Freie Universität Berlin
Fachbereich Mathematik und Informatik
Takustraße 9, 14195 Berlin

MASTER THESIS

USER POSITION PREDICTION IN 6-DOF MIXED REALITY APPLICATIONS USING ARTIFICIAL RECURRENT NEURAL NETWORK

VORHERSAGE DER BENUTZERPOSITION IN 6-DOF-MIXED-REALITY-ANWENDUNGEN UNTER VERWENDUNG EINES KÜNSTLICHEN REKURRENTEN NEURONALEN NETZWERKS

Oleksandra Baga

Freie Universität Berlin
Matrikelnummer 5480722
Master Computer Science
E-Mail: oleksandra.baga@gmail.com

Prof. Dr. First Supervisor
Fachbereich Mathematik und Informati
Freie Universität Berlin

Prof. Dr. Second Supervisor
Fachbereich Mathematik und Informati
Freie Universität Berlin

Statutory Declaration

I herewith formally declare that I have written the submitted master thesis independently. I did not use any outside support except for the quoted literature and other sources mentioned in the paper.

I clearly marked and separately listed all of the literature and all of the other sources which I employed when producing this academic work, either literally or in content.

I am aware that the violation of this regulation will lead to failure of the thesis.

29.06.2022..... Oleksandra Baga

Acknowledgments

This thesis was created in cooperation with the Fraunhofer Heinrich Hertz Institute.

First and foremost I would like to thank Prof FU Berlin, who supervised my thesis. Thank you very much for the helpful suggestions and constructive criticism.

A special thanks goes to the researcher Fraunhofer Heinrich Hertz Institute, Serhan Gül, who suggested an exciting topic for a research, which I was allowed to choose for my master thesis. I would like to express my sincere thanks for the commitment and the consultation during the preparation of this thesis.

Contents

List of Figures

Listings

List of Abbreviations

AR	Augmented Reality
CNN	Convolutional Neural Network
CPU	Central processing unit
DoF	Degree of freedome
DL	Deep Learning
KF	Kalman Filter
GRU	Gated Recurrent Unit
HMD	Head-Mounted-Display
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
LAT	Look ahead time
LSTM	Long-Short-Term Memory
M2P	Motion-to-Photon
MAE	Mean Absolut Error
MEC	Mobile Edge Computing
ML	Machine Learning
MR	Mixed Reality
NLP	Natural Language Processing
ReLu	Rectified Linear Unit
RNN	Recurrent Neural Network
RTT)	Round-trip time
SDG	Stochastic Gradient Descent
VR	Virtual Reality
3-DoF	Three degree of freedom
6-DoF	Six degree of freedom

1 Introduction

This thesis is focusing on designing and evaluation of the approach for the prediction of human head position in a 6-dimensional degree of freedom (6-DoF) of Extended Reality (XR) applications for a given look-ahead time (LAT) in order to reduce the Motion-to-Photon (M2P) latency of the network and computational delays. At the beginning of the work the existing 3-DoF as well as 6-DoF methods were analyzed, and their similarities differences were taken into account when a proposed Recurrent Neural Network-based predictor was developed. The investigation of different neural network architectures and the improvement of head motion prediction is the main goal of this thesis. Proposed approach was evaluated and at the end of the work the obtained results were discussed and the suggestions for future work were done.

The correct and fast head movement prediction is a key to provide a smooth and comfortable user experience in VR environment during head-mounted display (HDM) usage. The recent improvements in computer graphics, connectivity and the computational power of mobile devices simplified the progress in Virtual Reality (VR) technology. The way users can interact with their devices changed dramatically. With new technologies of VR environment user becomes the main driving force in deciding which portion of media content is being displayed to them at any time of interaction with VR Applications [13]. Until recently the high-quality experiences with modern Augmented Reality (AR) and VR systems were not widely presented in home usage and were mainly used in research labs or commercial setups. The hardware for displaying the VR environment was once extremely expensive but recent years became more broadly accessible and the 6-DoF VR headset designed for the end-user were released[1]. It is possible now to experience virtual reality scenes and watch new type of volumetric media at home and the market interest for development VR and AR applications expected to be huge next years.

Although all mentioned above improvements, rendering of volumetric content remains very demanding task for existing devices. It is possible to decrease the computational load on the client device by offloading of the task to a server infrastructure and than by sending the rendered 2D content instead of volumetric data [9]. The 2D view must correspond the current position and orientation of a user. Due to the added in this approach network latency and processing delays the rendered 2D image can appear even later on the display than with usage of local rendering system. The reducing the Motion-to-Photon (M2P) latency by prediction the future

[1] https://medium.com/@DAQRI/motion-to-photon-latency-in-mobile-ar-and-vr-99f82c480926

user position and orientation for a look-ahead time (LAT) at remote server and sending the corresponding rendered view to a client could be very effective solution for 6-DoF XR application with immersive media.

1.1 Problem statement

The existing on this moment virtual environments can be divided into two main groups. Depending on position of the user and their ability to move inside the VR environment the 3-DoF and 6-DoF.

While some efforts to reduce the computational latency on the client side are being already made, the new technique of the rendering on a cloud server was recently presented and covered in this thesis. However, cloud-based streaming further increases the delay and M2P latency. Thus it is important to create the method of the viewer's head pose and orientation for a look-ahead time (LAT) equal or larger to the M2P latency of the network round-trip time (RTT) the new challenges of the head motion prediction arises

1.2 Motivation for the research

1.3 Structure of the thesis

The organization of this thesis is as follows. The literature review chapter introduces the concepts of XR technologies and principles of motion prediction algorithms. It follows an overview of previous research.

Chapter 1 - Introduction.
The following chapter introduces the preprocessing pipeline, architecture, and evaluation process used to develop ahead motion prediction algorithm.

2 Background

This chapter introduces theoretical background of the presented research problem. First, the concept of mixed reality (MR) and the relation of this huge topic to the research field is presented, followed by an introduction of six degree of freedom (6-DoF) environment and the difference to the three degree of freedom (3-DoF). The term motion-to-photon latency (M2P) is covered, followed by a short discussion about an influence of M2P latency on the decreasing of user experience. The new developed cloud-based rendering and streaming approach is shortly discussed in this chapter. The last section of this chapter highlights challenges with the prediction of viewer's head pose that arises in modern XR applications in connection especially with the added network latency due the using of remote cloud server for computational offload.

2.1 Mixed reality

2.2 Six degrees of freedom

2.3 Motion-to-photon latency

2.4 Cloud-based volumetric video streaming

2.5 Challenges of head motion prediction

In the last decade, Recurrent Neural Network (RNN) algorithms have been adopted for motion prediction of 3D sequences.

3 Related work

This chapter introduces the background of the thesis to investigate head motion prediction. First, the concept of extended reality (XR) related to the research problem is presented, followed by introducing different head motion prediction approaches.

3.1 Time series methods

This chapter introduces the background of the thesis to investigate head motion prediction.

3.2 Kalman Filter

Microsoft HoloLens, known under development as Project Baraboo, are a pair of mixed reality smartglasses developed and manufactured by Microsoft.

3.3 Deep Learning Algorithms

present different NN architectures and explains why RNN is used

4 Data and Model

It allows web pages to be updated asynchronously by exchanging data with a web server behind the scenes.

4.1 6-DoF Dataset

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

4.1.1 Data collection from HMD

It allows web pages to be updated asynchronously by exchanging data with a web server behind the scenes.

4.1.2 Data Exploration

Data analysis AVG linear velocity position, plots

4.1.3 Data preprocessing

It allows web pages to be updated asynchronously by exchanging data with a web server behind the scenes.

4.2 Network Architecture

The history of artificial neural networks is filled with colorful, creative individuals from a variety of fields, many of whom struggled for decades to develop concepts that we now take for granted.

4.3 Network input

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

4.4 Training methods

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

Implementation and experiments

5

Here some code for my super neural network. The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

```
class StudentFactory(DjangoModelFactory):
        class Meta:
                model = Student
        student_card = factory.SubFactory(StudentCardFactory)
        first_name = factory.Faker('first_name')
        second_name = factory.Faker('last_name')
```

Listing 5.1: StudentFactory

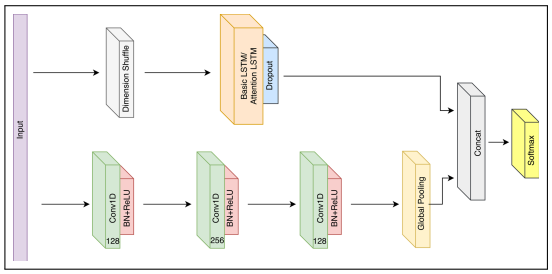

Figure 1: LSTM Fully Convolutional Networks for Time Series Classification

You might already know that you want to apply an established theory or set of theories to a specific context (for example, reading a literary text through the lens of critical race theory, or using social impact theory in a market research project).

5.1 Implementation

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

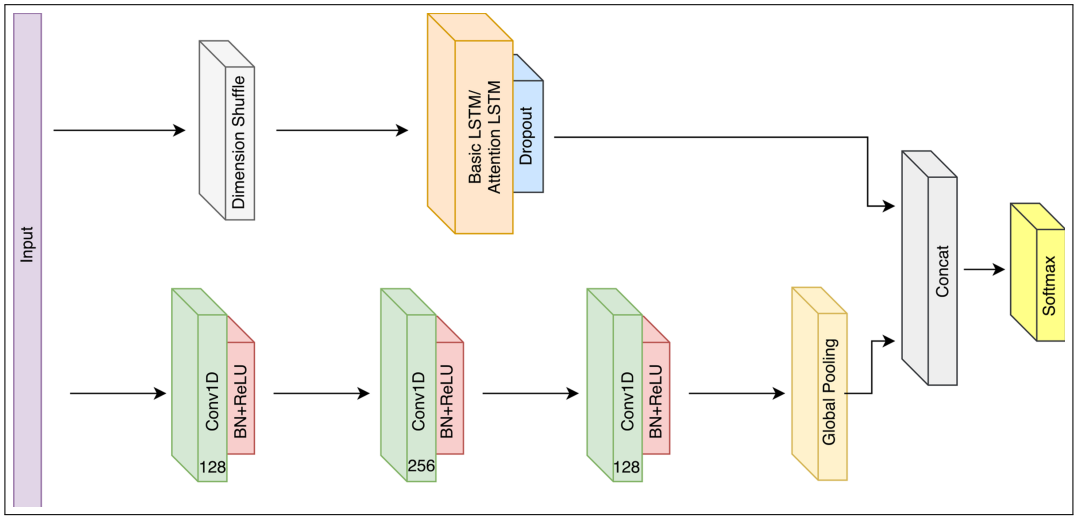

Figure 2: LSTM FCN WRAP

5.2 Experiments

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

5.3 Evaluation metrics

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

5.4 Results

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

6 Analysis

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

6.1 Limitations

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

6.2 Conclusion

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

6.3 Suggestions for future work

The artificial neural networks discussed in this text are only remotely related to their biological counterparts. In this section we will briefly describe those characteristics of brain function that have inspired the development of artificial neural networks.

Glossary

AJAX

AJAX (asynchrones Javascript und XML) ist der allgemeine Name für Technologien, mit denen asynchrone Anforderungen (ohne erneutes Laden von Seiten) an den Server gestellt und Daten ausgetauscht werden können. Da die Client- und Serverteile der Webanwendung in verschiedenen Programmiersprachen geschrieben sind, müssen zum Austausch von Informationen die Datenstrukturen (z. B. Listen und Wörterbücher), in denen sie gespeichert sind, in das JSON-Format konvertiert werden.

Bibliography

[1] A. Deniz Aladagli, Erhan Ekmekcioglu, Dmitri Jarnikov, and Ahmet Kondoz. *Predicting head trajectories in 360° virtual reality videos*. https://ieeexplore.ieee.org/document/8251913. date access on 29.03.22. 2017. DOI: 10.1109/IC3D.2017.8251913.

[2] R.S. Allison, L.R. Harris, M. Jenkin, U. Jasiobedzka, and J.E. Zacher. *Tolerance of temporal delay in virtual environments*. https://www.researchgate.net/publication/2945506. date access on 17.03.22. 2001. DOI: 10.1109/VR.2001.913793.

[3] Tamay Aykut, Christoph Burgmair, Mojtaba Leox Karimi, and Eckehard Steinbach. *Delay Compensation for a Telepresence System With 3D 360 Degree Vision Based on Deep Head Motion Prediction and Dynamic FoV Adaptation*. https://arxiv.org/abs/2007.14084. date access on 23.02.22. 2018. DOI: 10.1109/WACV.2018.00222.

[4] Yun-Kai Chang, Mai-Keh Chen, Yun-Lun Li, Hao-Ting Li, and Chen-Kuo Chiang. *6DoF Tracking in Virtual Reality by Deep RNN Model*. https://ieeexplore.ieee.org/document/9394069. date access on 02.04.22. 2020. DOI: 10.1109/IS3C50286.2020.00057.

[5] Howie Choset et al. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. the MIT Press, 2006, p. 630. ISBN: 978-0262033275.

[6] Alessandro Crivellari and Euro Beinat. *LSTM-Based Deep Learning Model for Predicting Individual Mobility Traces of Short-Term Foreign Tourists*. https://www.researchgate.net/publication/338377314. date access on 08.04.22. 2020. DOI: 10.3390/su12010349.

[7] Richard O. Duda, Peter E. Hart, and David G. Stork. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. Wiley-Interscience; 2. edition, 2000, p. 688. ISBN: 978-0471056690.

[8] Serhan Guel, Sebastian Bosse, Dimitri Podborski, Thomas Schierl, and Cornelius Hellge. *Kalman Filter-based Head Motion Prediction for Cloud-based Mixed Reality*. https://arxiv.org/abs/2007.14084. date access on 19.02.22. 2020. DOI: 0.1145/3394171.3413699.

[9] Serhan Gül, Dimitri Podborski, Thomas Buchholz, Thomas Schierl, and Cornelius Hellge. *Low-latency Cloud-based Volumetric Video Streaming Using Head Motion Prediction*. https://arxiv.org/abs/2001.06466. date access on 19.02.22. 2020. DOI: 0.1145/3394171.3413699.

[10] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. *LSTM Fully Convolutional Networks for Time Series Classification*. https://arxiv.org/abs/1709.05206. date access on 14.04.22. 2017. DOI: 10.48550/arXiv.1709.05206.

[11] Anh Nguyen, Zhisheng Yan, and Klara Nahrstedt. *Your Attention is Unique: Detecting 360-Degree Video Saliency in Head-Mounted Display for Head Movement Prediction*. https://www.researchgate.net/publication/328370817. date access on 15.03.22. 2018. DOI: 10.1145/3240508.3240669.

[12] Feng Qian, Lusheng Ji, Bo Han, and Vijay Gopalakrishnan. *Optimizing 360 video delivery over cellular networks*. https://dl.acm.org/doi/10.1145/2980055.2980056. date access on 12.03.22. 2016.

[13] Silvia Rossi, Irene Viola, Laura Toni, and Pablo Cesar. *A New Challenge: Behavioural Analysis Of 6-DOF User When Consuming Immersive Media*. https://ieeexplore.ieee.org/document/9506525. date access on 03.04.22. 2021. DOI: 10.1109/ICIP42928.2021.9506525.

[14] Silvia Rossi, Irene Viola, Laura Toni, and Pablo Cesar. *From 3-DoF to 6-DoF: New Metrics to Analyse Users Behaviour in Immersive Applications*. https://www.researchgate.net/publication/357172010. 1-7. 2021. date access on 13.04.22.

[15] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. the MIT Press, 2006, p. 672. ISBN: 978-0262201629.

[16] Dingqi Yang, Benjamin Fankhauser, Paolo Rosso, and Philippe Cudre-Mauroux. *Location Prediction over Sparse User Mobility Traces Using RNNs: Flashback in Hidden States*. https://www.researchgate.net/publication/338377314. date access on 07.04.22. 2020. DOI: 10.24963/ijcai.2020/302.

[17] Emin Zerman, Radhika Kulkarni, and Aljosa Smolic. *User Behaviour Analysis of Volumetric Video in Augmented Reality*. https://ieeexplore.ieee.org/document/9465456. date access on 13.04.22. 2021. DOI: 10.1109/QoMEX51781.2021.9465456.