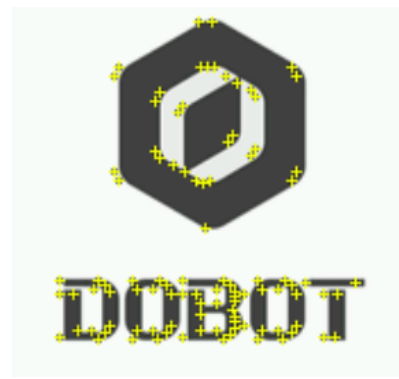
(b) Image with marked extracted features

Figure 2.6: Example of markers used for localization purposes

Next discussed type of localization is based on measurements obtained via external devices. There are many possible implementations of this approach but they could be categorised based upon the working principle.

Optical tracking of passive markers is best known in film industry. It is used in Motion Capture studios to transfer movement of actor to 3D modeled characters. It requires usage of multiple high-speed cameras with infrared illuminators fixed around the measurment area to traingulate a reflective marker position. Successful capture of tracked point by at least two camera give sub-milimeter precision. To avoid situation where marker is occluded by some obstacles localized object could be equipped with redundant ones. Disadvantage of such system is limitation of operating area caused by strength of the reflected light. There is also a way to increase to a certain extent range by using active markers which are light sources but this cause another problems with need to provide power to those.

From optical localization method there are also systems working without any markers. They are based on 3D depth cameras with are providing not only image but also give information about distance from objects. This allow to not only get position of tracked device but also could give feedback about scanned environment. This kind of systems generally have operation distance from 0.5 to 8 meters. Precision of positioning it is inversely proportional to it.

External localization could be also realized by measuring field strength or time of flight of electromagnetic waves. In this case at least three transmiters are needed to estabilish position. This method is slighty less accurate in compare to optical ones but have huge advantage in the form of beeing able to track device without direct visibility. Also range and refresh rate of such systems is significantly larger than optical ones.

(a) Motion capture studio
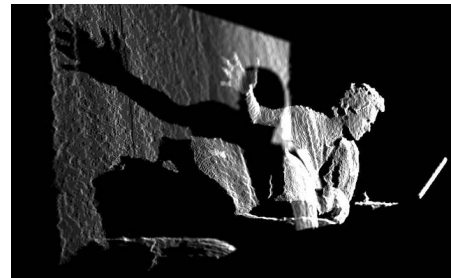
(b) 3D Camera Depth Data
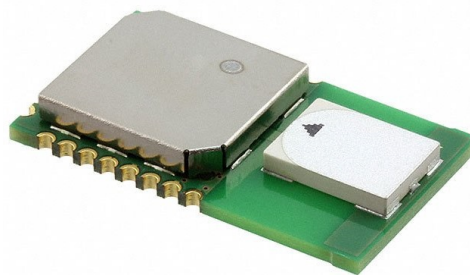
(c) UWB Localization module

Figure 2.7: External localization systems

Last presented type of positioning is Simultaneous Localization And Mapping (SLAM). This is technique with is using Laser Range Sensors or 3D Depth cameras to create map of environment. Basing on that data algorithm could calculate position and orientation in space. The big advantage of this solution is the ability to integrate all elements into one device with makes it almost limitless in terms of working area. Also additional gain is that the obtained data can be used to classify objects located in the environment. Unfortunately, such calculations takes huge amount of processing power so device need more expensive components to be able run in real-time. At Figure 2.8 could be seen 3D map of room created by using depth sensor and SLAM algorithm.

2.1.3   User control and interaction

Augmented Reality devices requires different approach to how user should use them. As the worker is no longer limited by his stationary desk there is need to change way of interaction. This section will discuss most common used methods.

Simplest HMD devices, where in the field of view we had only displayed text or images there is no need to add special types of interaction. In most cases on frame of device are placed buttons or small touchpad (Figure 2.9a) which allow user to control displayed content. This approach allows to reduce production costs so that it could have lower price tag and be able to get to more customers.

When device is providing ability to observe 3D objects in space then user need some controllers to be able to operate them. On the market most popular ones are using
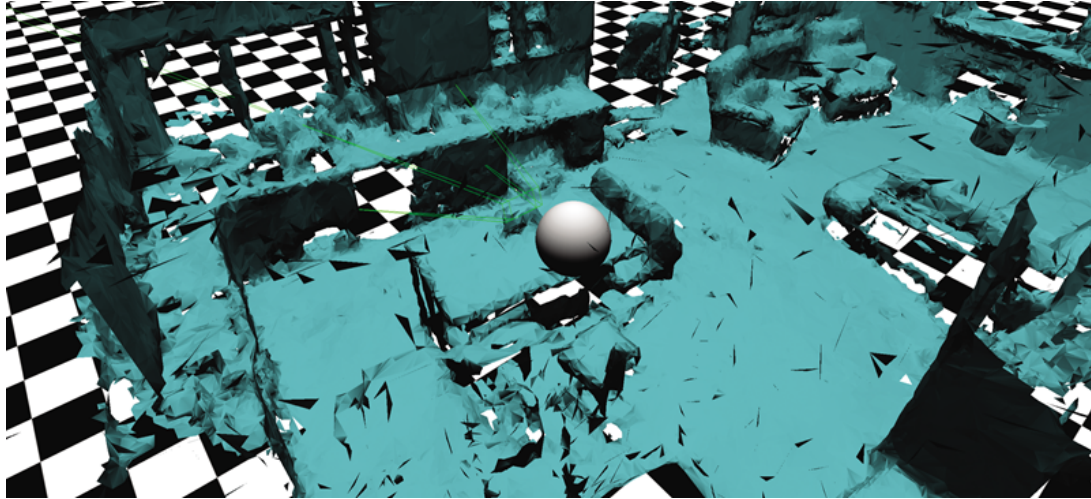
Figure 2.8: Room mapped using SLAM algorithm with usage of Microsoft Hololens

optical tracking (Figure 2.9b) but there are also those that uses electromagnetic fields. In addition, all are equipped with 6-DOF sensors to enchence stability of calculated position. This type of input is very precise but requires holding the controller which in some cases may hinder normal work.

The solution to this problem is to recognize gestures and track them (Figure 2.9c). This allow to operate with 3D environment while still having both hands free. Detection is carried out using depth cameras. Unfortunately like in case of using SLAM algorithms this method consume a lot of computing power so it is reserved only for the most expensive devices.

The last type of interaction with will be discussed is voice control. It is a very intricate topic due to diversity of humans dialects and accents. With current technology embedded computers can interpret only single words or predefined sentences however, using the resources of cloud services it is possible to extract information from context using natural language processing (NLP) algorithms. The disadvantage of this solution is requirement of continuous access to the Internet and quite slow response.

2.2   Products available on market

Augmented reality is still quite young technology. Omitting the fact that it was used in the army and civil aviation since the 1960s, it is now becoming available to industry and individual users. There are more and more companies trying their strength in this sector on the market. They use a variety of image display and interaction technologies, and develop newer and newer solutions. This section will show few choicen AR devices and softwares.

2.2.1   Hardware devices

The devices presented at table 2.1 are currently focused mainly on B2B cooperation. High prices and low maturity of the technology to the consumer market have an impact on this. Also there is no one unified environment between them so all applications are written to fulfill specific case. However, some devices can be bought in normal sales, so it is not a closed market for a regular customer.
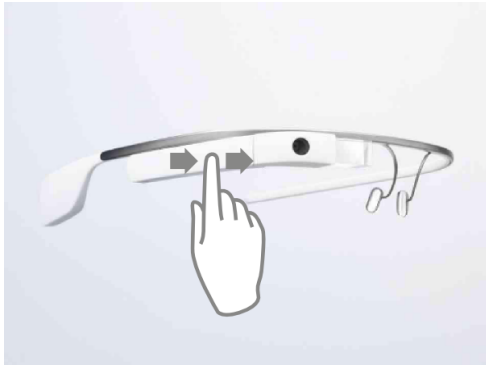
(a) Touchpad at Google Glass

(b) Mixed Reality Controllers

(c) Using gesture with Microsoft Hololens

2.2.2 Software solutions

The cheapest entry level to AR world is use of dedicated SDK's on computer, smartphone or tablet. On the market there are many ready to use solutions which cover different use cases. Most of them are commercial products, but there are also some open-source one. Table 2.2 shows the collected information about several of them.

Device name	Type of projection	Hardware	Sensors	Connectivity	Battery (life time)	Mechanical parameters	User input	Operating system	Capabilities	Price
Microsoft Hololens	Optical see-through 1920x1200 px per eye 40 degrees view angle	Intel x86 CPU Holographic Processing Unit 2 GB RAM 64 GB storage	Depth Sensor Camera 2MP Video Camera 4x environment understanding cameras IMU Ambient Light Sensor Microphone Onboard Speakers	Wi-Fi 802.11ac Bluetooth® 4.1 LE 3.5mm Jack micro-USB	4500 mAh 2-3 hours	Weight: 579 g Adjustable head band	Gesture input Wireless clicker Voice support Buttons on frame	Windows™ 10	Spatial sound Gaze tracking 6-DOF tracking Mapping of environment Teleoperation	$4500
Epson MOVERIO PRO BT-2000/2200	Optical see-through 960x540 px per eye 23 degrees view angle	TI OMAP 4460 1.2 GHz 1 GB RAM 8 GB internal storage 32 GB external microSD	5MP Stereo Camera GPS IMU Ambient Light Sensor Geomagnetic Sensor Microphone	Wi-Fi 802.11 b/g/n/a Bluetooth® 3.0 & BLE 3.5mm Jack micro-USB	2x 1240 mAh ~4 hours	Weight: 270-290 g Depending on version: - adjustable head band - rubber band	Wired controller Voice support	Android™ 4.0.4	Teleoperation 6-DOF tracking Task guidance	$2000
DAQRI Smart Glasses	Optical see-through 1360x768 px per eye 44 degrees view angle	6th Intel® Core™ m7 Dedicated vision processing unit 64 GB storage	Color Camera AR Tracking Camera Depth Sensor Camera IMU 2 Microphones with Active Noise Cancellation	Wi-Fi 802.11 a/b/g/n/ac Bluetooth® 2 USB 3.1 Type C Ports 3.5mm Jack	5800 mAh	Smart Glasses: 335 g Compute Pack: 496 g Adjustable head band	Gesture input Voice support	Custom solution	Mapping of environment Teleoperation 6-DOF tracking	Unknown
ARVIND SANJEEV LUMEN	Projection based 640x360px screen size: 10"-112"	Broadcom BCM2835 1GHz 512 MB RAM 32 GB external microSD	Color Camera Depth Sensor Camera	Wi-Fi 802.11 b/g/n Bluetooth® 4.1 LE micro-USB	1800 mAh ~30 min	Unknown	Gesture input	Linux	Objects recognition Mapping of environment	Unknown
Light Guide Systems Classic	Projection based	Any PC	Color camera	HDMI USB	AC Powered	Unknown	Camera feedback	Windows™	Augmentation of workstation Task guidance Automatic part assembly check Production time recording	Unknown
Magic Leap One	Optical see-through 1280x960 px per eye 50 degrees view angle	CPU: NVIDIA® Parker GPU: NVIDIA Pascal™ 8 GB RAM 128 GB Storage	Depth camera Environment understanding cameras IMU Microphone Onboard Speakers	Wi-Fi 802.11ac/b/g/n Bluetooth® 4.2 USB-C	~3 hours	Weight: 345 g	Gesture input 6-DOF tracked controllers Voice support	Lumin OS	Mapping of environment Teleoperation 6-DOF tracking	$2,295
Google Glass Enterprise Edition	Optical see-through 640360 px	Intel Atom processor 2 GB RAM 32 GB storage	5 MP camera IMU Ambient Light Sensor GPS & GLONASS Barometer Bone conduction transducer	Wi-Fi 802.11n/ac Bluetooth®	780 mAh ~8 hours	Weight: 36g	Voice commands Touchpad Mobile app	Android™ 4.4	Teleoperation Assistant Task guidance	$1,500

Table 2.1: List of selected AR goggles models

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
GPD in billions	235	225 bn	223 bn	323	423	523	624	725	826	924	1022

Table 2.2: Placeholder

Chapter 3

Applications of Augmented Reality

Chapter 4

Research of the subject

4.1 Used technologies

4.1.1 Unity

4.1.2 Vuforia

4.1.3 Robotic Operating System

4.2 Test results

4.2.1 Robotic arm

4.2.2 Mobile robot

Chapter 5

Summary

Bibliography