EAN 2019
57th conference on experimental stress analysis

Motion tracking in mixed reality

KRCMARIK David[1,a], CYRUS Jindrich[1,b], PETRU Michal[1,c] and KOCI Jan[1,d]

[1]Technical University of Liberec, Studentská 1402/2, 461 17, Czech Republic

[a]david.krcmarik@tul.cz, [b]jindrich.cyrus@tul.cz, [c]michal.petru@tul.cz, [d]jan.koci@tul.cz

Keywords: Tracking, Microsoft HoloLens, Communication unit, Industrial robots.

Abstract. We propose a viable approach for motion tracking of real objects within a holographic scene. Mixed reality applications are aspired to close cooperation of real and virtual worlds. For cooperation and interaction, we mainly need to know the precise position of real and virtual objects. These positions are likely to vary during the time. The movement can be tracked using traditional image recognition. However, this approach needs persistent visual contact with observed tracking image or a 3D object. Many mixed reality scenarios aim at industrial premises where tracked objects are equipped with a communication unit that can provide relevant information about movement and actual position. We propose a duplex communication interface based on TCP or UDP protocols that essentially helps with motion tracking of objects in the scene. The whole system can be refined by means of motion estimation based on the path already made and repetitive movements.

Introduction

The field of virtual and mixed reality devices is in recent years considerable widening. Most of the devices (HTC Vive, Oculus Rift) support virtual reality which does not require such a massive real-time image and sensor data processing as in the mixed reality. With the advent of Microsoft HoloLens [1] in 2016 new dynamically evolving environment started to gain attention. There are also other competitive platforms in the field of mixed reality (e.g. Magic Leap, Meta 2) which however does not provide such a rich experience as HoloLens. Most of the things with which human interacts are 3D objects. Commonly used 2D display screens can provide only a limited experience with such objects. Mixed reality provides true 3D visualization within real world hence deepens the perception of 3D objects and spatial interaction among them. Augmented reality by contrast is kind of mixed reality without interaction of virtual and real objects (e.g. Pokémon GO where real world is enhanced with new layer of perception).

Mixed reality combines real world objects with virtual objects into one scene. Head-mounted devices (HMD) are most commonly used since it is feasible in terms of technology and intuitive for users. The principle of combining real and virtual objects can be based either on video see-through (VST) [2] or optical see-through technology (OST). In video see-through approach the enviroment which is scanned with a RGBD (red-green-blue-depth) camera is enriched with virtual objects. The superimposed scene is then projected as in the case of virtual reality. OST devices are in our opinion more matured and natural with the advent of Microsoft HoloLens. User is able to see the surrounding environment immediatelly after placing the device on the head and the there is nearly no distortion of real world environment. The pros and cons of both concepts are described in many papers especially aimed to medical discipline [3].

When comparing mixed reality to virtual reality we think the future lies in mixed reality. The user is in direct contact with environment, there are significantly less reported sickness cases (characterized by eye strain, disorientation or even nausea) due to use of the device. OST-MDs are often self-contained and untethered, which means they does not need any additional computers, sensors or infrastructure.

Current challenges

In this paper we focus on OST-HMD Microsoft HoloLens because it is currently the most advanced device for mixed reality – Fig. 1. The device is used in many areas. Most of the applications are aimed to medicine and education.

Fig. 1: Microsoft HoloLens

In [4] an autopsy was performed with the help of HoloLens. As reported useful information was shown during the gross examination – radiographs, annotations, 3D scanned pathology specimens, whole slide image viewer, laboratory information system, etc. During the anatomic pathology various ways of communication with device were used (gestures, voice commands). Tele-collaboration (telepathology) allowing remotely access a senior pathologist for instructions was found very beneficial and possibly usable for other scenarios. The usage of mixed reality seems to have the biggest impact on the medicine – cases for neurosurgery [5] or cardiology [6] to name few are described. As already mentioned HoloLens proves to be especially rewarding in education. Many concepts in e.g. physics are hardly described since they are highly abstract – thermal conductivity, entropy. A control group of 59 students was used to test how the visualization of abstract physical quantities helps with cognitive understanding [7]. Results show small positive effect and authors pointed out that particular augmentation could significantly help for insight.

Particularly huge attention is made to other so far unusual fields of mixed reality usage. The term Industry 4.0 and Internet of things are often coined as future directions to which we need readjust our thinking in order to be competitive in 21 century. The number of devices connected to the Internet rapidly increases. Many companies incorporate sensors within their products sending information to servers in order to monitor the state of the device and possible predictive maintenance using so called big data [8]. This can save significant amount of resources and time. At the same time devices with bidirectional communication units can not only provide information but they can be controlled remotely by the end user or configured and adjusted by the provider for better performance. Since the amount of the devices increases mixed reality can provide a suitable way for their easy-to-use configuration even by untrained personnel [9]. HoloLens can help with discovering and selecting desired devices. The personnel has free hands to manipulate and adjust some features and at the same time needed manuals or teleassistance is available. Mixed reality with its ability of interconnecting real and virtual objects, connection to the Internet and large databases with information collected by Internet of Things (IoT) and availibility of remote technical assistance is a key component in developing Industry 4.0. An interesting project called Shipyard 4.0 was develeped which helps the engineers with interaction with product datasheets, maintenence procedures or quality control forms in real-world scenarios [10]. Both fog computing system and cloudlet approach were tested resulting in the cloudlet solution having better responsibility under high loads.

Important key to follow when designing immersive application for mixed reality are tele-immersion concepts – spatial immersion and presence [11]. In the paper we partly build on Minsky's tele-operated robot [12]. Mixed reality HoloLens provides two types of feedback to the user – visual and acoustical. We found that the most important aspect for visual feedback is location precision. It means that very disruptive for user is when the location of virtual object is not stable (in the terms of expected movement or stillness). On the other hand it can be beneficial when the user can immediately distinguish a real from a virtual via some cartoonish appearance or not perfectly rendered colors. 3D acoustical output significantly helps user with localization of virtual objects in the scene which are not directly visible and contribute to overall immersive feeling. Although there are motion controllers [13] available – Fig. 2, they do not provide haptic feedback to the user.

Fig. 2: Motion controllers

Haptic feedback [14] could be important for its intuitiveness and comfort. It allows the user to have also better performance when interaction with virtual content. In [15] and interesting concept enabling feeling of the material from which was made virtual object enabled the user to feel the weight. They have developed hRing which was put on the finger and they tracked the fingertips with the use of ARToolKit. A current state of the art in the field of HoloLens with haptic feedback is revised in [16]. It could be used for manipulation of virtual interfaces appended to real surfaces in the environment [17]. As already stated HoloLens are capable of recognition several gestures (build in gestures are 'tap', 'double tap' and 'drag and drop', system gesture is 'bloom') made by the user or even by other people in the scene. Since this feature is a software based more gestures could be possible added. In human-human interaction facial emotions are very important. Very interesting concept of detection user defined gestures is based on WiFi scanning There are even efforts [19] for gesture recognition of emotions based on databank of known correspondences of human feeling and their facial expression based on deep learning.

Security of usage HoloLens is often neglected. There is a common understanding that the data from device sensors have to be secured in order to avoid misuse from different applications possibly installed in the device or even from third subjects connected via WiFi or Bluetooth to HoloLens. It is important to realize that the users field of view can be also filled with much of information and objects coming from HoloLens and can possible be obtrusive or even dangerous since other information could be hidden (from real world or from other applications in HoloLens). This is called an virtual output security [20].

It was already mentioned that partly exceptional are HoloLens due to its ability precise itself localization and tracking of other real object in the scene. The fundamental concept is described in [21]. Unfortunately the whole thing is more complicated for OST devices [22] since one have to take into account the movement of eye. AR device have to be capable of 6-degree-of-freedom (6 DoF) movement tracking (x, y, z, roll, pith, yaw). Every pixel in the virtual scene (with index B) has to be transformed using position and rotation into a real scene (with index A) according to an equation

(1)

where R stand for rotation and t for translation. In the case of moving eye the coefficients for transformation (1) have to be heavily and periodically computed and updated. For that reason the field of view in HoloLens is limited in order to minimize eye movement – Fig. 3.

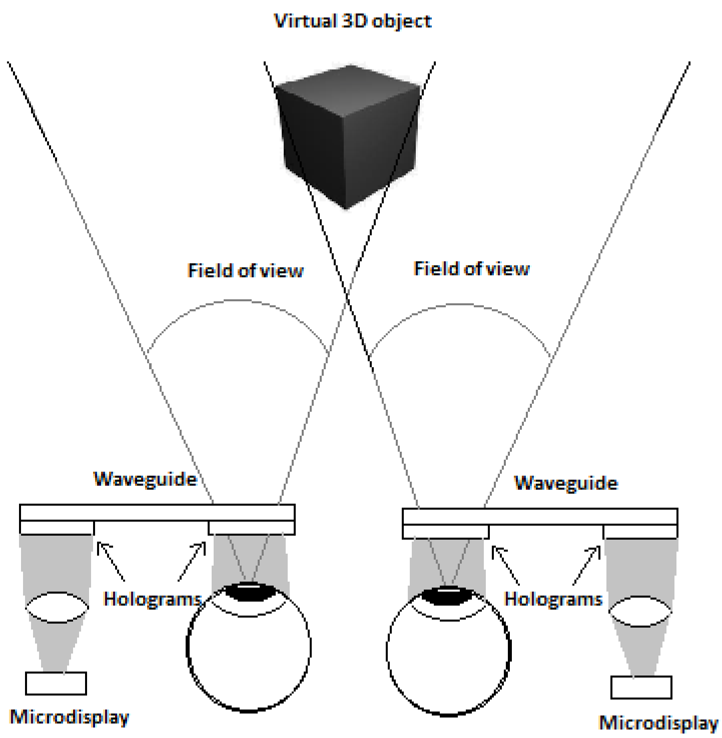

Fig. 3: Schematic view of moving eye problem for OST

If we want to used mixed reality platform even for applications where extremely accurate positioning and stability is demanded [23] we have to improve overall location error – for example using calibration [24]. Online temporal tracking [25] or well-know Kalman filter based solutions [26] seem to give promising results. Descriptor utilization is a concept where objects are compared from different views, classified and the system is in this way trained in such a way that only 16-dimensional descriptors can help with object recognition and their 3D pose estimation [27].

It is necessary to mention that there are also attempts to use low-cost RFID based devices which can to a certain extent track 3D objects in space [28]. A signal strength indicator (RSSI) was measured using commercial off-the-shelf device called RFTrack which uses discrete wavelet transform for detection target motion. Unfortunatelly precision of only 32 cm was reported. Tele-rehabilitation was tested using Microsoft Kinect (predecessor of HoloLens) based on optical system with the help of calibration [29]. The calibration is provided as complementary information concerning bone lengths, position of elbows etc. A Google Tango based project with additional visual markers [30] is another way of precise localization of moving devices in the scene which does not require any additional infrastructure.

As shown above there are many challenges in newly appearing platform of mixed reality. The thing of paramount importance remains precise localization which is not processor demanding and without needed ambient infrastructure.

Precise position and motion tracking

It is important to realize that in many situations where HoloLens are used there are usually a few devices, which can communicate with the rest of world. They can be intelligent devices robot-type or vehicle-type. HoloLens is a vital platform for simulation interaction between these devices and virtual world saving a lot of resources and time.

We can exploit such natural functionality to build a one-to-many bidirectional communication link possessing needed information about the scene. The whole setup assumes some kind of network for wireless connection – either WiFi or Bluetooth. Other communication standards can be easily implemented using separate converter like black-boxes. It is also necessary to count with the scenario that communication is temporally unavailable. In this case we use standard tracking camera for defined feature detection with the Vuforia backend for processing images. Previously the depth information was reported to be acquired using Kinect Fusion [31, 32]. In our setup we used only an industrial camera.

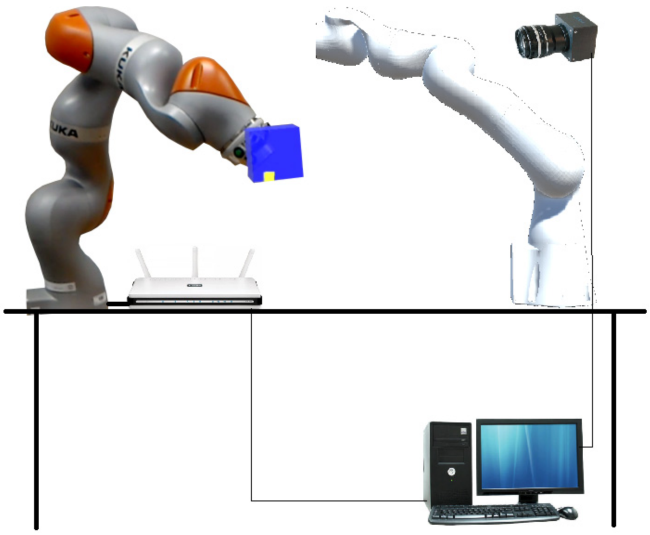

Fig. 4: Schematic view of setup

There was used high-throughput UDP communication in the HoloLens application to instantly report position of the end point of moving object. The goal of the whole project was to instruct the virtual robot using a gaze to pick particular cube (red or blue), handle the cube from a bowl to the real robot and the real robot took the virtual cube and placed it on belt. During the experiment, we have simulated network instability.

We have compared TCP and UDP base communication. Average time delay exhibited using TCP protocol was about 24 ms while for UDP communication it was only 7 ms.

Here is the example of the C# code used to establish the UDP communication:

```
private async void StartClient(bool sendWelcomeMessage = false) {
       try {
              //Create the DatagramSocket and establish a connection.
              clientDatagramSocket
                     = new Windows.Networking.Sockets.DatagramSocket();

              clientDatagramSocket.MessageReceived
                     += ClientDatagramSocket_MessageReceived;

              AddLog("client is about to bind...");
              await clientDatagramSocket
                            .BindServiceNameAsync(Cfg.UdpPort.ToString());
              AddLog(string.Format(
                            "client is bound to port number {0}", Cfg.UdpPort));

              if (sendWelcomeMessage) {
                     // Send a request to the echo server.
                     string request = "HoloLens are connected to you!";
                     // The server hostname that we connect to.
                     // only for confirmation message
                     var hostName = new Windows.Networking.HostName("SERVER");
                     using (var serverDatagramSocket
                            = new Windows.Networking.Sockets.DatagramSocket()) {
                            using (Stream outputStream
                                   = (await serverDatagramSocket.GetOutputStreamAsync
                                          (hostName,Cfg.UdpPort.ToString()))
                                          .AsStreamForWrite()) {
                                          using (var streamWriter
                                                 = new StreamWriter(outputStream)) {
                                          await    streamWriter.WriteLineAsync(request);
                                                       await streamWriter.FlushAsync();
                                          }
                            }
                     }
                     AddLog(string.Format(
                            "client sent the request: \"{0}\"", request));
              }
       }
}
```

The camera was an RGB Basler acA1600-20gc with Ethernet interface connected to a standard PC. We used Vuforia object tracking and this information helped during a loss of the communication between robot and HoloLens.

We have simulated also the loss of the communication on both lines – from camera and from robot to HoloLens. HoloLens immediately recognized this outage and switched to the previously learned trajectory. We assume that this scenario can apply to many cases with industrial robots, because the moving objects carry out periodical movements in the scene very often. The device is tracking the objects of interest in the scene and buffers the movements to the specific buffers (6 DoF described movement). If there is broken the communication line to HoloLens and the data about the actual robot position is missing, then the device tries to extrapolate the movement in order seamlessly provide visual content to the user.

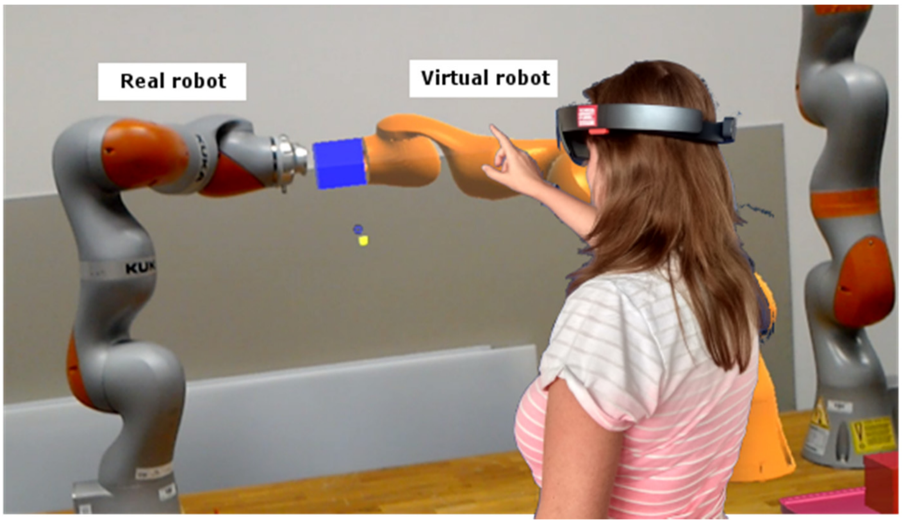

Fig. 5: Handling a virtual cube to the real robot

Summary

We presented a way of tracking moving object in the scene based on our already created infrastructure. Many applications with HoloLens are in environments, which are enriched with intelligent devices, or IoT enabled devices. Such environments can be readily exploited to provide necessary information of real object current location or movement directly to mixed reality devices. Hereby it is not necessary to perform demanding tracking using built-in sensors and process the data using processor-intensive algorithms directly within the mixed reality device. The strongest side of HoloLens is to visualize the virtual objects and to process spatial mapping in order to obtain the smooth and stable scene in the real time. Use of external precise data on mixed reality devices leads to the applications that consume fewer resources but are more efficient and produce breathtaking experience. Proposed method of very fast communication using UDP protocol can be a very handy tool, which is easily embedded and provides satisfying results. The time lag was as reported minimized so that user was unable to see any unnatural movements of virtual cube which was moved by the real robot.

Acknowledgements

The result was obtained through the financial support of the Ministry of Education, Youth and Sports of the Czech Republic and the European Union (European Structural and Investment Funds - Operational Programme Research, Development and Education) in the frames of the project "Modular platform for autonomous chassis of specialized electric vehicles for freight and equipment transportation", Reg. No. CZ.02.1.01/0.0/0.0/16_025/0007293.

References

[1] M.A. Hoffman, The future of three-dimensional thinking, Science 353 (2016) 876.

[2] J.C. Wang, et al., Video see-through augmented reality for oral and maxillofacial surgery, International Journal of Medical Robotics and Computer Assisted Surgery 13 (2017).

[3] J.M. Carmen, J. Calatrava, An augmented reality system for the treatment of phobia to small animals viewed via and optical see-through HMD: Comparison with a similar

system viewed via a video see-through HMD, International Journal of Human-Computer Interaction 27 (2011) 436-449.

[4] M.G. Hanna, et al., Augmented reality technology using Microsoft HoloLens in anatomic pathology, Archives of Pathology and Laboratory Medicine 142 (2018) 638-644.

[5] F. Incekara, et al., Clinical feasibility of a wearable mixed-reality device in neurosurgery, World Neurosurgery 118 (2018) E442-E427.

[6] V. Garcia-Vasquez, et al., Navigation and visualization with HoloLens in endovascular aortic repair, Innovative Surgical Sciences 3 (2018) 167.

[7] M.P. Strzys, et al., Physics holo.lab learning experience: Using smartglasses for augmented reality labwork to foster the concepts of the heat conduction, European Journal of Physics 39 (2018).

[8] J. Maktoubian, Proposing a streaming Big Data analytics (SBDA) platform for condition based maintenance (CBM) and monitoring transportation systems, EAI Endorsed Transactions on Scalable Information Systems 4 (2017).

[9] G. Alce, et al., AR as a user interface for the internet of things – comparing three interaction models, in: 16th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), France, 2017.

[10] T.M. Fernandez-Carames, et al., A fog computing and cloudlet based augmented reality system for the industry 4.0 shipyard, Sensors 18 (2018).

[11] S. Ohl, Tele-immersion concepts, IEEE Transactions on Visualization and Computer Graphics 24 (2018) 2827-2842.

[12] M. Minsky, Telepresence, Omni 4 (1980) 45-51.

[13] K.R. Moser, J.E. Swan, Evaluation of user-centric optical see-through head-mounted display calibration using a leap motion controller, in: IEEE Symposium on 3D User Interfaces (3DUI), South Carolina, USA, 2016.

[14] C. Pacchierotti, et al., Wearable haptic systems for the fingertip and the hand: Taxonomy review and perspectives, IEEE Transactions on Haptics 10 (2017) 580-600.

[15] S. Scheggi, G. Salvietti, D. Prattichizzo, Shape and weight rendering for haptic augmented reality, in: 19th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Italy, 2010.

[16] L. Meli, et al., Combining wearable finger haptics and augmented reality: User evaluation using and external camera and the Microsoft HoloLens, IEEE Robotics and Automation Letters 3 (2018) 4297-4304.

[17] R. Xiao, et al., MRTouch: Adding touch input to head-mounted mixed reality, in: 25th IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR), Germany, 2018.

[18] H. Abdelnasser, M. Youssef, K.A. Harras, A ubiquitous WiFi-based gesture recognition system, in: 34th IEEE Conference on Computer Communications (INFOCOM), Hong Kong, China, 2015.

[19] D. Mehta, M.F.H. Siddiqui, A.Y. Javaid, Facial emotion recognition: A survey and real-world user experiences in mixed reality, Sensors 18 (2018).

[20] K. Lebeck, et al., Securing augmented reality output, in: 38th IEEE Symposium on Security and Privacy (SP), California, USA, 2017.

[21] W. Robinett, R. Holloway, The visual display transformation for virtual reality, Presence Teleoperators and Virtual Environments 4 (1995).

[22] J. Grubert, et al., A survey of calibration methods for optical see-through head-mounted displays, IEEE Transactions on Visualization and Computer Graphics 24 (2018) 2469-2662.

[23] R. Vassallo, et al., Hologram stability evaluation for Microsoft® Hololens™, in: Conference on Medical Imaging – Image Perception, Observer Performance, and Technology, Florida, USA, 2017.

[24] M. Garon, et al., Real-time high resolution 3D data on the HoloLens, in: 15th IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Mexico, 2016.

[25] D.J. Tan, et al., A versatile learning-based 3D temporal tracker: Scalable, robust, online, in: IEEE International Conference on Computer Vision, Chile, 2015.

[26] Q. Yang, D.G. Taylor, G.D. Durgin, Kalman filter based localization and tracking estimation for HIMR RFID systems, in: IEEE International Conference on RFID, Florida, USA, 2018.

[27] P. Wohlard, V. Lepetit, Learning descriptors for object recognition and 3D pose estimation, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPT), Massachusetts, USA, 2015.

[28] L. Liyao, et al., Accurate device-free tracking using inexpensive RFIDs, Sensors 18 (2018) 2816.

[29] K. Yu, et al., On the accuracy of low-cost motion capture systems for range of motion measurements, in: Conference on Medical Imaging – Imaging Informatics for Healthcare, Research and Applications, Texas, USA, 2018.

[30] B. Marques, et al., Evaluating and enhancing Google Tango localization in indoor environments using fiducial markers, in: 18th IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), Portugal, 2018.

[31] R.A. Newcombe, et al., KinectFusion: Real-time dense surface mapping and tracking, in: 10th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR), Switzerland, 2011.

[32] S. Izadi, et al., KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera, in: ACM Symposium on User Interface Software an Technology (UIST), California, USA, 2011.