VISUAL 2019

The Fourth International Conference on Applications and Systems of Visual Paradigms

ISBN: 978-1-61208-724-5

June 30 – July 4, 2019

Rome, Italy

VISUAL 2019 Editors

William Hurst, Senior Lecturer, Liverpool John Moores University, UK

Vittoria Bruni, Researcher, Sapienza - Rome University, Italy

VISUAL 2019

Foreword

The Fourth International Conference on Applications and Systems of Visual Paradigms (VISUAL 2019), held between June 30 – July 4, 2019 - Rome, Italy continued the inaugural event in putting together complementary domains were visual approaches are considered in a synergetic view.

Visual paradigms were developed on the basis of understanding the brain's and eye's functions. They spread over computation, environment representation, autonomous devices, data presentation, and software/hardware approaches. The advent of Big Data, high speed images/camera, complexity and ubiquity of applications and services raises several requests on integrating visual-based solutions in cross-domain applications.

We take here the opportunity to warmly thank all the members of the VISUAL 2019 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to VISUAL 2019. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the VISUAL 2019 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that VISUAL 2019 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of visual oriented technologies.

We are convinced that the participants found the event useful and communications very open. We also hope that Rome provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city

VISUAL 2019 Chairs

VISUAL Steering Committee

Vijayan K. Asari, University of Dayton, USA
Robert S. Laramee, Swansea University, UK
Luciano Pereira Soares, Insper, Brazil
Jibonananda Sanyal, Oak Ridge National Laboratory | Mississippi State University, USA
Irene Yu-Hua Gu, Chalmers University of Technology, Sweden

VISUAL Industry/Research Advisory Committee

Kresimir Matkovic, VRVis Research Center, Vienna, Austria
Jinrong Xie, eBay Inc., USA
Afzal Godil, National Institute of Standards and Technology, Gaithersburg, USA
Silvia Biasotti, Consiglio Nazionale delle Ricerche, Genova, Italy
Carolin Helbig, Helmholtz Centre for Environmental Research (UFZ), Germany

VISUAL 2019

Committee

VISUAL Steering Committee

Vijayan K. Asari, University of Dayton, USA
Robert S. Laramee, Swansea University, UK
Luciano Pereira Soares, Insper, Brazil
Jibonananda Sanyal, Oak Ridge National Laboratory | Mississippi State University, USA
Irene Yu-Hua Gu, Chalmers University of Technology, Sweden

VISUAL Industry/Research Advisory Committee

Kresimir Matkovic, VRVis Research Center, Vienna, Austria
Jinrong Xie, eBay Inc., USA
Afzal Godil, National Institute of Standards and Technology, Gaithersburg, USA
Silvia Biasotti, Consiglio Nazionale delle Ricerche, Genova, Italy
Carolin Helbig, Helmholtz Centre for Environmental Research (UFZ), Germany

VISUAL 2019 Technical Program Committee

Amr Abdel-Dayem, Laurentian University, Canada
Andrew Adamatzky, University of the West of England, Bristol, UK
Athos Agapiou, Eratosthenes Research Centre / Cyprus University of Technology, Cyprus
Konstantin Aksyonov, Ural Federal University, Russia
Mokhled S. AlTarawneh, Mutah University, Jordan
Leyteris Anastasovitis, Information Technologies Institute - Centre for Research and Technology Hellas (ITI-CERTH) / University of Macedonia, Greece
Djamila Aouada, University of Luxembourg, Luxembourg
Vijayan K. Asari, University of Dayton, USA
George Baciu, The Hong Kong Polytechnic University, Hong Kong
Yufang Bao, Fayetteville State University, USA
Ardhendu Behera, Edge Hill University, UK
Saeid Belkasim, Georgia State University, Atlanta, USA
Achraf Ben-Hamadou, Digital Research Center of Sfax, Tunisia
Yannick Benezeth, Univ. Bourgogne Franche-Comté, France
Stefano Berretti, University of Florence, Italy
Silvia Biasotti, Consiglio Nazionale delle Ricerche, Genova, Italy
Hans-Peter Bischof, Rochester Institute of Technology, USA
Kadi Bouatouch, IRISA / Unviversity of REnnes 1, France
Vittoria Bruni, University of Rome "La Sapienza", Italy
Franco Alberto Cardillo, Institute for Computational Linguistics - National Research Council, Pisa, Italy
Jacopo Cavazza, Istituto Italiano di Tecnologia, Italy

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Simulating Household Electricity Consumption

Mutinta Mwansa, William Hurst, Carl Chalmers, Yuanyuan Shen, Casimiro A. Curbelo Montañez
Department of Computer Science
Liverpool John Moores University
Byrom Street
Liverpool, L3 3AF, UK
Email: M.Mwansa@2017.ljmu.ac.uk, {W.Hurst, C.Chalmers, Y.Shen, C.A.CurbeloMontanez}@ljmu.ac.uk

***Abstract*— The smart grid is an advanced infrastructure that leverages communication technology, data analytics and cloud computing to control the distribution and consumption of energy. Smart grid systems include producers, consumers and actors, to ensure a resource saving and efficient electrical network. Within the smart grid, the smart meter records the consumption of electricity in private homes and businesses accurately. The data generated can be used to provide an insight into social demographics, household behaviour patterns, social clusters, general energy consumption patterns and a variety of value-added services. However, one of the biggest challenges for researchers in this area is the access to smart meter datasets. This is because real world datasets contain sensitive consumer information and, therefore, privacy is a key concern. Therefore, this paper focuses on simulating realistic data collected from the residential smart meter. As such, this paper presents a simulation of a home environment and the data produced. The validity of the data is justified through a visual comparison with a real-world smart meter dataset.**

Keywords- Smart Meters; Profiling; Simulation; Visualisation.

I. INTRODUCTION

Smart meters are a core component of the smart grid. Typically, they reduce financial losses, operational costs and enable energy suppliers to forecast customer demand [1]. As a result, smart meters are being implemented on a global scale. Many countries such as the UK, USA, Australia and Italy are already advanced in their smart meter implementation. Additionally, Sweden is one of the first countries in Europe to carry out metering reform and large-scale smart meter roll out. Before the reform, electricity consumption data for small customers is typically read on a yearly basis and billing is estimated based on the previous year's consumption, instead of actual meter readings. Consumer demand for timely and correct billing is the main driver for smart meter deployment [2].The smart meter system is equipped with a large number of sensors and actuators placed in all parts of the grid to monitor and control the operational characteristics and behaviour. Based on the data collected from these sensors, smart meter entities and electricity suppliers (utility companies) are able to make more insightful and better decisions. For example, they are able to manage and optimise the electricity flows, forecasting users' demand for electricity and balancing the grid more efficiently; and even detect when there is abnormal energy usage in homes. The potential research implications of access to this data is significant. For example, considerable research has been implemented into the use of smart meter data for remote healthcare monitoring [3]. Whereas, other research, has focused primarily on load balancing to support the efficiency of the grid and resource allocation [4]. The remainder of the paper is as follows. A background research on smart meter systems is put forward in Section 2. Subsequently, the research aims and objectives and the methodology is discussed in Section 3. Section 3 also presents a sample of the data collected from our smart meter case study. Section 4 discusses the methodology and techniques used for profiling users. The paper is concluded in Section 5. In particular, this paper focuses on the smart meter and investigates the novel approaches for consumer profiling and for the consumers to monitor energy usage in real time.

II. BACKGROUND RESEARCH

A smart meter is an electronic device that records the consumption of energy with high accuracy. However, smart meter is part of the much wider Advanced Metering Infrastructure (AMI).

A. The Advanced Metering Infrastructure (AMI)

An AMI is comprised of systems and networks that receive data from smart meters; and it facilitates the bidirectional communication between the consumer and the rest of the smart grid stake holders. It reduces the traditional need for energy usage readings to be collected manually [7]. Therefore, the smart meter is able to communicate with a gateway through a Home Area Network (HAN), Wide Area Network (WAN) or a NAN, which is outlined as shown in Figure 1.

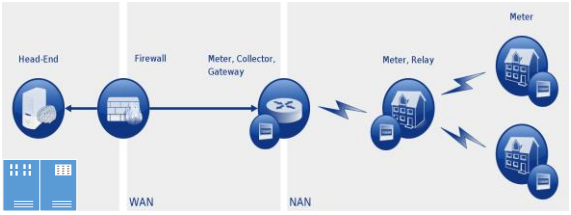


Figure 1. Advanced Metering Infrastructure

The HAN is housed inside the consumer premises and is made up of different devices, e.g., Meters, Thermostats, Electric storage devices, ZigBee transmitters. All of the

 ISBN: 978-1-61208-724-5

acquired data is sent to the Meter Data Management System (MDMS), which is responsible for storing, managing and analysing the data [8]. The MDMS sits within the data and communications layer of the AMI. This component is an advanced software platform, which deploys data analytics while facilitating various AMI applications, including:

- Managing metered consumption data.
- Outage management.
- Demand and response.
- Remote connect / disconnect.
- Smart meter events and billing [9].

This information can be shared with consumers, partners, market operators and regulators. The Wide Area Network (WAN) handles the communication between the utility companies and the HAN. The Head-End System (HES), also known as the meter control system, is located within a metering company network.

B. Machine Learning Techniques for Profiling

The first step to profile behaviour from the data produced by smart meters is to model and understand the normal patterns. Therefore, the field of machine learning provides methodologies that are ideally suited to the task of extracting knowledge from these data. A parametric approach often used is linear regression, which predicts a real valued output based on one or more input values. Prediction of a single output variable from a single input variable is called "univariate linear regression" whereas "multivariate linear regression" indicates multiple features. This module is used to define a linear regression method, which trains a model using a labelled dataset. The trained model can then be used to make predictions. Regression analysis is usually the best option and the fastest method to analyse the consumption data of buildings [10]. Among the statistical approaches, regression techniques deserve attention due to:

- Relative ease to implement.
- Interpretability of the results.
- The requirement of less computational power than other statistical approaches (genetic algorithms, neural networks, support vectors machine).
- Satisfactory prediction ability.
- Increased availability of data through smart metering.

Linear regression is a statistical analysis method used to model the relationship between two variables via fitting a linear equation to observed data. This relationship can be identified between the independent (explanatory) and dependent (response) variables. The response must be continuous, whereas the independent variables may be either continuous, binary or categorical. Linear regression can be expressed as:

$$y = \beta_0 + \beta_1 x \ , \qquad (1)$$

Where Y is the dependent variable, and X is the independent or explanatory variable and the betas (β_0 and β_1) are the coefficients that we need to identify to make predictions. This is demonstrated in Figure 2.

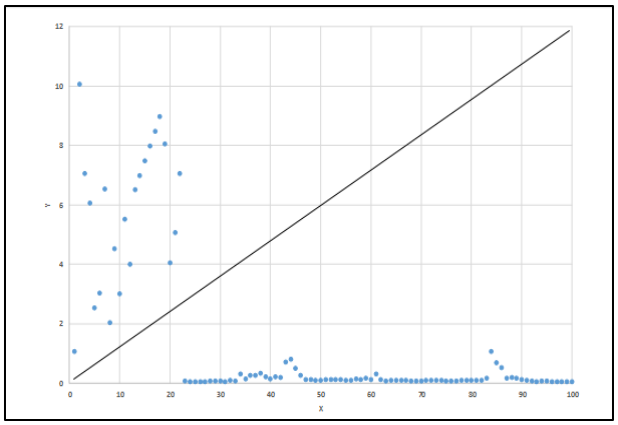


Figure 2. Simple Linear Regression

The line is modelled based on the linear equation shown in Figure 2 fitted to our data. Here, x and y are known variables from our data used to estimate the parameters β_0 and β_1 of the regression line. Typically, the parameters are estimated by minimising the least squares or the sum of squared errors. Therefore, for n observations, the linear regression model can be defined as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1,2,...,n. \qquad (2)$$

The random variable ε represents the error term in the model, a statistical term that represents random fluctuations and measurement errors among other factors out of our control [11].

Techniques, such as linear regression, can be used to predict future energy trends. For example, by using the smart meter data readings taken from one month, it would be possible to predict with a relative accuracy the expected consumption for the following month. In the following sub-section, a demonstration of the data that is collected from smart meters and how it can be analysed to model user behaviour is presented.

III. CASE STUDY: REAL-WORLD DATA

The electronic meters for electricity (smart meters) are undergoing an increasing deployment in private homes all over the world. As a consequence, an ever growing physical communication network, made up of millions of local meters, has been established, whose considerable advantages are so far in favour primarily, if not solely, of the energy distributors, since they are enabled at simplified, more efficient, and less costly transactions with the customers, e.g., for meter reading, billing, and energy supply administration. The detail and granularity of the data collected can be used in so many ways by utility companies, the future challenges faced is the issue of data storage and data management costs which prevent initiatives from becoming widely adopted. The amount of data produced by

 ISBN: 978-1-61208-724-5

two million smart meter customers reaches upwards of 22 gigabytes per day [15]. Naturally, it is a significant challenge to manage this data; which may include the selection, deployment, monitoring, and analysis processes.

C. Data Description

Any real-time information processing usually requires cloud computing [16]. Any delay may cause a serious consequence in the whole system, which has to be avoided as much as possible. The dataset used in this research is comprised of one-month's energy readings from 5 different users. Table 1 demonstrates a sample of smart meter data collected over a period of one month (January) for a single home occupant. The general supply of energy used on a daily basis (the energy consumed) is measured in kilo watts per hour (KWH) and can be described as what is used to bill the customer. Table 1 shows an example of energy reading of an individual household meter. Data is collected over a 30 min time interval period and the "energy delivered" in KWH. The customer key is the primary key used to identify the consumer while the End Date Time highlights the time and date of the acquired reading. Both the general supply and off peak supply are recorded based on the specified tariff.

Table I. SMART METER DATA SAMPLE

CUSTOMER_KEY	End Date time	General Supply KWH	Off Peak KWH	Year
8410148	1/1/13 0:29	0.081	0	2013
8410148	1/1/13 0:59	0.079	0	2013
8410148	1/1/13 1:29	0.082	0	2013
8410148	1/1/13 1:59	0.085	0	2013
8410148	1/1/13 2:29	0.073	0	2013
8410148	1/1/13 2:59	0.07	0	2013
8410148	1/1/13 3:29	0.07	0	2013

As above, a sample of the dataset is presented in the visualisation in Figure 3. In this case, five user's energy consumption over a five hour period is displayed.

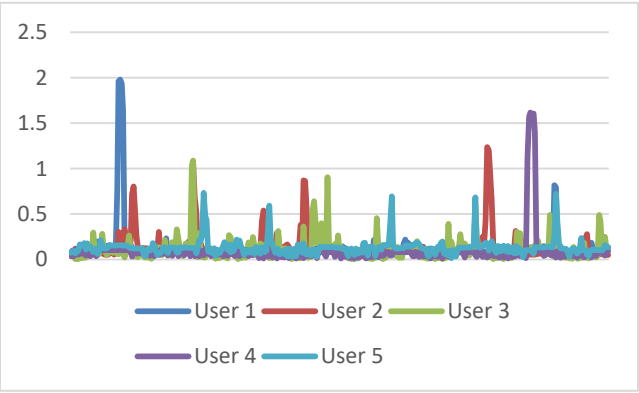


Figure 3. Visualisation of Data Sample

The y-axis displays the energy usage reading, while the x-axis displays the time in half-hours. As there are five hours, there are ten time stamps on the x-axis. The total dataset for the five users is plotted in a scatter matrix (Figure 4), which shows the correlation between their individual energy usage patterns.

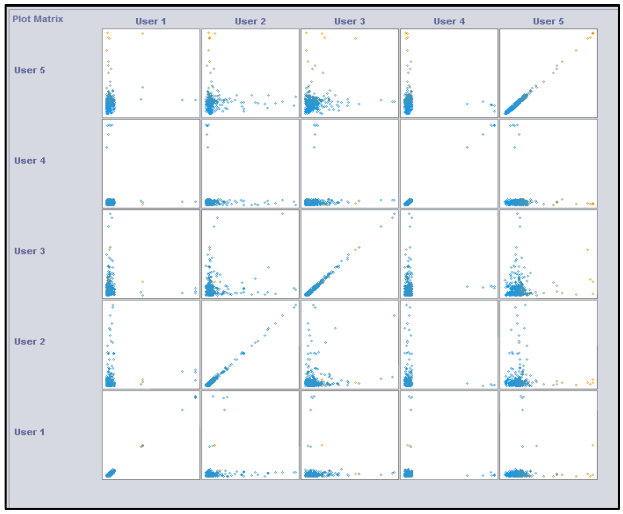


Figure 4. A generic pattern captured for 5 different users showing energy usage.

In order to visualise and analyse the total energy usage patterns over a much longer period, the smart meter data is loaded into a data model. The software used for this task is Microsoft Power BI [17]. Figure 5 presents an example of a much larger dataset that is comprised of seventy thousand household meter readings showing the energy usage and the behaviour trend over a period of 12 months. Here, the general distribution of energy readings highlights the energy consumptions levels for different households. This type of data visualisation could give suggestion to the number of occupants living in a given premise. Houses with increased energy usage are more likely to have an increased number of occupants or devices [12].

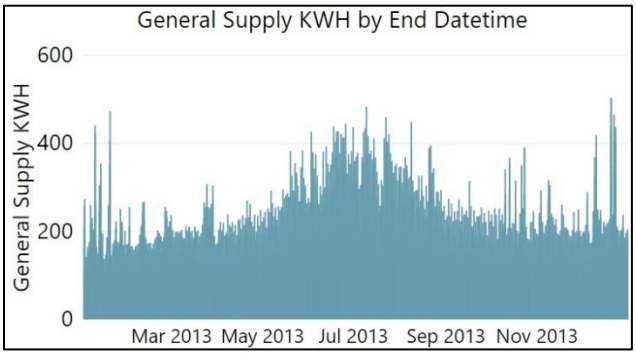


Figure 5. A generic pattern captured for 5 different users showing energy usage.

However, this information can be used by others, either maliciously or inadvertently to ascertain an insight into an individual's home life. For example, activities or occupancies of a home for specific periods can be determined. In a general, analysis of granular smart meter energy data could result in 1) invasion of privacy; 2)

 ISBN: 978-1-61208-724-5

unwanted publicity and embarrassment (e.g., public disclosure of private facts of people's daily living lifestyles).

The security policies governing the reliability of the smart grid depend on appropriate connectivity protocols and the national institute of standards and technology being the reference model proposed [16]. Recognizing the urgent need for standards to support Smart Grid interoperability and security, NIST developed a three-phase plan. 1) Identify an initial set of standards that would promote the rapid development of the Smart Grid, 2) establish a robust framework for the sustaining development of many additional standards, and 3) establish the a framework for the conformity testing infrastructure that is needed.

III. CASE STUDY: SIMULATION DATA

As previously discussed, access to smart meter data is limited. In this section, a case study on the simulation of smart meter data is presented.

D. Simulationd Design.

Figure 6 displays a model of a simulated home. The home was designed to resemble a moderate family home, with a standard set of appliances.

Figure 6. Simulated Home

The number and type of appliances present in the home are customised in Figure 7.

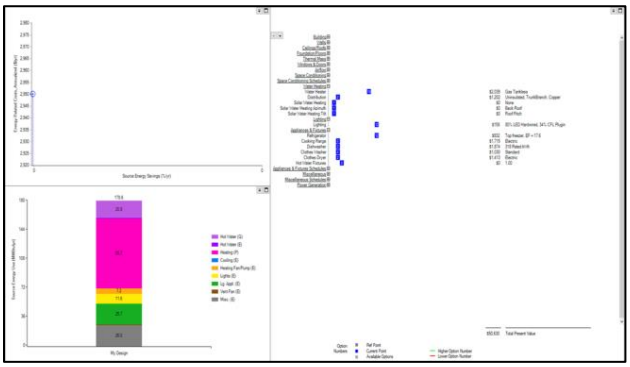


Figure 7. Home Customisation

The properties of the simulated home are detailed in Table II. Other properties, such as insulation, construction materials, weather sheathing and exterior finish are included in the simulation but omitted from Table II.

Table II. SIMULATED SMART HOME PROPERTIES

Input	Value
Project Type	Standard
Application Type	New Construction
Building Type	Multi-Family
Analysis Mode	Design
Reference Building	My Design
Sim Engine	EnergyPlus
Building: Finished Floor Area	4500
Building: Bedrooms	18
Building: Bathrooms	8

By running the simulation, data is constructed for a given simulation period. A visualisation of the energy readings from the simulated home is presented in Figure 8. The months are displayed on the x-axis and the total energy usage values for a given day are displayed on the y-axis.

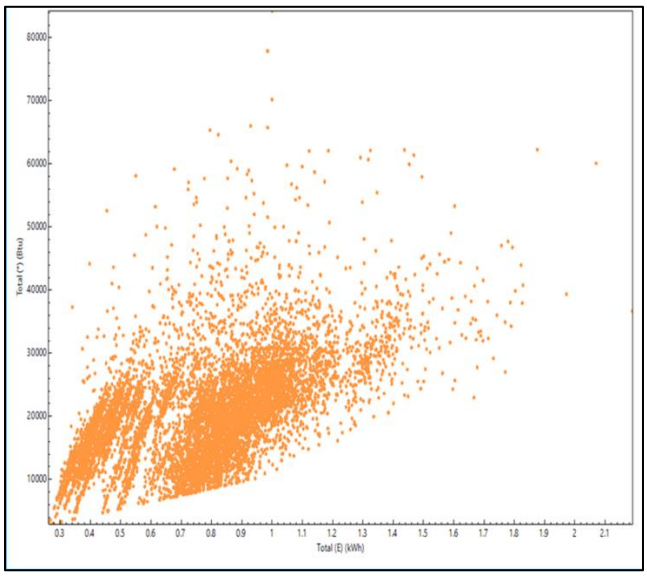


Figure 8. Simulation Data Visualisation Sample

In the following section, a discussion is presented on the use of real-world smart meter data in comparison to the simulated data generated by the artificial home.

IV. DISCUSSION

Figure 9 displays real-world energy readings from a single home. The home is picked at random from the data set, but it meets the following criteria:

- Data shows that there is an occupant in the premises
- They live in a standard house

Each morning demonstrates a sudden change in user behaviour. The energy usage, in KWH, is shown in the y-axis while the time the reading was taken is shown in the x-axis. The graphs indicate the time when the consumer becomes active in the morning. These activity start times vary depending on each user and readings are captured for

the whole 24 hour period. These types of behaviour can be attributed to the consumer's morning, afternoon and evening activities and is a key indicator for understanding and identifying alterations in routine.

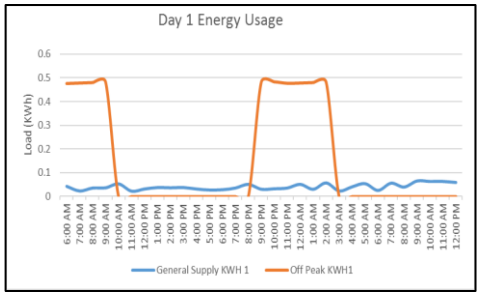


Figure 9a. Measured power load over a 24 hour period on day 1 showing occupant using energy mostly in the moning and late evenings.

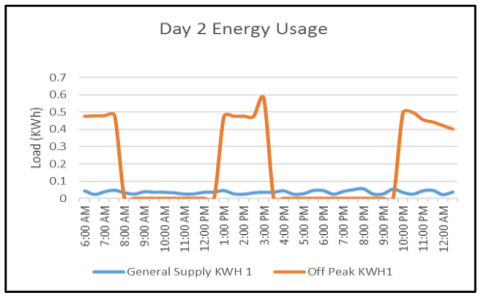


Figure 9b. Measured power load over a 24 hour period on day 2 showing occupant using energy in the morning, mid-afternoon and late evenings.

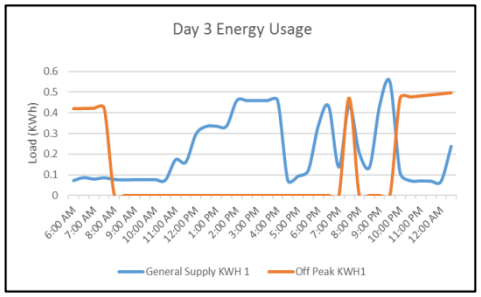


Figure 9c. Measured power load over a 24 hour period on day 3 showing occupant using energy throughout the day.

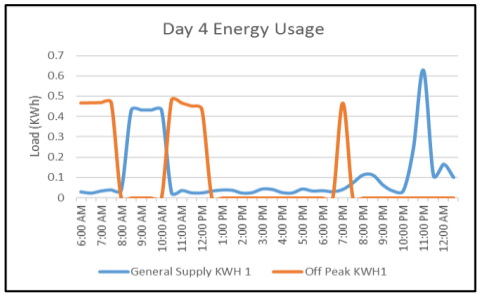


Figure 9d. Measured power load over a 24 hour period on day 4 showing occupant using energy in the morning and late evenings.

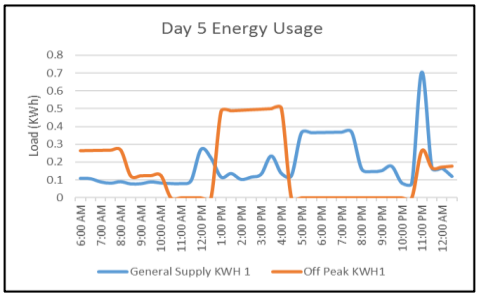


Figure 9e. Measured power load over a 24 hour period on day 3 showing occupant using energy throughout the day.

By the efficient analysis of the energy use data, different energy consumption patterns of different household are demonstrated below and corresponding energy use behavioural characteristics are identified. Along the x-axis is the hours of the day over a 24 hour period with the readings taken for each day for 5 consecutive days for each house. The y-axis values refer to the energy usage in kilowatts (KWh). Using this data, we will build up a pattern of expected behaviours and identify a trend. Figure 10 displays the simulated data home patterns over a 12-month period. The green line is the total energy usage, whereas the orange and the blue line depicts the individual energy readings for unit 1 and unit 2 in the simulation study. Similarly to the real-world data, on visual inspection, patterns in the energy consumption are apparent. For example, similar spikes in the energy use on a weekly basis, lower energy use in the summer months and higher energy use during winter.

Figure 10. Simulated Home Graph Visualisation

V. CONCLUSION

In this paper, several visualisations of electricity consumption patterns both from a real-world dataset and from a simulated smart home have been conducted. Smart meters produce considerable volumes of data, presenting the opportunities for utilities to enhance customer service, lower costs and improve energy efficiency; and for customers to

 ISBN: 978-1-61208-724-5

reduce their bills and save energy. The availability of such information results in more informed consumers who can better self-manage their electricity usage, choose low energy saving appliances and thus contribute to the reduction of greenhouse gas emissions. Therefore, we conclude that the analytic techniques and methodology proposed in this paper can be practical and useful to benefit the consumers and utility companies as well as the governments. Our ambition is that our work will lead to a realistic simulated smart meter dataset that other researchers can use for investigations into home profiling.

This research project is funded in part by the EPRSC - EP/R020922/1. Owing to the ethical sensitive nature of this research, the data underlying this publication cannot be made openly available.

REFERENCES

[1] Skopik, F. and Smith, P.D. eds., 2015.Smart grid security: Innovative solutions for a modernized smart grid . Syngress.

[2] J.Zheng, D.W. Gao, and L.Lin, " Smart meters in smart grid: An overview. In Green Technologies," Conference IEEE, pp. 57-64, April, 2013.

[3] E.D. Knapp and J.T. Longil, " Industrial Network Security: Securing Critical Infrastructure Networks and Smart Grid, SCADA and other Industrial Control Systems," Chapter, 12, 2015.

[4] Espinoza, M., Joye, C., Belmans, R, and De Moor, B., 2005. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. IEEE Transactions on Power Systems, 20(3), pp.1622-1630.

[5] Bedingfield, S., Alahakoon, D., Genegedera, H, and Chilamkurti, N., Multi-granular electricity consumer load profiling for smart homes using a scalable big data algorithm. Sustainable cities and society, 40, pp.611-624,2018.

[6] Virone G, et al. Behavioral patterns of older adults in assisted living. IEEE Trans Inf Technol Biomed 12(3), pp. 387–398,2008.

[7] M.Popa, Data Collecting from Smart Meters in an Advanced Metering Infrastructure, Proceedings of 15th International Conference on Intelligent Engineering Systems, pp. 137-142,2011.

[8] D. Niyato and P.Wang, "Cooperative transmission for meter data collection in smart grid," IEEE Communications Magazine, vol. 40, pp. 137-142, 2012.

[9] Y.T. Hoi, et al. "The Generic Design of a High-Traffic Advanced Metering Infrastructure Using ZigBee," vol. 10, pp. 836-844, 2014.

[10] Malekian, R et al. A novel smart ECO model for energy consumption optimization. Elektronika ir Elektrotechnika, 21(6), pp.75-80, 2015.

[11] A. C. Rencher and G. B. Schaalje, Linear Models in Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2007.

[12] Bartusch C, Odlare M, Wallin F and Wester L. Exploring variance in residential electricity consumption: Household features and building properties. Applied Energy 2012;92:637–643

[13] World Health organisation on Ageing 2015, accessed 19 march 2018,<http://www.dw.com/en/aging-health-who-call-to-action-as-world-population-over-60-set-to-double-by-2050/a-18751636#>

[14] Chalmers, C., Hurst, W., Mackay, M., & Fergus, P. Smart Meter Profiling For Health Applications. In the Proceedings of the International Joint Conference on Neural Networks, July 2015

[15] Choo, K.K.R., 2011. The cyber threat landscape: Challenges and future research directions. Computers & Security, 30(8), pp.719-731.

[16] Advanced metering Security Threat Model (DRAFT); R Robinson, J McDonald,B Singletary, D Highfill,N Greenfield,M Gilmore

[17] T. Lachev, and E. Price, 2018. Applied Microsoft Power BI: Bring your data to life!. Prologika Press.

[18] Coalton Bennet and Darren Highfill. Networking AMI Smart Meters, IEEE Energy 2030, November 2008.

[19] Bera, S., Misra, S. and Rodrigues, J.J., Cloud computing applications for smart grid: A survey. IEEE Transactions on Parallel and Distributed Systems, 26(5), pp.1477-1494, 2012.

 ISBN: 978-1-61208-724-5

Augmented Reality for Enhancing Life Science Education

John Barrow, Conor Forker
Institute of Medical Sciences,
University of Aberdeen,
Aberdeen, AB25 2ZD, UK
e-mail: {j.barrow,
c.forker.15}@abdn.ac.uk

Andrew Sands, Darryl O'Hare
Imagin3D, Office 4,
CTH, Sci Tech Daresbury,
WA4 4FS, UK
e-mail: {devteam,
darryl}@imagin3d.co.uk

William Hurst
Department of Computer Science,
Liverpool John Moores University,
Liverpool,
L3 3AF, UK
e-mail: W.Hurst@ljmu.ac.uk

***Abstract*— Augmented Reality (AR) has the opportunity to be a disruptive technology in the delivery of educational materials at all levels, from public outreach activities to expert level teaching at undergraduate and postgraduate levels. The attractiveness of AR as a teaching tool is its ability to deliver a blended learning experience created from the mixing of the virtual and real environments or materials in the classroom. This allows students to learn in a variety of ways to mix didactic, experiential and kinaesthetic learning. We have developed, and are in the process of developing, AR applications that aim to transform the learning space into one that is highly interactive, so this paper will discuss the potential impact of such teaching interventions on higher education.**

Keywords- Augmented reality; education; visualisation.

I. INTRODUCTION

One contemporary paradigm in higher education is the tailoring of educational resources towards a so-called "digitally native" audience who, for the most part, have grown up surrounded by digital technology. The students entering higher education today are demanding a high standard of education that incorporates the digital world in which they live. There is sometimes a tendency to aspire to incorporate digital technologies into everything that is carried out in a higher education institution, but it is not digital natives rather digitally aware students that we should be developing [1]. Furthermore, students require an ability to be prepared for the future, especially in a fast-paced, ever changing world were digital literacy, and skills associated with it are seen as essential [2]. Further to this, the life sciences are one such area where this requirement for a digital skill set is a necessity as much of the current research relies on digitalisation of data. For example, this could be in the form of genomic data or population statistics of various patient sub-groups [3]. This paper will discuss the development and testing of an augmented reality (AR) app for teaching metabolism, especially glucose metabolism and insulin signalling. Section II discusses the background to the project; Section III describes the implementation of the AR app; Section IV shows results from testing; and Section V discusses the conclusions of this work.

A. VR/AR in Education

As a consequence of the required skills set a life science undergraduate needs and the inherent requirement to develop digitally-driven approaches to delivering a high quality education, the use of augmented reality and virtual reality has become an emerging theme. A simple literature search in the PubMed (The National Center for Biotechnology Information) database for the terms virtual reality and augmented reality in education shows the rapid growth of both areas (Figure 1).

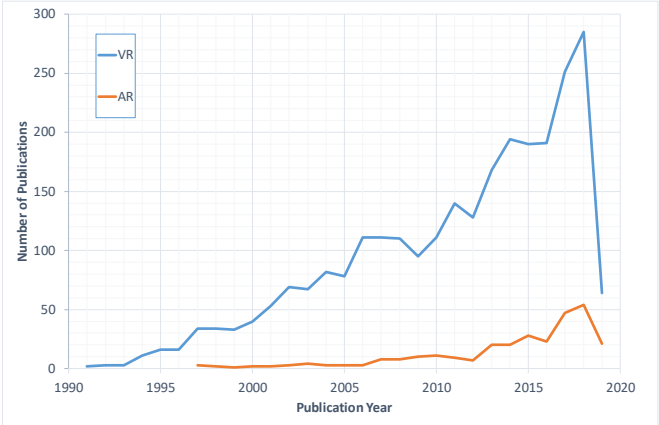


Figure 1. PubMed results by year for virtual and augmented reality publications.

Search terms of "virtual reality AND education" and "augmented reality AND education" were used to identify the total number of publications in each year from the first year that a publication appeared in the PubMed database. VR (blue line), virtual reality; AR (red line), augmented reality. Note – there was one publication for augmented reality dated 1989, which is excluded from the data shown.

B. Aims

This paper will describe the results of a consultation on the use of digital visualisation technologies in the teaching of life science subjects and the creation and testing of an AR application to aid in the teaching of metabolism, specifically linked to glucose and insulin signalling.

C. Study Methods

The present study was conducted in the School of Medicine, Medical Sciences and Nutrition at the University of Aberdeen. Ninety participants were self-selected by completing a survey as part of a second year undergraduate biochemistry course that forms part of the core curriculum for most life science degree programmes. Of the ninety questionnaire participants, eight were randomly selected to

 ISBN: 978-1-61208-724-5

take part in a focus group and seven volunteered to test an AR application linked to the teaching of metabolism.

The questionnaire aimed to identify student preferences in the teaching structure-function relationships in the life science subject areas. The focus group aimed to question participants on their preference for digital technology and how it can integrate with their teaching. The testing of the AR application aimed to identify their views of AR in education having been exposed for the first time to this kind of technology in the classroom.

II. BACKGROUND

Virtual and augmented realities (VR and AR respectively) allow immersive and visual experiences for the user. In VR, this will be via the use of a head-mounted display that is either a standalone device or a device tethered to a computer that drives the visualisation hardware. VR can also be delivered using smartphone technology and a headset viewer that allows the user to display a VR image on the smartphone screen. When using tethered or standalone VR headsets, there is a requirement for high-end computer hardware to run the software, but for smartphone VR the requirement for accelerated hardware is less so and the headset smartphone holders are also relatively inexpensive or can even be created by the user (e.g., Google Cardboard). In AR, the experience is different from VR as it allows the user to overlay digital content in the real-world environment and interact with that content. This offers some distinct advantages over VR as it allows the user experience to be shared amongst groups rather than being a single-user experience, and it also provides an opportunity for users to mix learning styles when the AR is combined with more traditional forms of teaching materials such as texts or lecture slides. This mixed approach could provide a powerful tool that satisfies many learner styles, allows collaborative learning, and provides increased scope to bring subjects to life in a way that has not been possible before.

A. VR/AR in Educational Context

In an educational context, VR offers some distinct advantages over standard teaching practices in that it can allow students to simulate scenarios, such as surgical training of medical students [4], or allow students to understand abstract concepts that are not visible like protein structure and function [5]. AR also offers advantages as it can allow the delivery of mixed methods teaching were students have traditional learning from written materials coupled with visualisations of the processes involved [6], or provide interactions between real-world objects and the digital visualisations [7].

In many education settings where budgets are often constrained or limited, the use of smartphone technology to deliver VR or AR experiences becomes a more attractive proposition. This also ties in with more and more students having their own devices that are capable of delivering high quality digital experiences as smartphone technology becomes increasingly more powerful. Moreover, students are required in many higher education institutions to use their own devices to record attendance, interact with classes through online voting systems and other institutional resources such as timetables and virtual learning environments. This means there is a real opportunity to develop classroom activities that make use of the ‘bring your own device’ (BYOD) model. It does also present challenges as BYOD means there will be variable technologies in circulation in any given student cohort. Operating systems, hardware specifications and graphics capabilities will vary widely, so careful consideration may need to be given if students were to use their own devices for VR or AR applications.

B. A TPACK Model of AR Education

There is growing evidence that digital visualisations help students understand abstract concepts, which can be viewed through the lens of the technological pedagogical content knowledge (TPACK) framework [8]. This framework (see Figure 2) highlights the importance of the interplay between technology, discipline knowledge and teaching practice to deliver a modern and relevant programme of study, especially in the life sciences were digital technology is crucial in virtually all research areas. In higher education institutions, the expectation is that subject-specific knowledge and expertise is provided by academic staff alongside effective teaching practices. The one area where there is perhaps some variation is in the ability of the academic to embed technology into the classroom and provide students with a modern curriculum that integrates technology into their learning. AR offers just such an opportunity without changing traditional curricula significantly, as it allows the blending of instructive teaching with digital visualisations.

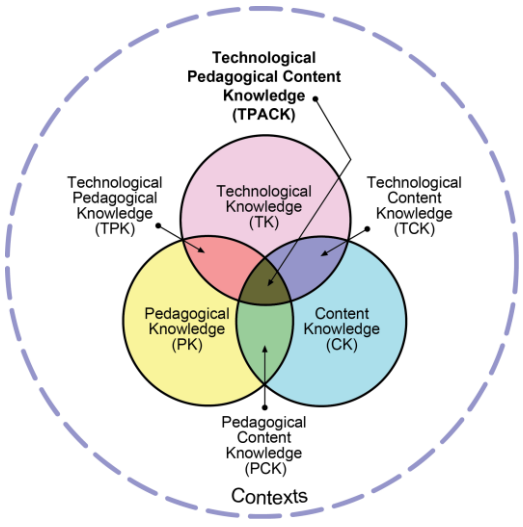


Figure 2. Technological Pedagogical Content Knowledge Framework [9].

C. 3D Literacy and AR in Life Sciences Teaching

The ability to understand three-dimensional structures is crucial in virtually all life science disciplines, from a gross anatomical level to the sub-cellular and molecular levels. One could describe this skill set as a ‘3D literacy’ [10] that

plays a pivotal role in a full appreciation of the concepts that underpin biological processes and functionality. Traditional teaching methods in the life sciences will often be didactic in practice, with possible inclusion of models to highlight structural aspects of how processes work. In anatomy and physiology disciplines, this will often take the form of physical cadaveric specimens, but where this is not possible, teaching will rely on physical models to highlight structure-function relationships in the human body. In the molecular life science disciplines, this concept of structure and function is just as valid, and as such, various physical chemical structure models will often be used to explain the processes involved at a sub-cellular level. The use of models coupled with some functionality can teach molecular concepts well, with two examples being the concept of polarity of water molecules or structure-function relationships in proteins [11][12].

More recently, it has been possible to create models using 3D printing. The advantages of this technology are that the teacher can create models of virtually any kind of structure to aid in their teaching, but it does have drawbacks as it can be time consuming and technically difficult to create certain structures (e.g., very thin or overhanging structures) due to constraints in the 3D printing process. There are a huge variety of examples that utilise 3D printing and as with traditional models used in teaching they have been used in all areas of the life sciences were structure-function relationships are important for understanding [13][14]. All of the above teaching examples have several drawbacks, be that expense, time to create models, lack of functionality or movement in the models or that these models do not allow for mass participation in larger class sizes due to the limited number of models available.

AR has the possibility of addressing some of the issues with more traditional forms of structure-function teaching. If we follow the design principles set out by Dunleavy [15] then AR has the possibility to: 1. Enable and then challenge; 2. Drive by gamified story; and 3. See the unseen. All three of these principles can be relatively easily achieved using AR. There are many areas were AR has been implemented with varying degrees of success, but it holds most promise in those subjects were an appreciation of three dimensional space and structure is crucial for a full understanding of the subject. It has therefore been most successfully employed in subjects such as anatomy were it is crucial that students understand the spatial arrangements of tissues and organs, and were cadaveric material is not always available [16]. AR has also been used in more abstract subjects such as structural biology were students will understand molecular and sub-cellular processes much better if they can appreciate how the structure of molecules often dictate their function [17].

III. IMPLEMENTATION

The implementation of the AR application presented in this research was a three-stage process, involving 1) Modelling the different 3D assets; 2) Texturing and setting up the game-engine mechanics and 3) AR implementation.

A. *Modelling*

Nine different 3D models were created for the application. These are presented in Table I.

TABLE I. SCENE MODELS AND THEIR ROLES.

Model	**Role**
Character	Core element, in the application produced in high detail (73k polygons), as displayed in Figure 3.
Chocolate bar	Minor model, present to demonstrate eating.
Intestines and Gut	Core component for the application. Created in detail to show overall gut/digestive system construction.
Veins/Arteries	Core component for demonstrating the biological processes taking place, but with low poly (300 polygons).
Blood Vessels	Low poly (50 polygons each), animated objects in the scene to demonstrate blood flow and scale.
Cells	The cells are prominent in the application. They are very low poly, with the details being added through the texturing.
Muscle Cells	The muscle cells, again are low poly, and are the recipients of the glucose and insulin molecules.
Glucose & Insulin Molecules	Low poly, small objects in the scene but core to demonstrating the biology.

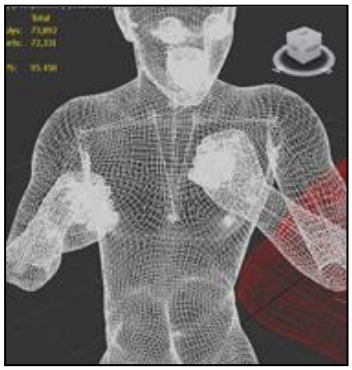


Figure 3. 3D Character Mesh.

Following the modelling process, textures were added within the game-engine environment directly. By texturing within the game engine, it allowed for the inclusion of shaders on the models; resulting in a higher quality of texturing and realism.

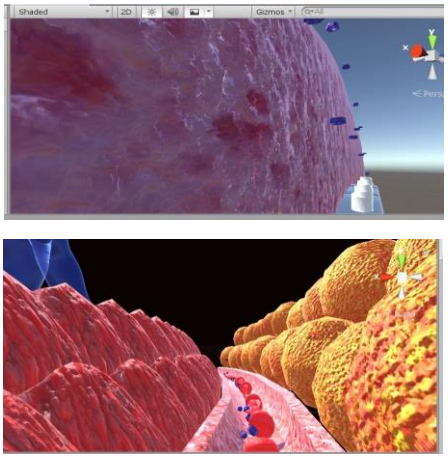
B. *Texturing and Game Engine Mechanics*

The texturing process was crucial for adding detail to the 3D modelling process and creating a relatability for the

students when operating the AR application. An effective texturing process also ensured that the models maintained a low poly count, as the detail was generated by the textures rather than the 3D models themselves. A low poly count is necessary to ensure that the AR application runs smoothly on hand held devices (i.e., tablets and smart phones), which have a limited processing and graphics capability in comparison with a PC/Laptop.

As Figure 4 displays, the character is semi-transparent, in order to allow the intestines and digestive system to be visible. Food is also animated travelling down the throat into the stomach, as the character bites the chocolate bar. An organic texturing was applied to the various cells and tissues as shown in Figure 5.

Figure 4. Character Texturing.



(a)

(b)

Figure 5. Organic Texturing with the Game Engine. (a) Cells, (b) Blood Vessel with Blood Cells (Centre), Muscle Cells Right) and Pancreas Cells (Left).

High-resolution texture images were used. With the model being AR, it is possible to zoom in and view the assets at close inspection. Bump map and height maps were also applied to the models so that they did not appear flat on projection. The other technical challenge involved the UVW mapping as many of the models are spherical in appearance. The final scene composition is displayed in Figure 6. The composition consists of a character, with food passing down the throat as an animation; a close up view of the digestive system enclosed in a box next to the character; and a close up view of the biological process taking place within the blood stream.

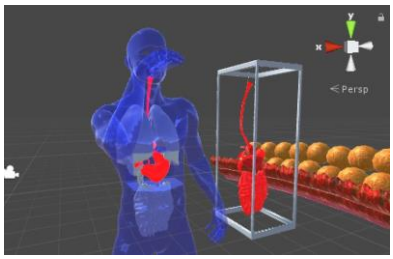
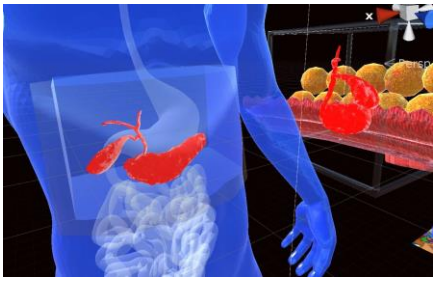


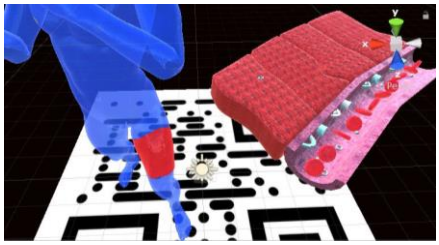
Figure 6. Final Composition.

C. AR Setup

The AR was set up using standard black and white QR code markers to project the models on (as shown in Figure 7b). However, under testing the models often ended up projecting with glitches or delays. The model project on QR is displayed in Figure 7.



(a)



(b)

Figure 7. Model Project in AR. (a) Up-close texturing (b) model projected on QR code in Game Engine.

Instead, advanced QR markers were used to improve the stability of the projection, as displayed in Figures 7a and 8. The app was then deployed on both Android and Windows tablets, as displayed in Figure 8.

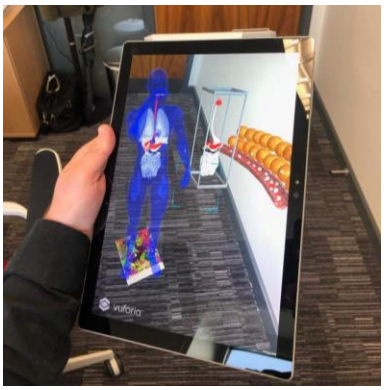


Figure 8. App Functioning on Windows Tablet.

 ISBN: 978-1-61208-724-5

Following the deployment of the app, it was then field-tested in a classroom environment at the University of Aberdeen.

IV. RESULTS

Participants in the test completed an online questionnaire asking them to think about their views on gamification and the use of digital technology in their education as part of a second year undergraduate biochemistry course. 90 students (out of a class of 150) participated in the questionnaire and as a demographic, the vast majority (95%) were 19-24 years of age (i.e., an age group considered to be 'digital natives').

A. Initial Questionnaire Findings

When asked if they enjoy lectures, on a 5-point Likert scale, 40% enjoyed roughly half of their lectures and 49% enjoyed almost all lectures. 76% of participants agreed or strongly agreed that lecturers were good at explaining abstract concepts, suggesting that most students can understand and follow the lecture material being delivered, however only 22% strongly agreed so there is a large proportion who could have their learning enhanced at least partially. Similarly, a large proportion of the 90 participants agreed that lecturers made classes exciting and engaging with 63% agreeing or strongly agreeing to this statement, with 27% neutral showing that there is a decrease in student satisfaction compared to lecturers explaining abstract concepts well. 51% of participants agreed or strongly agreed that they make detailed lecture notes, with 25% disagreeing with this statement and 23% neutral, showing that approximately half of the students do not make detailed notes of the lecture material. 38% of participants were neutral and 19% agreed or strongly agreed that PowerPoint slides were boring, suggesting that there could be room to improve teaching provision away from standard teaching practices.

In answer to the statement that they prefer a hands-on approach to learning than being in lectures, 45% agreed or strongly agreed, 25% disagreed and 30% were neutral, showing the majority would possibly benefit from additional teaching methods. When asked the question of how lectures could be made more enjoyable, the free text responses were analysed, and the top 5 words identified in the responses were "interactive", "engaging", "examples", "videos" and "interaction". This is promising for the use of AR in the classroom as it shows that for engagement initiatives to be a success they should incorporate interactivity.

When asked if they consider themselves a "gamer", 42% and 19% strongly disagreed or disagreed respectively, and only 22% agreed or strongly agreed, which shows that any technology implemented into their learning should be adapted for the novice gamer audience and be easy to understand and interact with. Overall, 98% of participants own a smartphone (either Android or iOS), 93% have a laptop and 40% have a tablet, which means that only a very small percentage of the class do not have smartphones, so would minimise the cost if students were to use their own devices.

When asked if the use of games in class will help student learning, the majority (52%) agreed or strongly agreed that it would, and only a small fraction (15%) disagreed or strongly disagreed. This coupled with 92% agreeing or strongly agreeing that being able to "see" abstract concepts would help their learning, would strongly suggest that a highly visual and interactive AR approach would provide students with an excellent learning resource.

When asked the open-ended question of what they would like included in an educational app, the most common remarks were "keep it engaging", "abstract concepts", "3D structures" and "providing an interactive experience with more difficult theories". The results are plotted in a word cloud in Figure 9, where the larger the word, the more often is appeared in the feedback comments. Clear requirements for the application to be 'engaging' and 'interactive' are prominent in the feedback.

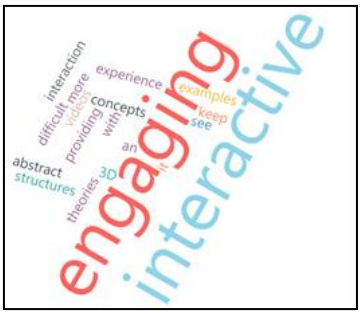


Figure 9. Pre-Survey Feedback Comments.

B. Focus Group

Based on the questionnaire feedback, we next organised a focus group made up of eight randomly chosen participants who indicated they wanted to take part in further discussions. The focus group was intended to gather views and opinions on the role that visualisation apps could play in teaching on the second year undergraduate biochemistry course. The following questions, which emerged from the questionnaire data, were used to encourage discussion: 1. Do you have difficulty visualising and learning abstract concepts? 2. How do you currently learn these difficult topics? 3. What would be the best use of gamification to teach lecture topics?

In summary, AR was favoured over VR and the participants would prefer any AR content to be used in a tutorial rather than lectures so that information from lectures could be consolidated rather than being taught for the first time. Several participants would also like to be able to pick up the AR app if required rather than it being a compulsory session that they must attend, which would allow them to use the additional visualisations if they required them to aid their learning. Accurate use of 3D models would be preferential over simplified models or models that do not resemble the actual structures being visualised.

C. AR App Usage Feedback

Following development of the AR app, it was tested with seven randomly selected participants (different individuals from those who participated in the focus group) to understand if students would prefer to use AR in their

education. The AR app was developed to run alongside written materials that explained how three processes work in a specific metabolic process. 43% of participants strongly agreed and 57% agreed that the AR app motivated them to understand the processes being visualised, with nobody remaining neutral or disagreeing. 58% of participants strongly agreed or agreed that the AR app enhanced their learning and 42% where neutral. 57% strongly agreed and 43% agreed that the class was enjoyable and fun, and they rated the class overall with the same scores. When asked if they would like to see more AR in the classroom, 86% responded "absolutely" and 14% said they were not sure as it would depend on the class.

When asked which subject areas AR technology would be most useful for, the following subject areas: enzyme activity and protein structure, biochemistry, molecular biology, genetics, bioenergetics, immunology, physiology, neuroscience, enzyme activity and enzyme structure. Additionally, participants stated, "I think it's a great idea, and would definitely help information we're taught traditionally just from being able to visualise processes"; "it would be very useful to see how enzyme activity and protein conformation works"; "it's really cool and [would] be a great asset in classes and for learning concepts which are more abstract to understand.

Figure 10. Post-Survey Feedback Comments.

It also helps people to gaze into the future and feel like they are part of…learning in a new direction". Again, the feedback was plotted in a word cloud as displayed in Figure 10. Key positive words, such as 'great', 'future', 'helps' are prominent in the feedback.

V. Conclusion and Future Work

In conclusion, the application proved successful with end-users and received primarily positive feedback. The deployment of the application on Android and Windows also meant the up-take of the application had a wide reach. In the future, we will investigate the deployment of the application on an iOS device.

There were some challenges with the initial model implementation as it did not show the correct processes, but this was overcome with clear communication between the development team and the scientists on the project.

There are also alternative approaches to delivering this kind of visualisation in class using VR, which will also be explored in the future, but as already described, the use of VR would change the class dynamics as it is a highly personalised experience even though it may be more immersive.

References

[1] M. Henderson, N. Selwyn, G. Finger, and R. Aston, "Students' everyday engagement with digital technology in university: exploring patterns of use and 'usefulness,'" J. High. Educ. Policy Manag., vol. 37, no. 3, pp. 308–319, May 2015.

[2] L. C. Larson and T. N. Miller, "21st Century Skills: Prepare Students for the Future," Kappa Delta Pi Rec., vol. 47, no. 3, pp. 121–123, Apr. 2011.

[3] V. Marx, "The big challenges of big data," Nature, vol. 498, no. 7453, pp. 255–260, Jun. 2013.

[4] N. E. Seymour et al., "Virtual Reality Training Improves Operating Room Performance: Results of a Randomized, Double-Blinded Study," Ann. Surg., vol. 236, no. 4, p. 458, 2002.

[5] M.-J. Garcia-Bonete, M. Jensen, and G. Katona, "A practical guide to developing virtual and augmented reality exercises for teaching structural biology," Biochem. Mol. Biol. Educ., vol. 47, no. 1, pp. 16–24, Jan. 2019.

[6] A. Sugiura, T. Kitama, M. Toyoura, and X. Mao, "The Use of Augmented Reality Technology in Medical Specimen Museum Tours," Anat. Sci. Educ., Nov. 2018.

[7] D. Behmke et al., "Augmented Reality Chemistry: Transforming 2-D Molecular Representations into Interactive 3-D Structures," Proc. Interdiscip. STEM Teach. Learn. Conf., vol. 2, no. 1, Jan. 2018.

[8] M. J. Koehler, P. Mishra, and W. Cain, "What is Technological Pedagogical Content Knowledge (TPACK)?," J. Educ., vol. 193, no. 3, pp. 13–19, Oct. 2013.

[9] Reproduced by permission of the publisher, © 2012 by tpack.org

[10] J. R. Roberts, E. Hagedorn, P. Dillenburg, M. Patrick, and T. Herman, "Physical models enhance molecular three-dimensional literacy in an introductory biochemistry course*," Biochem. Mol. Biol. Educ., vol. 33, no. 2, pp. 105–110, Nov. 2006.

[11] T. Herman et al., "Tactile teaching: Exploring protein structure/function using physical models," Biochem. Mol. Biol. Educ., vol. 34, no. 4, pp. 247–254, Jul. 2006.

[12] J. L. H. Johnson, S. H. Yalkowsky, and Ed Vitz, "A three-dimensional model for water," J. Chem. Educ., vol. 79, no. 9, p. 1088, Sep. 2002.

[13] S. C. Meyer, "3D Printing of Protein Models in an Undergraduate Laboratory: Leucine Zippers," J. Chem. Educ., vol. 92, no. 12, pp. 2120–2125, Dec. 2015.

[14] M. Vaccarezza and V. Papa, "3D printing: a valuable resource in human anatomy education," Anat. Sci. Int., vol. 90, no. 1, pp. 64–65, Jan. 2015.

[15] M. Dunleavy, "Design Principles for Augmented Reality Learning," TechTrends, vol. 58, no. 1, pp. 28–34, Jan. 2014.

[16] R. Barmaki et al., "Enhancement of Anatomical Education Using Augmented Reality: An Empirical Study of Body Painting," Anat. Sci. Educ., Jan. 2019.

[17] P. Safadel and D. White, "Facilitating Molecular Biology Teaching by Using Augmented Reality (AR) and Protein Data Bank (PDB)," TechTrends, pp. 1–6, Oct. 2018.

Industrial Augmented Reality (IAR) as an Approach for Device Identification within a Manufacturing Plant for Property Alteration Purpose

Tshepo Godfrey Kukuni
Department of Electrical, Electronic, and Computer Engineering
Central University of Technology, Free State
Bloemfontein, South Africa
email: tgkukuni@gmail.com

Ben Kotze
Department of Electrical, Electronic, and Computer Engineering
Central University of Technology, Free State
Bloemfontein, South Africa
e-mail: bkotze@cut.ac.za

***Abstract*—In this article, a possible solution to the identification and detection of components in a process is investigated within a controlled network environment. This seems necessary as the component is not always identifiable with small or no previous knowledge to the design and implementing phase. Augmented Reality utilising identifiers such as machine vision is investigated rather than the current Quick Response codes or Radio Frequency Identification. If identified at its position, the data and device details are then available to the user for viewing or editing purposes. This project is still in its startup phase thus real data will be addressed and discuss in a follow-up article.**

Keywords- industrial augmented reality; image processing; virtual reality; mixed reality; augmented reality; RFID; QR codes; manufacturing plants; SCADA.

I. INTRODUCTION

In recent years, the demand for software resources has increased drastically due to the rise in smartphones hardware [1]. As a result, researchers have predicted that Mixed Reality (MR) will have the potential to play a very big role in our future daily lives, such as in education, medical, production, etc. It is with such reasons that Virtual Reality (VR), which is a component of the MR technology, can be transmuted into Augmented Reality (AR) by adding real elements such as live video feeds to the virtual world.

Acquiring information about the surrounding objects effectively is a crucial factor for many people including people with disabilities such as the blind. The introduction of the AR application, which is denoted as a powerful user interface technology that augments the user`s environment with computer-generated entities can accomplish this task by making use of AR technologies [2].

The information about reality becomes more interactive between the user`s perceptions of the real and virtual world in an AR environment. The reality happens through the application of the real-time object detection and recognition algorithms that will enable recognition of the surrounding objects in the real environment in order to align the computer-generated images with these objects in an AR view [3].

In comparison with VR, users can see virtual and real-world objects concurrently in an AR system. Since both VR and AR are virtual objects related phenomena's, the concept of enhancing the illusions that the virtual objects are present in a real scene has led to more research attention focusing on the occlusion problem.

The problem with occlusion occurs when real objects are in front of the virtual objects in a scene [4]. However, the information can be inserted in a contextual-dependant way, which therefore allows AR to act as a substitute for the traditional assembly.

The concepts of utilising AR technology as a possible Quick Response (QR) code improved system through the use of Machine Learning (ML), by integrating a well-known AR toolkit, is proposed as a good way to obtain good reputation data. Furthermore, this process can be automated, and robust tracking can be achieved utilising this method [5].

The basic concepts in AR applications depend on the identification of real-world objects on the screen by tracking them, then augmenting the scene with an artificial object. Tracking is often combined with some estimation of the correct 2D or 3D world coordinates for proper placement of augmentation in the scene. The application can further be discriminated to build artificial markers for object detection and those with the ability to use "natural" image features [6] - [8].

The AR application is to be developed and incorporated in the manufacturing plant, using data that is obtained from the sensors that are identical to the Supervisory Control and Data Acquisition (SCADA) system data. This data will accord the opportunity for reading and altering of certain properties of the devices within the manufacturing plant.

 ISBN: 978-1-61208-724-5

No real applications of this process technology have been found for implementation, thus the reason for the study.

This paper proposes an optimal and efficient model utilising Machine Vision (MV) to detect and identify devices based on their positions within the manufacturing environment with the aid of the AR application. Problem statement outline the reasons for conducting this study, then the aim and objectives of how this study will be conducted.

Furthermore, the methodology section outlines the procedure in which the development of the prototype will be carried out and lastly the conclusion section outlines the overview of the feasibility of the study.

II. PROBLEM STATEMENT

Despite the greatness that AR brings in the Industry 4.0 platform, there have been constraints associated with AR technology. The predominant constraints are categorised into two, namely technology and environment.

Technological constraints on mobile AR are aligned with the resources on most smart devices. This constraint is substantiated predominantly with limited memory, limited computational capability as well as the limited graphics capability.

It is with such reasons that Industrial Augmented Reality (IAR) capabilities need to further be investigated concerning object detection, recognition, and identification of devices within a manufacturing environment.

In a controlled network, the identification of components in the process is difficult without the knowledge and background in the design and implementation process. Thus, the concept of device identification with the aid of AR application utilising identifiers such as MV other than QR codes and Radio-Frequency Identification (RFID) codes needs to be investigated.

The challenge currently experienced by industry is the development, implementation, and integration of MR systems such as VR and AR systems that will have the capability to augment manufacturing operations and deliver cost-effective, time-efficient and ameliorate the quality of service (QoS) and products.

This research seeks to develop an efficient and optimal model to use AR Application, as well as MV to detect, identify and alter the device properties in the manufacturing environment.

A. *The objectives of this study*

- To prove that vision-id could be used for device identification rather than tags or QR codes.
- To use AR application to detect errors, faults or device malfunctions in the manufacturing plant.
- To develop a test bench that will operate as a manufacturing plant.
- To determine the details of the devices by means of their position in the process within the manufacturing environment utilising MV technology.

B. *Original contributions expected from the research*

- Includes recognition of automation structures/ components by means of vision for identification of such process.
- By means of identification, the user will have the capability to view and adjust the parameters of the process in the scaled plant.

III. LITERATURE REVIEW

In recent years, computers have gained a great deal of popularity due to their capability to make human life much easier, and this has also affected the success of businesses that are universally linked to the decisive approach for the establishment of innovative technologies. As the growth and improvement in technology inflated, mobile devices also gained widespread recognition and are considered a technological game-changer due to their powerful processors, etc.

Milgram et al. [7] present the AR definition as a continuum figure where he highlights the (closeness of the AR system to the real environment) and Augmented Virtual (AV - closer to the virtual environment), which both lie between the real and the virtual environment as highlighted in Figure 1.

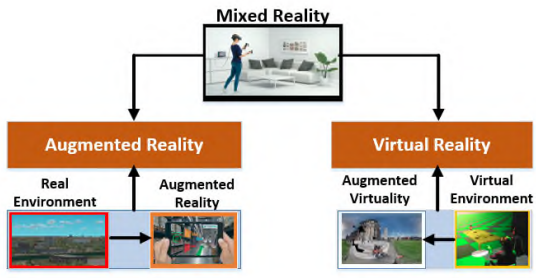


Figure 1. Reality-Virtuality Continuum [7].

 ISBN: 978-1-61208-724-5

Milgram`s definition of AR is supported by Azuma and Stapleton; however, despite support from other authors, Milgram did not cover the 3D aspect of the AR technology. Furthermore, to this discussion, Gausemeier [8], developed an application with interest in the information display.

The study focused on image-based object recognition and tracking for AR applications for inserting information into the user's field of view through a mobile device (MD) system.

Gausemeier et al. model development are aligned with the 3D-CAD file to the 3D model of the derived object. From the video point of view, Gausemeier used streams, edges, color and textile information. However, Gausemeier`s method is based on extracted pictures for filtering and comparison as compared to utilization of a live video in real-time for object detection and identification.

Silva al et. [9] in his paper outlines the method of utilising Bayesian Network Model, by calculating the probability of an event. Silva`s model is a simplification model based on Jensen Bayesian network model [10].

Silva classifies this model as a simplified model by discarding the third element which in this case is the texture and focuses his study only on the color and shape.

Antonijevic et al. [11] highlight the combination of IEC 61850 feature with augmented reality technology for providing added value visualisation capabilities in the substation automation domain. However, Antonijevic study despite using IEC 61850 as the communication tool, the AR application is still based on device identification utilising QR codes rather than MV.

Dos Reis et al. [12] present an application that uses a paranomic augmented environment to extend the information shown to power systems operators supporting data interpretation, monitoring, and manipulation. However, dos Reis, study is based on a live environment and proved the feasibility of IAR.

Furthermore, to IAR application, Marcincin et al. [13] present new attitude in imaging of combined working environment and its practicality for realising the principle of the utilisation of half-silvered surface, which provides the advantages for displaying AR objects directly in the working view of the user and free motion without hardware device connectivity.

IV. METHODOLOGY

In recent years, the QR codes have been identified as an improvement technology to Bar Code (BC) technology. The introduction of QR technology came about the limitations that are experienced in BC technology such as storage capacity and character type.

The QR codes introduction came about the capabilities of encoding and decoding of the different types of data such as binary, numeric, alphanumeric, etc. as indicated in Figure 2. QR codes are pinpointed to have a significant problem associated with slow QR detection.

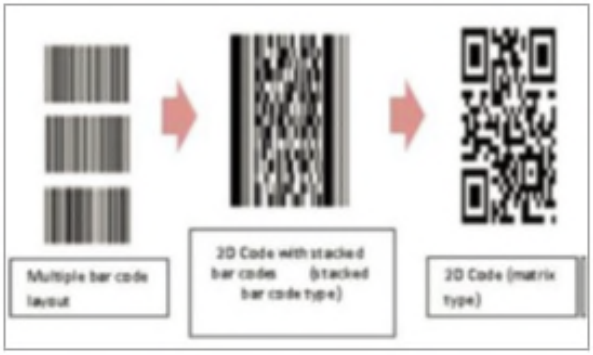


Figure 2. Multiple bar code to 2D [14].

A key upcoming type of technology (MR) will be used in the same principle in which QR technology has upgraded the BC technology.

The device tracking will be utilising the high-resolution camera, MV and OpenCV library. While the AR application, which is regarded as the computer technology that enhances real environments through visual represented information which has the following features:

- Object tracking;
- Ability to superimpose virtual objects on to user`s view of a real-world scene;
- The positioning of virtual objects in a real-world scene;
- To combine real-world images with virtual images in real-time.

MV is defined as knowledge and approach cast-off to provide an image-based automatic examination for quality control, process control and robot guidance [15]. MV is regarded as a real-world component as highlighted in Figure 3.

There are several shortcomings such as light intensity for the camera to sense. However, such shortcomings will be depicted by the Arduino which acts as the control system due to its interface with sensors and sending of data to the SCADA were the user can monitor and be alerted of any anomalies within the factory. However, the very same information can be obtained using the AR app.

 ISBN: 978-1-61208-724-5

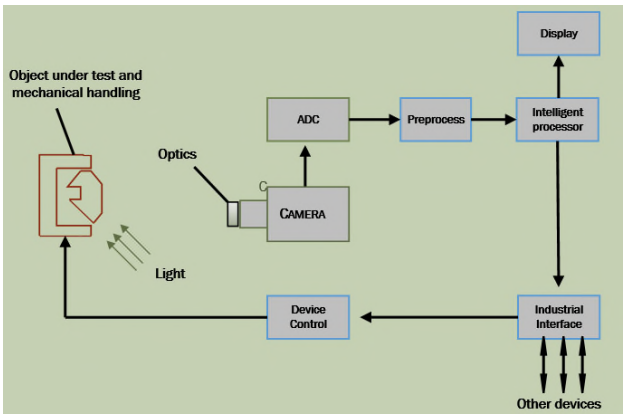


Figure 3: Industrial machine vision system for inspection [16]

Industrial Machine Vision (IMV) comprises of the following components: Lighting, Optics, Image sensing, and Image processing.

As the sensing happens, AR applications will be used to identify the devices based on their positions and outlines their details and the Artificial Intelligence (AI) of the system will kick-off in the background. Then, the AR application will be activated to detect and identify the devices while communicating with the SCADA to have identical data.

The sensory part of the system will have the input functionalities for reading in the manufacturing plant`s temperature, humidity, light intensity, and motor speed. This sensory data will register out the pneumatic outputs on the SCADA App and this data can also be obtained using the AR App.

The light sensor will be a critical sensory part of the study since the study is based on MV. The reasons are that MV depends on image acquisition utilising cameras with high resolutions whilst light depends on the light sensors. The light sensor will control the light intensity of the plant to allow the camera to operate optimally by using vision to capture accurate data. With this said, the study will be divided into 3 sections namely; sensing, automation and vision.

Automation will be accomplished by applying the Artificial Intelligence (AI) algorithm that will learn the environment as well as the behavior of the devices through Machine Learning (ML) application. This will be accomplished by the development of a vision application based on open-source software called OpenCV due to the cost affiliated with this software. C++ based software called OpenCV provides the tools needed to solve computer vision problems and also can process low-level image functions and high-level algorithm [17].

In addition, OpenCV will provide a common infrastructure for computer vision applications and also accelerate the use of machine perception in the manufacturing environment.

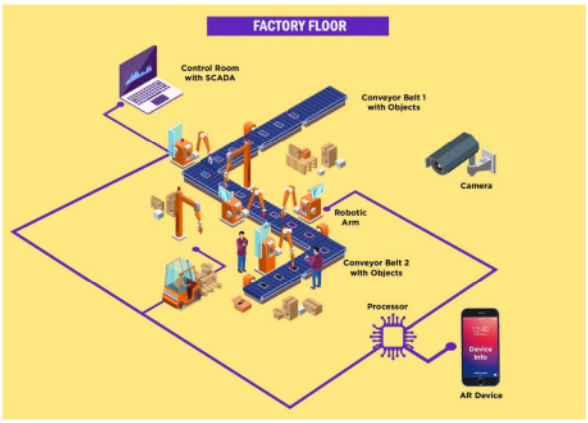


Figure 4: Proposed architecture solution for AR identification solution in Manufacturing Environment

Figure 4 highlights the proposed system for the AR application solution interfaced with a SCADA system for device detection and identification based on the AR application developed in Unity3D on an Android platform.

The SCADA application will be developed in C language through the use of Arduino microcontroller which will be interfaced to the physical sensors. Furthermore, the Graphical User Interface (GUI) will be developed in C# language using Microsoft Visual Studio 2012. MV will be realised by the use of an open source library called OpenCV technology. This MV processing will be realised through the use of a high-quality resolution camera and it will be used to identify objects on the conveyer belt for the Robotic Arm (RA) to pick.

SCADA will be linked to all the system components of the manufacturing plant such as conveyer belt, motors driving the conveyer belt, RA as well as the sensors that will provide the manufacturing environments with data such as motor failure, motor speed, light intensity, temperature, and plant humidity. These sensors will indicate when to activate the cooling system etc, or how negatively the light intensity affects the vision camera. In addition to the sensor's responsiveness, the SCADA will save the information for later analysis. However, data analysis is not accounted for in this study.

MV will be accomplished by the use of a high-resolution camera which will be placed on a fixed position on top of a servo motor to monitor, detect and identify the

 ISBN: 978-1-61208-724-5

devices (motors and a robotic arm) within the manufacturing plant.

Several OpenCV algorithms such as Canny edge detection, image recognition, color detection, and shape detection will be utilised on one side of the study utilising MV. Due to the progress in AR technology, related researches along with advanced computer hardware and software has let to AR technologies in getting more attention.

An AR Application will be built with several functionalities such as identification, detection, and recognition of devices (motors and a robotic arm). An AI application will be embedded in the background of the App for device recognition.

The application designed for this study will be used for detecting the factory phenomena such as temperature and humidity obtained by SCADA technology. Then, SCADA will act as the control system to control the movement of the motors to direct the conveyer belt within the factory. The object recognition, identification, and detection will be based on the motors as indicated in Figure 4 for detection and identification of motor information such as motor speed, motor direction, motor model, etc. However, an embedded identifier will be used as the focal point to the motor and robotic arm but in an interacted manner with the motor information. The device details obtained from the SCADA will match the device details provided by the AR application to improve productivity and have an earlier notification of device malfunction within the manufacturing plant.

V. CONCLUSION

This paper has proposed an efficient and optimal model that considers the use of AR for use in manufacturing plants.

- An approach in device detection and identification is presented where image processing through image acquisition is utilised for device identification. The knowledge of the devices and their positions within the manufacturing plant will be known through the use of an Android device that will identify those devices utilising AR technology.
- MV will be used to identify objects of a certain shape and color. Furthermore, all the sensory data from the sensors will be displayed on the SCADA system.

ACKNOWLEDGMENT

Acknowledgment for financial assistance and academic assistance from Central University of Technology, Free-State (CUT, FS), Black Engineers Enterprise (BEE), Calvary and Selosesha YMG

REFERENCES

[1] C.-F. Lin, S.-W. Lo, P.-S. Pa and C.-S. Fuh, "Mobile Application Design of Augmented Reality-Digital Pet," p. 21.

[2] G. Reitmayr and D. Schmalstieg, "Location-based Applications for Mobile Augmented Reality," vol. 18, p. 1, 2003.

[3] V. Beglov, "Object information based on marker recognition," Finland, 2013, p. 1.

[4] Y. Tian, T. Guan and C. Wang, "Real-Time occlusion handling in Augmented Reality Based on Object Tracking," p. 2886, 29 March 2010.

[5] H. Kato and M. Billinghurst, "Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conference System," *In IEEE and ACM International Workshop on Augmented Reality*, p. 4, October 1999.

[6] D. Wagner and D. Schmalstieg, "First steps towards handheld augmented reality," *In ISWC`03*, p. 3, November 2005.

[7] P. Milgramn, H. Takemura, A. Utsumi, and F. Kishino, "Augmented Reality: A class of displays on the reality reality-virtuality continuum," *Telemanipulator and Telepresence Technologies*, pp. 282-292, 1994.

[8] J. Gausemeier, J. Freund, C. Matysczok, B. Bruederlin, and D. Beier, "Development of a real-time image-based object recognition method for mobile ar-devices," *In AFRIGRAPH 03: Proceedings of the 2nd international conference on Computer graphics, virtual Reality, visualization and interaction in Africa*, pp. 133-139, 2003.

[9] R.L. Silva, P.S. Rodrigues, D. Mazala, and G.Giraldi " Applying object Recognition and tracking to augmented reality for information visualization," p.5,3 June 2014.

[10] F. V. Jensen and T. D. Nielsen, Bayesian Networks and Decision Graphs, Second ed., 2001, pp. 41-42.

[11] A. Antonijevi, S. Sucic, and H. Keserica, "Augmented Reality for Substation Automation by Utilizing IEC 61850 Communication," p. 3, 8 March 2018.

[12] P. R. dos Reis, D. L. G. Junior, A. S. de Araujo, G. B. Junior, A. C. Silva and A. C. de Paiva, "Visualization of Power Systems Based on Panoramic Augmented Environments," p. 1, 30 September 2014.

[13] J. N. Marcincin, J. Barna, M. Janak, and L. N. Marcincinova, "Augmented Reality Aided Manufacturing," p. 27, 2013.

[14] T. G. Amaral, V. F. Pires, J. F. Martins, A. J. Pires and M. M. Crisostomo, "Image Processing based classifier for detection and diagnosis of induction motor stator fault," p. 7, 11 February 2014.

[15] D. Wave, "Two-dimensional code fromthe bar code," [Online]. Available: https://www.qrcode.com/en/index.html. [Accessed 29 May 2019].

 ISBN: 978-1-61208-724-5

[16] S. Sathiyamoorthy, "Industrial Application of Machine Vision," vol. 03, no. 01, p. 1, January 2014.

[17] B. G. Batchelor and D. W. Braggins, "Commercial Vision Systems," in *Commercial Vision: Theory and Industrial Applications,* 1992, p. 406.

 ISBN: 978-1-61208-724-5

Profiling with Smart Meter Data in a Virtual Reality Setting

William Hurst, Casimiro A. Curbelo Montañez
Department of Computer Science,
Liverpool John Moores University
Liverpool, UK
Email: {W.Hurst, C.A.CurbeloMontanez}@ljmu.ac.uk

***Abstract*— Visualising complex data facilitates a more comprehensive stage for conveying knowledge. Data-to-day, we are surrounded by data. Each one of us is also a regular creator of data. For example, even simply surfing the internet, and following links, generates information that is collated by the website owner. Similarly, the introduction of the smart meter has meant that we are now generators of data from using electivity and gas in our home. Organisations are finding increasingly more interesting ways to manipulate datasets such as this. For example, smart meter data is being increasingly used for predicting load balancing within the smart grid and for the development of remote healthcare monitoring services. It is clear that by visualising complex datasets, finding answers to complex questions in an understandable manner becomes possible. Yet, interpreting large datasets using virtual reality is a concept that is still in its infancy. Therefore, in this paper, a visualisation of smarter meter data in a virtual reality setting, is demonstrated. The aim of the work is to i) outline an approach for data visualisation in virtual reality and ii) demonstrate how a virtual assistive environment can be created for remote healthcare monitoring.**

Keywords- Data Visualisation; Virtusal Reality, Smart Meter.

I. INTRODUCTION

Today, information generated from different domains continues to accumulate and is being collected at increasing rates as a consequence of living in a big data era [1]. This field is expected to keep growing and get more complex so it is important to innovate continually and develop new ways of understanding the gathered information [2]. The intricacy of the models used for analysis increase with the data complexity; this makes it even more of a challenge to deliver effective communication and visualisation of the data to end users. Therefore, in order to identify patterns and to provide insights about the data's architecture it is key to understand the information in a more convenient way. This should be achieved via the development of meaningful tools for the analyst and standard users.

The visual representation of data plays an important role when presenting complex findings in an informative and engaging way. This, combined with advanced analytics, can be integrated in methods to support the creation of interactive and animated graphics on different platforms, including desktops and various mobile computing devices [3]. However, less traditional visualisation methods, such as those using immersive Virtual Reality (VR) platforms, represent a powerful and innovative approach for multi-dimensional data visualisation that can outperform traditional desktop visualisation tools [4].

Smart meters are a rich source of granular electricity consumption data. This has raised considerable attention in the recent years on a global scale due to the numerous advantages smart meters provide [5]. Supported by the Advance Metering Infrastructure (AMI) [6], smart meters enable real-time monitoring of energy usage by recording electrical data such as voltage, frequency and energy consumption information [7]. This high-resolution data collected from smart meters can ultimately provide valuable information on the electricity consumption behaviours and lifestyle of the consumers. Therefore, allowing the development of remote monitoring systems to assess independent living in populations with Dementia or Alzheimer's disease [8]. In this sense, institutions such as the National Health Service (NHS) in the United Kingdom (UK) are able to use the data collected remotely to explore the data in a novel way and provide an assessment of the patient. Based on these ideas, a virtual assistive environment concept is simulated in this paper for remote healthcare monitoring using data collected from smart meters.

This paper focuses on the visualisation of smart meter data in a virtual reality setting to maximise the perception of the data scape geometry and provide a more intuitive way to explore high dimensionality and abstraction inherent in the data. The remainder of the paper is as follows. Section II presents a background discussion on visual data analytics in VR and highlights related projects. Section III outlines the methodology adopted for this work. Section IV presents the implementation and a discussion on the work. The paper is concluded in Section V.

II. BACKGROUND

VR interfaces have been broadly used in many fields including scientific visualisation with numerous commercial and academic software systems developed in the field of physics, astronomy, biology, medicine, and engineering among others [9]. The benefits of using such technology provides a better understanding and manipulation of the data which facilitates a more efficient and comprehensive analysis [10]. In this sense, VR technologies have the potential to assist decision makers when dealing with analytical tasks. Users can be immersed in the dataset to explore it from a different perspective, with the possibility of extracting knowledge from the inside-out instead of from the outside-in as typically conducted using 2D techniques.

 ISBN: 978-1-61208-724-5

A. *Visual Data Analytics using Virtual Reality (VR)*

Despite the success of VR in scientific visualisation, it is still in its infancy in the field of information visualisation. VR environments: immersive (specifically head mounted displays) and non-immersive (desktop) 3D worlds, where a virtual world is enhanced with abstract information.

Traditionally, Science Visualization [11] and Information Visualisation [12] have been the main areas of visualisations. In the first case, data from scientific experiments is represented using three-dimensional visualisations, with various uses in biology, medicine, architecture and meteorology among other fields. On the other hand, Information Visualisation emerged to facilitate the comprehension and interpretation of the data to users utilising graphics. In addition to these two areas of visualisation, the field of Visual Analytics [13] has emerged in the past few decades. Visual analytics (VA) combines visualisation, data mining and analysis methods with suitable user interaction. To provide advanced insights in the data, especially high dimensional data. Users can be immersed into the data via Immersive Analytics (IA), which is derived from the VR and VA fields, and uses stereoscopic visualisation to immerse an individual into a virtual environment.

VR has been used to model statistical visualisations in large sets of data points [14]. The authors in [14], for example, developed an application for statistical analysis using the C2 immersive VR environment [15], and then compared it against a more traditional workstation-based tool for high-dimensional data visualisation, XGobi [16]. Users were tested in how well they detected and selected clusters, intrinsic dimensionality and radial sparsity, using several graphic methods to analyse the data (i.e., brushing and grand tour). Results demonstrated that the added dimension provided by C2 enables the users to make better decisions about the structure of high dimensional data. This, in turn, indicates the benefits of using C2 environments to improve user's productivity for structure and feature detection tasks in comparison with XGobi. System experience plays an important role when analysing the data in favour of those users with more experience interacting with desktop environments. Slower interactions in the C2 system were apparent for people with no previous experience in VR. However, the intuitive nature of virtual environments can accelerate the learning process when using immersive environments such as C2.

The role of VR (and also Augmented Reality (AR) and Mixed Reality (MR)) in Big Data visualisation has been highlighted in [17]. The authors provide an overview of past and current visualisation methods in the field of Big Data while discussing important challenges and solutions towards the future of Big Data visualisation using immersive analytics (the combination of VR and Visual Analytics). These challenges are related to current technology development, as well as human limitations [18]. Advances using such techniques, will ultimately help improving human challenges related to their ability to manage the data, extract information and gain knowledge from it.

In the utility domain, VR has vaguely been explored with some exceptions where AR instead of VR has been used [19][20]. Therefore, this represents an ideal opportunity to investigate how Immersive Analytics can be used to extract knowledge from smart meter readings (i.e., anomalies). Angrisani et al. [19] utilised augmented reality to develop an approach to improve home power consumption awareness. The authors included sensors into the system, designated to measure power factor, current and active power consumed by household appliances. Experiments conducted using the AR reality system allowed users to access electrical consumption associated with appliances' corresponding load in a simple an easy way rendering the power values in a smart device (smartphone or tablet). Based on the information reported, the users were able to decide whether to switch the appliances off or not. The experiments conducted showed the potential of AR in energy monitoring within household context.

B. *Virtual Reality Applications*

Other VR applications have been explored in our previous work where we demonstrate the use of VR for training and productivity enhancement. For example, in [21], a VR crane simulation is outlined, as presented in Figure 1. The aim of the application is to train drivers/operators in a safe environment, improving the productivity behind the training stage of crane operation. The application allows the user to move the crane around and the view changes dynamically. The prototype shows where a crane can be positioned within a real-world 3D virtual environment, taking into account the swing and rotation of the crane beam.

Figure 1. Virtual Crane Simulation [13]

In [22], a VR proton beam therapy unit is constructed using an actual building information model of a proton beam therapy treatment unit. The application acts as a metric for supporting patients by providing an opportunity for the individual to be prepared mentally for the treatment process.

Figure 2. Virtual Proton Beam therapy Unit [14]

 ISBN: 978-1-61208-724-5

The application presented in Figure 2 has the potential to be a staff training metric. The interaction is simplified and provides an effective platform for inducting new staff into the treatment room and processes involved. Both projects are a clear demonstration of the role VR can play in the way we understand out environment, and it is a clear transformative technology for training and communication.

Building on this background investigation and our related VR development work, in the following section the methodology adopted for this project is presented.

III. METHODOLOGY

In the methodology, an overview of the process flow adopted for this research is presented. The data employed in the visualisation and the techniques used to structure the data for use in a VR setting are also outlined.

A. Process Flow

The process involves a six-step pipeline, as outlined in Figure 3.

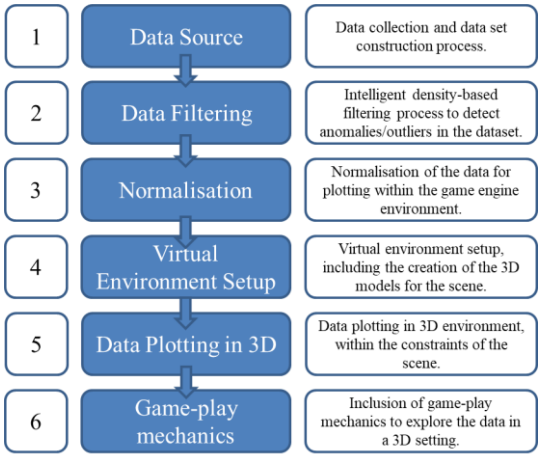


Figure 3. Process Flow

The data source in this paper is comprised of smart meter data from 10 homes, collected over twelve months. The energy readings are taken at 30-minute intervals. An overview of the data is presented in Figure 4, which displays the total energy usage for one individual in the dataset. The y-axis displays the KiloWatt Hour (KHW) energy usage and the x-axis is the rowID for the energy reading within the dataset.

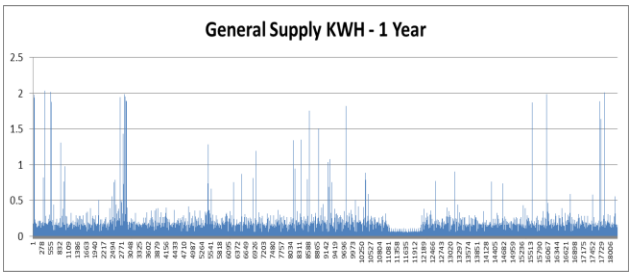


Figure 4. Data Overview

The x-axis, therefore, corresponds to time and displays the progression of the energy usage over one-year. However, it is the anomalous points in the data that are of specific interest. For example, patterns or changes in energy consumption, which deviate from the norm that may be indicated by an anomalous point in the dataset. Given the size of the dataset, simply visualising the raw data would not be an ideal metric for exploring the data in a virtual environment. For that reason, a Local Outlier Factor (LOF) algorithm is applied to the data.

B. LOF Clustering

The LOF process filters the data. Anomalies then stand out from the overall dataset. In order to calculate a LOF anomaly score, the number of variants according to the mathematical combination is calculated in (1). The LOF anomaly score measures the local deviation of density through determining how isolated the value given by *k*-nearest neighbours.

$$\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{k(k-1)\ldots 1} \tag{1}$$

A value of 1 indicates that an object is comparable to its neighbours (inlier). Likewise, a value below 1 indicates a dense region. A value significantly above 1 indicates an anomaly (outlier). Any value above 2 is considered, as we are interested in the higher-value outliers in the dataset to ensure that they are clear anomalies. The dataset, presented in Figure 4, is presented subsequently as a LOF plot in Figure 5. Through visualising the anomalies in this way, outliers can be highlighted more clearly, through deviations from the dense regions of data. Individual anomalous data points, with an associated time stamp, stand out by having a greater outlier value.

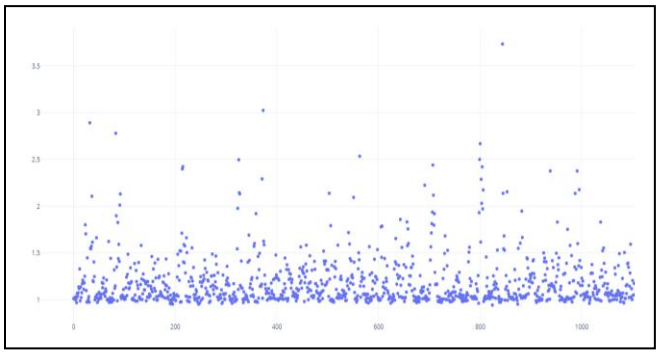


Figure 5. LOF Visualisation

Once the data had been processed through the LOF algorithm, it can be inserted into the VR environment.

IV. IMPLEMENTATION

For the 3D development, Unity game engine is employed. Models are created externally in a 3D modelling environment and imported into the scene.

 ISBN: 978-1-61208-724-5

A. *Import Data into Unity*

Inserting the raw LOF results into Unity process is displayed in Figure 6. The data points are represented by a simple smart meter 3D model in the 3D space. The LOF results are not constrained to a small space, and are disbursed of a significantly large area, which cannot be explored easily in a VR environment.

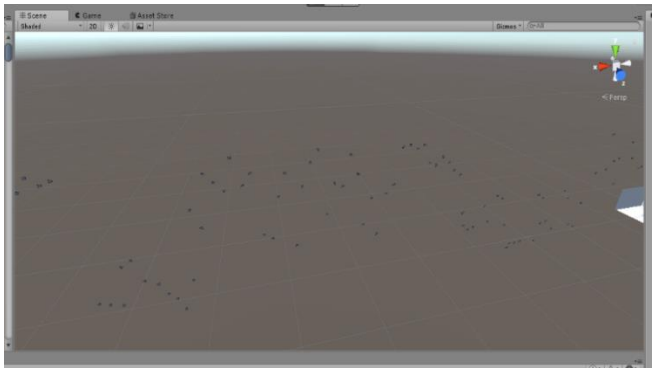


Figure 6[a]. Plotting Data in unity

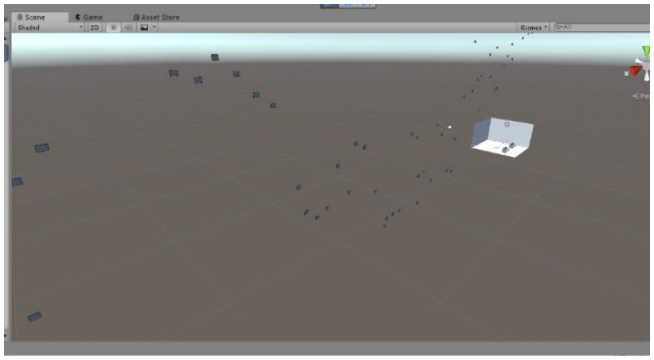


Figure 6[b]. Plotting Data in unity

The 3D data exploration room is visible in Figure 6b. This demonstrates the scale of the data plotted into the world. Ideally, the data should resemble a 3D plot of the LOF anomaly scores, such as the one displayed in Figure 7. Where, as before the x-axis shows the row ID from the dataset, the z-axis displays the anomaly score and the y-axis shows the density. In this case, the data is confined to a small 'explore-able' space.

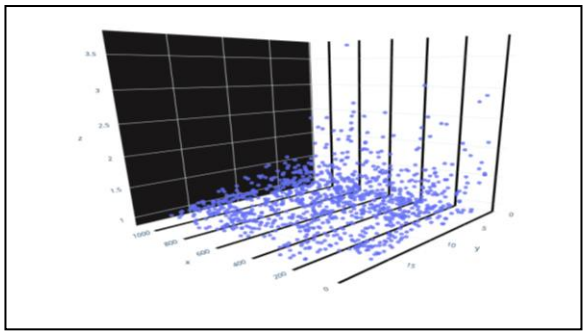


Figure 7. 3D LOF Plot

B. *Normalising th Data on Import*

Therefore, in order to constrain the data points to a room environment that would allow the user to explore the data requires a normalisation process. Within Unity, a 'data plot' game object is inserted, which correlates the data points to an x, y, z, co-ordinate in the 3D space. This is displayed in Figure 8. The process is achieved by scaling all the values between 0-1, with the maximum and minimum values from the dataset used to define the size of the data plot in the environment.

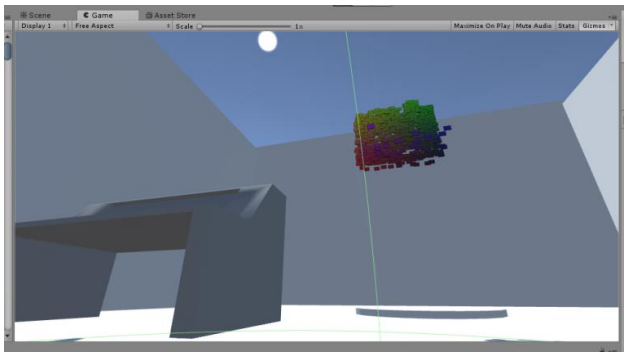


Figure 8. Data Normalisation

At this point, a First Person Shooter (FPS) asset is included in the game world, to allow the player to explore the environment and the dataset. The data plot is also moved closer to the floor, as displayed in Figure 9.

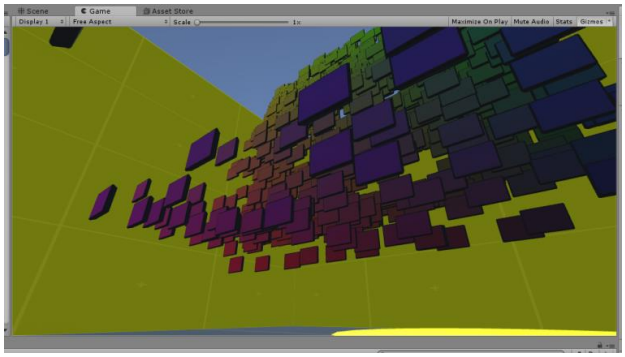


Figure 9. FPS View of the Data Plot

In an ideal setting, the game world would allow for multiple users, who can explore remotely and discuss the data patterns being visualised. To simulate this concept, characters are added, as displayed in Figures 10 and 11.

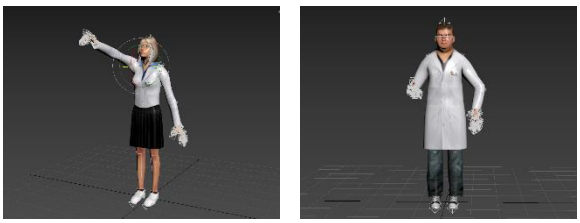


Figure 10. Scene Character Examples

 ISBN: 978-1-61208-724-5

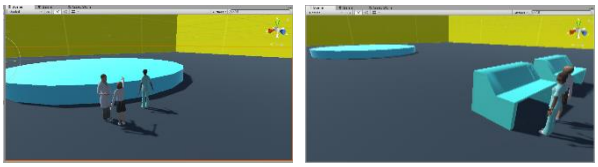


Figure 11[a]. Characters in Scene

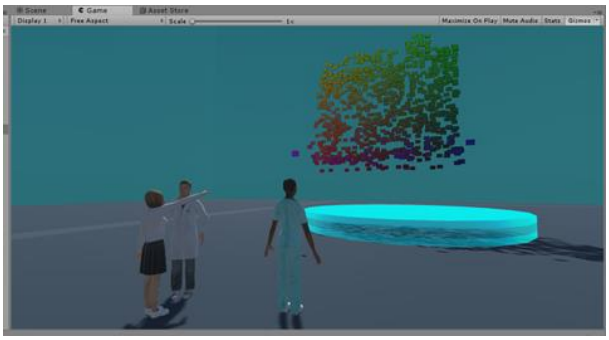


Figure 11[b]. Characters in Scene

The final prototype scene is displayed in Figure 11b. The centre disk is projected as a hologram within the scene. This is implemented using a shader in Unity. The interaction within the environment is rudamentary and can be presented using the standard KLM interaction times, as presented in Table II.

TABLE I. KLM INTERACTION TIMES

KLM Interaction Times - PC		
Task	Time (seconds)	Abbreviation
Mental preparation	1.35	M
Home on Keyboard/Mouse	0.40	H
Pointing	1.10	P
Press left click	0.10	Kl
Press right click	0.10	Kr
Turn left (A Key)	0.28	Kl
Turn right (D key)	0.28	Kr
Move Backward (W key)	0.28	Kb
Move Forward (S Key)	0.28	Kf

For example, walking around the room from the console to the data plot would inolve the following steps

$$Task\ A\ =\ M + H + \left(Sb\ *\ n\right) \tag{2}$$

C. Discussion

The next stage of the development is three-fold. 1) To set up the real-time insertion of data into the environment and 2) to integrate smart meter data at lower sampled intervals, as actual patient monitoring from 30-minute samples is a considerable challenge. For example, in related work, by using 10-second intervals the detection of household appliances is possible [7][8]. 3) At this stage we will employ supervised machine-learning algorithms, such as a Support Vector Machine (SVM) and the Bayes Point Machine binary classifier, to detect actual device interactions. Algorithms will be developed to construct device usage, time, day and device combinations. This forms the premise for actual activity construction within the home environment. This means the VR process could be personalise for individual use. In order to associate devices with behaviours, selected algorithms for behavioural modelling may include the back-propagation trained feed-forward neural network classifier, the levenberg-marquardt trained feed-forward neural net classifier, for example. These techniques are well-known algorithms and are selected for their ability to learn normal and abnormal values in a dataset [6-8]. By using the above techniques, the ambition of the work is to set up a real-time remote patient monitoring VR application.

V. CONCLUSION AND FUTURE WORK

This paper presents proof of concept demonstration of the use of VR and the integration of a dataset. The future direction of this work will include adding interactivity with the data point, so that users will be able to view the time stamp and the anomaly score of the data point. In addition, we will also experiment with the inclusion of other datasets so that the user will be able to view more than one person at once for the comparison purposes in the game environment. The NPC characters will be replaced by actual avatars of the users allowing clinicians to monitor patients remotely from a shared 3D environment.

ACKNOWLEDGEMENTS

This research project and the special track session are both funded by the EPRSC - EP/R020922/1. Owing to the ethical sensitive nature of this research, the data underlying this publication cannot be made openly available.

REFERENCES

[1] Y. Lin, H. Wang, J. Li, and H. Gao, Data source selection for information integration in big data era, Inf. Sci. (Ny)., vol. 479, pp. 197–213, Apr. 2019.

[2] J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, Expanding the Frontiers of Visual Analytics and Visualization, 1st ed. London: Springer London, 2012.

[3] Intel IT Center, Big Data Visualization: Turning Big Data Into Big Insights: The Rise of Visualization-based Data Discovery Tools, Intel IT Cent., no. March, pp. 1–12, 2013.

[4] L. Wang, G. Wang, and C. A. Alexander, Big Data and Visualization: Methods, Challenges and Technology Progress, Digit. Technol., vol. 1, no. 1, pp. 33–38, 2015.

[5] Y. Wang, Q. Chen, T. Hong, and C. Kang, Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges, IEEE Trans. Smart Grid, pp. 1–1, Feb. 2018.

[6] R. Rashed Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, A survey on Advanced Metering Infrastructure, Int. J. Electr. Power Energy Syst., vol. 63, pp. 473–484, Dec. 2014.

[7] C. Chalmers, W. Hurst, M. Mackay, and P. Fergus, A Smart Health Monitoring Technology, in Intelligent Computing Theories and Application. ICIC 2016. Lecture Notes in Computer Science, Springer, Cham, 2016, pp. 832–842.

[8] C. Chalmers, W. Hurst, M. Mackay, and P. Fergus, Smart meter profiling for health applications, in 2015 International Joint Conference on Neural Networks (IJCNN), 2015, pp. 1–7.

 ISBN: 978-1-61208-724-5

[9] M. Lanzagorta, L. Rosenblum, E. Kuo, and R. Rosenberg, Using Virtual Reality to Visualize Scientific, Engineering, and Medical Data, Sci. Vis. Conf. 1997, p. 161, 1997.

[10] R. J. Garcia-Hernandez, C. Anthes, M. Wiedemann, and D. Kranzlmuller, Perspectives for using virtual reality to extend visual data mining in information visualization, in 2016 IEEE Aerospace Conference, 2016, vol. 2016–June, pp. 1–11.

[11] H. Wright, Introduction to scientific visualization. 2007.

[12] C. Ware, Information Visualization - Perception for Design, vol. 53, no. 9. 2013.

[13] J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, Expanding the Frontiers of Visual Analytics and Visualization, 1st ed. London: Springer London, 2012.

[14] L. Arms, D. Cook, and C. Cruz-Neira, The benefits of statistical visualization in an immersive environment, in Proceedings IEEE Virtual Reality (Cat. No. 99CB36316), 1999, pp. 88–95.

[15] L. Nelson, D. Cook, and C. Cruz-Neira, Xgobi vs the c2: Results of an experiment comparing data visualization in a 3-d immersive virtual reality environment with a 2-d workstation display, Comput. Stat., 1999.

[16] D. F. Swayne, D. Cook, and A. Buja, XGobi: Interactive Dynamic Data Visualization in the X Window System, J. Comput. Graph. Stat., vol. 7, no. 1, p. 113, Mar. 1998.

[17] E. Olshannikova, A. Ometov, Y. Koucheryavy, and T. Olsson, Visualizing Big Data with augmented and virtual reality: challenges and research agenda, J. Big Data, vol. 2, no. 1, p. 22, Dec. 2015.

[18] B. Sommer, M. Baaden, M. Krone, and A. Woods, From Virtual Reality to Immersive Analytics in Bioinformatics, J. Integr. Bioinform., vol. 15, no. 2, pp. 4–9, Jun. 2018.

[19] L. Angrisani, F. Bonavolontà, A. Liccardo, R. Moriello, and F. Serino, Smart Power Meters in Augmented Reality Environment for Electricity Consumption Awareness, Energies, vol. 11, no. 9, p. 2303, Sep. 2018.

[20] How EDF Energy is using augmented reality in its UK smart meter rollout | Engerati - The Smart Energy Network. [Online]. Available: https://www.engerati.com/smart-infrastructure/article/ict-data-management/how-edf-energy-using-augmented-reality-its-uk-smart. [Accessed: 05-Jun-2018].

[21] W. Hurst, N. Shone and D. Tully., Investigations into the Development of a Knowledge Transfer Platform for Business Productivity, Submitted to the Third IEEE International Congress on Information and Communication Technology, pp. 159-164, 2018

[22] W. Hurst, K., Latham, D. O'Hare, A. Sands and R. Gandy, A Case Study on the Development of a Virtual Reality Proton Beam Therapy Unit, 5th Springer International Conference on Interactive Digital Media, 2018

 ISBN: 978-1-61208-724-5

Visualising Network Anomalies in an Unsupervised Manner Using Deep Network Autoencoders

Matthew Banton, Nathan Shone, William Hurst, Qi Shi
Department of Computer Science
Liverpool John Moores University
Liverpool, UK
e-mail: m.d.banton@2017.ljmu.ac.uk, {n.shone, w.hurst, q.shi}@ljmu.ac.uk

***Abstract*—As network data continues to grow in volume, it is important that network administrators have the tools to be able to identify anomalous network flows and malicious activity. However, it is just as important that tools allow the administrator to visualise this activity in relation to other benign activity. As such, this paper will propose a method to not only identify malicious activity, but also visualise the activity and how it relates to other network activity (both benign and malicious).**

Keywords-Autoencoder; Visualisation; k-NN; Deep Learning

I. INTRODUCTION

Computer networks are increasingly important to people's daily lives, and the rate of devices connecting to IP networks is increasing. The Cisco Visual Networking Index predicts that by 2021 there will be 3.5 networked devices per capita, up from 2.3 per capita in 2016 [1]. Most of this increase is coming from mobile devices and comes with a corresponding increase in the volume of data being used, with the amount of data in existence expected to increase to 44ZB by 2020 and 3.3ZB of data being transmitted across IP networks per year by 2021 [1].

Visualising this data presents a challenge for the network administrator, and visualising attacks presents an even bigger challenge. It is not enough to simply know an attack is happening, an administrator needs to know from where an attack is originating, what kind of attack it is, what its target is, and what other kind of systems may have been affected. The 2018 Cost of a Data Breach Study [2] found that companies that contained a breach within 30 days saved over $1 million compared to those that took over 30 days. By making data clearer, breaches can be contained faster, which can save companies money. The mean time to contain a breach was found to be 69 days.

Current systems allow the administrator to see an overview of a network and can include statistics and relevant details such as total network traffic, or even traffic over certain connections. However, these graphs typically lack a security view specifically, and are more focused on letting an administrator see which services and equipment may be under strain. Alternatively, security-focused services tend to provide anomaly detection and alert administrators to the presence of suspicious activity rather than providing clear visualisation combined with the normal network activity [3][4][7].

This paper will propose a method to visualise anomalous network activity. Anomalous activity will be detected using an Autoencoder network, which feeds into a k-NN (k-Nearest Neighbour) classifier. The results of this will be plotted onto a force-directed graph, which will highlight anomalous nodes clearly for the network administrator.

Deep learning can aid with this significantly. k-NN has been used in various methods of anomaly detection in the past [5], however accuracy has frequently been an issue as the noise of most network data means k-NN methods can fail to adequately classify data [6]. The Autoencoder deep network can aid with this by reducing the amount of noise in the data and increasing classification accuracy.

As such this paper shall propose a novel deep network to review network data and plot it in a form that allows a network administrator to see any anomalous activity, along with other relevant network details.

The rest of this paper is structured as follows. In Section 2, we will discuss other relevant work within machine learning and visualisation. In Section 3, we will discuss the methodology used within the experiment, including a brief description of unsupervised methods and why they are being used, as well as more detail about the Autoencoder and k-NN methods that will be used. Section 4 includes more detailed methodology and initial experimentation, including a detailed discussion of the structure of the model and rationale for any choices made, as well as initial results. Section 5 includes conclusions and future work.

II. RELATED WORK

Several tools already exist to allow administrators to view network activity, structure or alerts. However, many of these are unintuitive or omit important information. For instance, the OpenDaylight SDN (Software Defined Network) Controller [7] comes with a visualisation tool that allows the administrator to see a representation of all switches and nodes connected to the network, and how they are connected. It does not provide the administrator with any additional useful information, such as traffic volume over certain trunks, nor does it provide the administrator with any security alerts as no kind of Intrusion Detection System (IDS) is included. Alternatively, there are systems like Snort [8], which do not include any visualisation at all, simply alerting administrators that suspicious traffic has been detected. While its popularity speaks to its usefulness, it is purely text based, and sends alerts independent of other

 ISBN: 978-1-61208-724-5

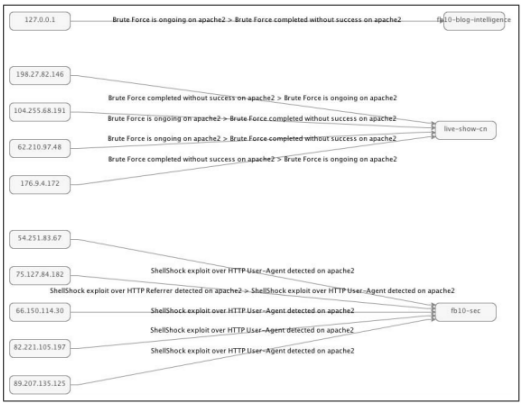


Figure 1. Line Graph model produced from SNORT output.

Figure 2. Input data hits for the 5x5 SOM.

network conditions or traffic, making it easy for an administrator relying on it to miss other important information.

Researchers have tried to address this in several ways. For instance, in [9], the authors propose a system to visualise Snort logs by converting them into line graphs, showing the victims, attackers and types of attacks. This improves on the basic Snort logs, as it becomes easier to see which alerts are related (by host, or system) and which may be completely coincidental. However, it is dependent on the logs of Snort and therefore misses other potentially important information that may be needed (such as overall network load). In addition to this, on larger systems the graphs start to become harder to read as more and more data needs to be included within them, an example is shown in Figure 1.

Within [10] the authors use k-means clustering to group network data into groups (normal and anomalous), and then take the data from the anomalous cluster and use the same process to separate it into Transmission Control Protocol (TCP), User Datagram Protocol (UDP) and Internet Control Message Protocol (ICMP) traffic. After taking the cluster, which contains a mix of all three protocols they finally design a ruleset designed on this cluster and apply it to the testing set, which results in a detection rate of attacks above 80% for all five data types (normal, Denial of Service (DoS), Remote to Local (R2L), User to Root (U2R) and probe) except for U2R.

For unsupervised learning, the authors of [11] use Robust Autoencoders (RAE), which is an Autoencoder that splits the training data (X) into normal and outlier elements (L and S), such that $X = L+S$. The purpose of this is to avoid fitting anomalous or rare data, which should prevent underfitting of normal data, and help anomalous data be highlighted more easily. They find that RAEs are an effective way to reduce false positives and do have the benefit of not underfitting normal data. However, the approach was only used for port scan type attacks, and so may not scale as well when looking for other attack types.

Potluri, *et al.* [12] evaluate Stacked Autoencoders and Deep Belief Networks (DBN) as feature reducers, with Softmax Regression and Support Vector Machine classifiers. They found that the stacked Autoencoders achieve higher levels of accuracy than the DBN when classifying fewer classes, whilst DBN and Soft-max Regression achieved higher accuracy with more classes.

Alom and Taha [13] look at attack detection using Auto Encoders and Restricted Boltzmann Machine (RBM) for feature extraction and dimensionality reduction, and combine it with k-means clustering for classification. They show that the combination can produce an accuracy of 92.12% with 9 features, or 90.86% with only 3 features. This compares to using k-means alone with 39 features for 87.72% accuracy or an Extreme Learning algorithm (again with 39 features) gaining an accuracy of 89.17%. This shows the potential of Autoencoders for categorising sets with extremely limited data sets, something that becomes more important in SDN environments with limited flow features.

Palomo *et al.* [14] decide to use a self-organising map to group and highlight network data. Using real network data captured from four subnets of a university network, they create a dataset from 150,871 samples, where 1 sample is a single packet. Each sample consisted of nine features, namely IP source address, IP destination address, protocol, source port, destination port, date, time, packet length and delta time. They find that self-organising maps can be an effective way to group similar network data, highlighting suspicious network data clusters. While the clusters do represent distinct network activity, the size of each node is determined by the amount of traffic that cluster exhibits. This means that the comparatively small amounts of anomalous network activity could be confused with other benign network protocols that generate low volumes of traffic if an administrator were to not examine the details more closely. For example, in Figure 2 nodes 18, 19 and 20 look very similar, however node 19 represents benign Address Resolution Protocol (ARP) traffic while nodes 18 and 20 represent suspicious activity from Russia and Italy.

III. PROPOSED METHODOLOGY

As we have seen, visualisation of network states is an ongoing research area, with many papers and projects

proposing different ways to visualise the current state and data being transmitted. However, research into visualising network anomalies has not kept pace with this work. This paper intends to propose a method to both detect and visualise network anomalies, making it easier to see what kind of attacks network administrators are dealing with, thus allowing them to react quicker.

Our system proposes using an Autoencoder deep network to reduce the features of the SDN, and then using k-Nearest Neighbour to sort the resulting data. As has been shown by [15] and [16], reducing noise in data can improve accuracy for k-NN, and other shallow learning methods, and this is an important stage in our model.

The k-NN can then classify the reduced data into groups, allowing related data to be grouped together. This is then placed into a force directed graph, which will show the administrator the related flows in a clear and concise manner.

While other researchers have proposed similar unsupervised models to this, results are only ever given in a table or using a ROC curve. The method proposed allows the administrator to quickly see which flows are malicious, and which ones are benign.

A. Unsupervised Methods

The use of both Autoencoders and k-NN means that the system is unsupervised. Unsupervised machine learning has benefits in not needing labelled data to train the models. Labelled data within network environments can be difficult to access, and typically requires skilled administrators or other Network Intrusion Detection System (NIDS) to label the data appropriately. As such, many researchers have proposed that using unsupervised methods is more appropriate for intrusion detection. Unsupervised methods tend to have lower accuracy and more false positives than supervised methods, as shown by [17] and [18]. However, this is not necessarily always the case, as shown by [11] and [13], where the authors show an unsupervised method using recurrent Autoencoders can be effective when attempting to detect port scans and show that they can gain lower false positives.

B. Autoencoders

An Autoencoder is a neural network designed to learn the features of a set of data. Within an Autoencoder the desired output is the input itself. So for input I and output O, $I = O$. However, there are also one or more hidden layers that are smaller than the input, forcing the network to encode a representation of the input which can then be decoded into the output.

This forces the model to learn a representation of the input data that it can use to attempt to recreate the eventual output. The goal of this is to reduce noise or unneeded features in an automatic manner. This has shown to be effective within network security, as network data tends to include a lot of noise that is not relevant to classifying the data [11] [19]. Within this context, noise refers to data that is unimportant to the overall network wellbeing. Within larger or more complete datasets this is often low level network admin data (e.g., Dynamic Host Configuration Protocol (DHCP) joins and parts), but within higher levels this can persist with for example, benign retransmissions of data.

C. k-Nearest Neighbor

As noted, the purpose of the Autoencoders is to reduce the noise within the data, not to classify any data. To classify the data we will use a k-NN algorithm. This unsupervised algorithm classifies data based upon a plurality vote of its nearest neighbours as to which class it belongs in.

D. Plotting the Graph

The output of the model is a list of x, y coordinates for each neighbour, on each flow (so for 5 neighbours, each flow will have 5 sets of x, y coordinates). From here, each coordinate can have its results averaged (creating an average x, y coordinate for each flow) and these are added as nodes to the graph. The final step is to create the links. The same averages are run, however each time a node is created with the same x, y coordinate as another flow, a link is made between them. The result is a graph that joins similar flows together, while dissimilar malicious flows will be separate, making them easier to identify.

IV. Initial Experimentation

In this section, we will describe the experiments undertaken for this research, whilst detailing more about the proposed method and reasoning behind the choices made.

A. The Dataset

The dataset used for this research is the real network data from the University of Twente [20]. The dataset consists of connection monitoring for multiple SSH servers, which is organised into flows matching the IP Flow Information Export (IPFIX) [21] standard. The data is not labelled, however, with the use of the unsupervised learning technique proposed this is not a problem. The dataset consists of network flows recorded on four routers and includes Date first seen, Duration, Protocol, Source IP Address and Port, Destination IP Address and Port, Number of Packets, Bytes and Number of flows. Also included is the logs from the SSH servers, which allows us to create a dataset of mixed log and flow data. Pre-processing was performed in order to convert text data (protocol, date) into a numerical form. Due to the size of the dataset, a 5% subset was used. This was split into training and testing subsets, with approximately 66.66% training to 33.33% testing, in order to make twice as much training as testing data. Simple random sampling was used to select the 5% of the dataset to be used, as well as to determine whether the record would be part of the training or testing datasets, and this was accomplished using the Python random function. The dataset consists of primarily benign data, however slightly less than 2% are malicious flows that include brute force or dictionary attacks. The dataset is therefore not evenly

 ISBN: 978-1-61208-724-5

distributed, but is a fair representation of what other SSH server network traffic may look like.

B. The Model

As stated, we took 66.66% split of the dataset and used it to train the Autoencoder part of the network, whilst leaving the remaining data for testing. The output from the Autoencoder is fed to the k-NN algorithm. Again, k-NN classifies the data it receives based upon a plurality vote of its nearest neighbours. A representation of the model is shown in Figure 3.

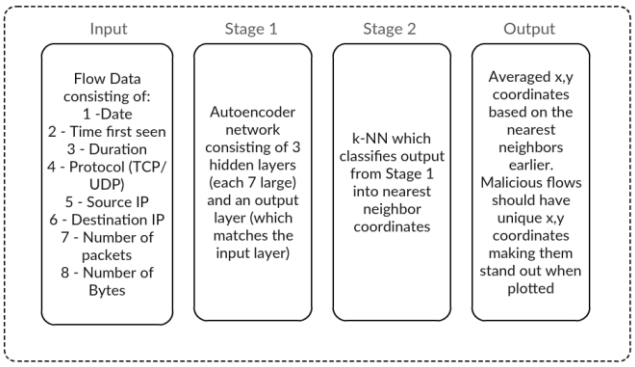


Figure 3. A representation of the model being used.

1) The Autoencoder

The Autoencoder network uses a size 8-7-7-7-8 in order to reduce noise within the data, where the sizes 8 are the input and output respectively, and the 7-7-7 is the middle hidden layers. This is based off similar model shapes in [15], who use a similar structure, in addition to our own testing where we found adding more layers leads to overfitting on the majority class. This structure should allow the model to remove noise without losing valuable data. Additionally, 500 epochs were chosen to train, along with a batch size of 200 and a learning rate of 0.01. These values were based off [11] and [22] who choose very similar values for their models, however as these works were not using the same datasets further optimisation could be done. With additional testing on the dataset itself, a batch size of 200 and 500 epochs showed no signs of overfitting, so these values were chosen as final values.

2) k-NN

After training within the Autoencoder network the output was given to the k-NN algorithm to train and classify. The dataset consists of both anomalous "brute force" access attempts and regular access attempts (with occasional legitimate access attempts rejected due to admin errors). As such, the k-NN model should produce two primary clusters, one of legitimate access attempts and one (significantly smaller) of illegitimate access attempts. The model was set to give the five closest neighbours to the input. A larger value could have been set; however, this would increase the amount of time it takes to process the model, and would increase the complexity of the final graph. There is a possibility that having too many neighbours would produce a graph that does not show the different outputs as the similarities of the benign and malicious data would work to pull them together.

3) Tools Used

GPU-based Tensorflow running on an NVIDIA RTX2080 Ti 11GB GPU was used to construct the Autoencoder, with the results of the Autoencoder going to train the k-NN. The k-NN was coded using sklearn, and processed on an i9-9900 CPU. The output of the k-NN was a text file with the node and nearest neighbors to it. This is converted into a JSON file and then imported into the NVD3 generated graph.

Finally, some areas of the model have not been fully optimised and further accuracy could be gained as a result of further optimisation. In particular, the number of epochs and learning rate were chosen based upon common values in other works, which used different datasets. Some testing was done to ensure these values were still relevant, but could still be optimised further. The number of epochs could likely be increased above 500, without signs of overfitting, but this will come with a corresponding increase in the amount of time to train and test the model. As has been noted in Sections 1 and 2, time is an important aspect in intrusion detection, and the quicker intrusions can be detected and contained, the more money can be saved.

C. Results

In Figure 4 we can see the result of the graph that came from the k-NN without the use of the Autoencoder network. As can be seen, anomalous results are not as easy to identify, they are separate nodes, however due to the loose clustering of benign flows, and it could be easy to miss malicious flows. Figure 5 shows the results from the model using the Autoencoder network to reduce the noise. The malicious flows are more notable due to the aggressive clustering of benign flows, and a busy administrator would be able to note them and gain useful and additional information from the suspicious flows. In Figure 5 benign nodes have clearly joined up, while the malicious nodes are separate, while in Figure 4 the benign nodes are not as joined up, making them harder to identify at a glance. Shown is the result for flows found within a 2 hour period on 1st Feb, as the NVD3 process used had trouble managing more flows than this.

 ISBN: 978-1-61208-724-5

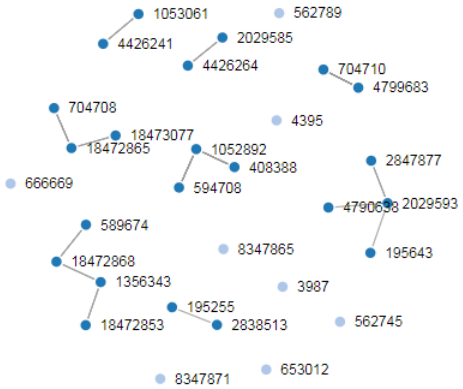


Figure 4. Shallow learning graph.

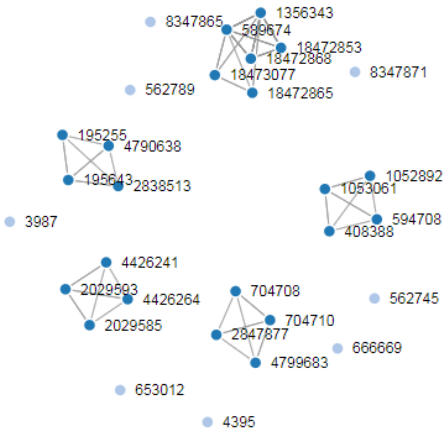


Figure 5. Deep learning graph.

V. CONCLUSIONS AND FURTHER WORK

In this Section, we will provide further discussion on the results, as well as outlining the future direction of our work.

A. *Conclusions*

As Figure 5 shows, Autoencoder networks are an effective method to reduce noise in network data, and highlight anomalies that can then be detected by a shallow learning algorithm, in this case k-NN. We have shown that by mapping the output of the k-NN onto a force directed graph, we can more easily visualise the malicious flows, and how they interact with other flows. Additionally, this graph allows mouse rollover to give more information about the flow, allowing the administrator to quickly identify malicious flows, flows that are related to it, and gain additional information, such as IP address of the source and destination, ports and volume of data. Force graphs are clearly an effective method of conveying this information to the administrator in a clear and concise manner, which if implemented in a production environment could reduce the risk of administrator error and speed up response time. Figure 4 shows the equivalent shallow learning graph.

B. *Further Research*

Other types of unsupervised learning could be used for classification and noise reduction. We chose k-NN as the classifier due to its speed and ability to gain accurate results. However, k-means could offer similar benefits and results, and has been shown to be effective when paired with models that reduce the dimensionality of the data it is classifying.

Unsupervised versions of DBN could also be effective. When being used to classify data in a supervised manner, DBNs have been shown to produce highly accurate results on common datasets, but DBNs do not have to be trained in a supervised manner. Using a RBM network instead of Autoencoder might result in higher accuracy, although again the amount of time to train and test the network would need to be considered.

This paper has focused on unsupervised methods, and it is the authors' belief that unsupervised machine learning is preferable over supervised methods, simply because of the difficulty in obtaining fully labelled datasets for production environments. Obtaining labelled data typically requires highly skilled administrators to manually review the training data, and label accordingly. This is a time consuming and, given the expertise required, often expensive process, that is still prone to error. However, often supervised methods do provide higher accuracy, and in a world where some breaches are detected over a year from the initial attack (and, it must be assumed, some are never detected) it could be argued that the extra cost of supervised methods is worth the extra accuracy. With this in mind, methods such as Support Vector Machines or Softmax classifiers could be considered, especially Softmax where the output

 ISBN: 978-1-61208-724-5

probabilities could be mapped directly onto the force directed graph.

REFERENCES

[1] Cisco, "Cisco Visual Networking Index : Forecast and Methodology , 2016 – 2021 Investor relations Search jobs," p. 2021, 2017.

[2] P. Institute, "2018 Cost of a Data Breach Study: Global Interview," no. July, p. 15, 2018

[3] "SIEM | LogRhythm." [Online]. Available: https://logrhythm.com/products/siem/. [Accessed: 30-May-2019]

[4] W. C. Lin, S. W. Ke, and C. F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," Knowledge-Based Syst., vol. 78, no. 1, 2015

[5] B. Xu, S. Chen, H. Zhang, and T. Wu, "Incremental k-NN SVM method in intrusion detection," Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS, vol. 2017-Novem, pp. 712–717, 2018

[6] T. O. Project, "OpenDaylight," 2019. [Online]. Available: https://www.opendaylight.org/. [Accessed: 30-May-2019].

[7] Cisco, "Snort." [Online]. Available: https://www.snort.org/. [Accessed: 30-May-2019].

[8] Cisco, "Snort." [Online]. Available: https://www.snort.org/. [Accessed: 30-May-2019]

[9] A. Azodi, F. Cheng, and C. Meinel, "Towards better attack path visualizations based on deep normalization of host/network IDS alerts," Proc. - Int. Conf. Adv. Inf. Netw. Appl. AINA, vol. 2016–May, pp. 1064–1071, 2016.

[10] B. Langthasa, B. Acharya, and S. Sarmah, "Classification of network traffic in LAN," 2015 Int. Conf. Electron. Des. Comput. Networks Autom. Verif., pp. 92–99, 2015.

[11] G. Kotani and Y. Sekiya, "Unsupervised Scanning Behavior Detection Based on Distribution of Network Traffic Features Using Robust Autoencoders," 2018 IEEE Int. Conf. Data Min. Work., pp. 35–38, 2018.

[12] S. Potluri, N. F. Henry, and C. Diedrich, "Evaluation of hybrid deep learning techniques for ensuring security in networked control systems," 2017 22nd IEEE Int. Conf. Emerg. Technol. Fact. Autom., pp. 1–8, 2017.

[13] M. Z. Alom and T. M. Taha, "Network Intrusion Detection for Cyber Security using Unsupervised Deep Learning Approaches," in 2017 IEEE National Aerospace and Electronics Conference (NAECON), 2017, vol. 2017–June, pp. 2379–2027.

[14] E. J. Palomo, J. North, D. Elizondo, R. M. Luque, and T. Watson, "Visualisation of network forensics traffic data with a self-organising map for qualitative features," Proc. Int. Jt. Conf. Neural Networks, pp. 1740–1747, 2011.

[15] P. Vincent and H. Larochelle, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion Pierre-Antoine Manzagol," J. Mach. Learn. Res., vol. 11, pp. 3371–3408, 2010.

[16] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," vol. 313, no. July, pp. 504–508, 2006.

[17] P. Laskov, P. Dussel, C. Schafer, and K. Rieck, Learning intrusion detection: Supervised or unsupervised?, vol. 3617 of Le. Springer Berlin / Heidelberg, 2005.

[18] R. C. Staudemeyer, "Applying long short-term memory recurrent neural networks to intrusion detection," Sacj, no. 56, pp. 136–154, 2015.

[19] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," Pattern Recognit., vol. 58, 2016.

[20] R. Hofstede, L. Hendriks, A. Sperotto, and A. Pras, "SSH Compromise Detection using NetFlow/IPFIX," ACM SIGCOMM Comput. Commun. Rev., vol. 44, no. 5, pp. 20–26, 2014.

[21] R. Hofstede *et al*., "Flow monitoring explained: From packet capture to data analysis with NetFlow and IPFIX," IEEE Commun. Surv. Tutorials, vol. 16, no. 4, pp. 2037–2064, 2014

[22] Y. Chuan-long, Z. Yue-fei, F. Jin-long, and H. Xin-zheng, "A Deep Learning Approach for Intrusion Detection using Recurrent Neural Networks," IEEE Access, vol. 3536, no. c, pp. 1–1, 2017.

 ISBN: 978-1-61208-724-5

A Review on the Development of a Virtual Reality Learning Environment for Medical Simulation and Training

Kieran Latham, Patryk Kot, Atif Wariach, Dhiya Al-Jumeily.
Faculty of Engineering and Technology
Liverpool John Moores University
Liverpool, United Kingdom
Email: {K.Latham, P.Kot, A.I.Waraich, D.Aljumeily}@ljmu.ac.uk

Mani Puthuran, Arun Chandran
Consultant Interventional Neuroradiologist
NHS The Walton Centre
Liverpool, United Kingdom
Email: {Mani.Puthuran, Arun.Chandran}@thewaltoncentre.nhs.uk

***Abstract*—This paper aims to discuss specific immersive Virtual Reality Medical Training platforms developed as research projects in the use of Virtual Reality for Medical Training. It looks at the technology that is utilised by different applications and investigates the methodologies employed in the development of immersive applications. This paper identifies some critical strengths for Virtual Reality Medical Training applications, such as the reduction of risk to practitioners and patients, and the ability to simulate complex scenarios that in real-world practice would be hard to reproduce.**

Keywords- Gamification; Immersive Application; Medical Training Simulation; Virtual Reality Learning Environment.

I. INTRODUCTION

Technological advancement in the area of Virtual Reality (VR) and medical training, has seen a surge in the number of institutions investigating the development of an immersive application for the pedagogical dissemination of anatomical knowledge and the practical training of medical procedures, both at commercial and research level. Many institutions are coming up with ways to utilise VR to enhance the learning experience of medical practitioners by creating immersive simulations for users to practice. As identified by the Virtual Reality Society [1], a Virtual Reality Learning Environment (VRLE) is a platform, which allows users to engage with content actively and can improve the user's ability to develop better cognitive abilities, spatial awareness and even be used to perform hazardous tasks in a risk-free environment. Another form of immersion is the Cave Automatic Virtual Environment (CAVE) [2]. In this environment, projectors are directed to between three and six of the walls of a room-sized cube, and typical setups include either having projection screens installed to receive camera projections of the virtual world, with the user wearing 3D glasses to distinguish better what is around them. However, companies are coming up with CAVE systems that can utilise Flat Panel LCD screens and optical tracking for a much more immersive and dynamic CAVE system [3]. In the medical industry, especially, many institutions are looking for immersive solutions to disseminating medical knowledge in more interactive ways and to guiding medical professionals through complicated procedures. These procedures require repetitive and consistent training of the user in order to improve the success rates of procedural cases [4]. Section II of this paper will discuss the advantages a VRLE has on medical simulation, and the effect each one has on the quality of training provided. Section III discusses some significant works related to Immersive VR medical platforms and looks at the tools and techniques utilised within the research projects. The methods involved in assessing the effectiveness of the virtual reality simulator, in terms of heuristics and quality of training by providing users with training scenarios and ways of providing qualitative and quantitative feedback. Section IV will conclude the paper by discussing the findings of the research. How users utilised the applications and the benefits and limitations identified by the researchers. As well as potential avenues for future investigation.

II. THE PURPOSE OF A VRLE

A vital aspect of the discussed immersive VR applications are the advantages they offer to traditional non-immersive VR training platforms. While it is beneficial to the user to provide them with an immersive training experience in which to improve their technical skills and coordination, There are multiple advantages to immersive VR applications as identified below.

A. User Engagement and Immersion

The fundamental purpose of a VRLE is to provide users with a content-rich learning environment for them to train in and better engage with content related to what they are learning. As identified by Byl *et al* [5], where qualitative discussion backed by user experience questionnaires identified that the virtual simulator provided a high-quality experience which when combined with its efficient usability made for an exciting experience that maintained the user's engagement. It demonstrates that a VRLE will benefit a user's development when it provides an immersive experience and improves the quality of training.

 ISBN: 978-1-61208-724-5

B. Bespoke Training Platforms

Another benefit of a VRLE is that it can be tailored to provide users with training in a specific subject or role, allowing for training programs to be developed and utilised, evidenced by Lam *et al* [6]. They identified that the use of a training simulator developed for cataract surgery allowed trainees to learn a specific set of skills required to perform this type of procedure effectively. It is clear from this that a VRLE has the potential to enhance the quality of learning, by allowing for the creation of a simulator with specific training scenarios in mind that can allow trainees to learn skills essential to their role effectively.

C. Risk-Free Training and Assessment

A standard belief across most VR platforms is that they provide users with a risk-free environment to train in that allows them to associate themselves with hazardous situations with little to no chance of actual harm occurring. The notion of risk-free training identified by Li *et al.,* explains that an advantage of medical training within a VRLE is that the user is provided with a safe virtual environment to train inside. In this project, training Clinicians perform laparoscopic procedures in a virtual environment, allowing them to develop an understanding of the steps involved in the procedure [7]. The trainees are also able to develop a better understanding of the surgical equipment and the human body through Virtual Reality training. The user can associate themselves with the complicated procedures and associate themselves with the risks involved that are likely to cause complications for the patient. In addition to this, the ethical approval requirements for simulating training are less demanding than if a trainee was to practice on a live patient, so that is an added advantage to virtual simulation.

III. Related Works

The purpose of this research project is to provide interventional radiologists with a VRLE in Endovascular Surgery, and as such, background research has been conducted into the area of existing medical VRLE and has identified several publications pertinent to this research project. These articles provide an insight into the methodologies they employed, how they approached their testing scenarios, and any conclusions they identified through the completion of any experiments. Identified through background research is that VR is used to identify medical simulators on a PC, regardless of whether these applications utilise immersive technology such as head-mounted displays or haptic feedback controllers. As such, this review will look at applications specifically for VR technology, looking at how researchers developed these applications, how they tested the simulators, the feedback received, and any noticeable differences using VR.

A. Tools and Techniques

The identification of what software and hardware were used to develop these immersive applications is crucial to the development of a VR training simulator. By identifying essential tools and techniques will help in creating a workflow to develop an application of similar nature and provide an understanding of what is available and commonly used in the development of medical simulations. The following is a list of such tools.

1) Süncksen et al [8]: The purpose of this research project was to identify whether a VRLE created that recreated radiographic procedures could improve the skills of practising radiologists. When developing the VRLE designed for x-ray imaging, authors opted to use Unity 3D game engine, which is a powerful engine used to create interactive visual applications [9], the completed application provides users with a visual representation of an operating theatre with c-arm interaction for radiography. Radiographic images produced are based on Computed Tomography (CT) and as such, use complex datasets commonly used in the medical industry. The DICOM Toolkit (DCMTK) is a plugin developed for Unity that allows for the importing of Digital Imaging and Communications in Medicine (DICOM) datasets, which are datasets produced by radiography [10]. The purpose of this is to convert DICOM data into readable datasheets to be visualised in-game inside Unity. For the virtual reality implementation of the application, the researchers opted for the HTC Vive [11]. The Vive utilises two stationary sensors that can track the user in an open environment and provide a full room-scaled environment, allowing users to walk around the environment and interact with objects more realistically.

2) Harrington et al [12]: Researchers developed a VRLE as a novel approach to providing Doctors with a training platform to improve their critical thinking and decision-making regarding patient care. The simulator developed by Harrington, C.M., *et al* for decision making utilised the Unity 3D game engine for the development of the simulator. Additionally, for extended functionality and platform support for Oculus VR in Unity, Oculus Utilities was installed to Unity, the toolkit provided additional features for use with the Gear VR. Furthermore, Autodesk 3DS Max 2014 was used for the 3D modelling and design of the virtual environment used in the simulator and provided the VRLE for the application. The simulator was designed to run on the Samsung Gear VR HMD, which is powered by the Oculus platform and consists of a Gear VR HMD. This setup allows the user to interact with the simulation, listen to what is said, assess the situation around them and make decisions.

3) Byl et al [5]: The simulator developed in this research project is a novel platform for the training and improvement of medical ultrasound imaging and spatial cognition for doctors. In order to develop this simulator, the researchers opted to follow a similar method to *Süncksen, M., et al* [8]. The researcher decided to use the Unity 3D game engine for the development of the immersive application, which is used to present the virtual world and training scenario that will

 ISBN: 978-1-61208-724-5

provide users with a content-rich learning environment. Additionally, the application made use of a "visualization toolkit for artificial and medical volumetric image data". It works like the DCMTK toolkit used in the x-ray imaging software, which allows for the realistic simulation of medical imaging in the application to improve immersion. In this instance, however, the toolkit appears to be homogenous to the Department of Applied Sciences at Flensburg University [5]. The application also utilises the HTC Vive HMD and haptic feedback controllers, which allows them to look around the virtual environment and interact with objects within, as part of the training scenarios.

TABLE I. IDENTIFIED TECHNOLOGIES FOR VR DEVELOPMENT

Project	Tools		
	Platform	***Engine***	***Plugins***
1) X-Ray Imaging [8]	HTC VIVE	Unity 3D C#	DCMTK [10]
2) Critical Decision Making [12]	Gear VR	Unity 3D C#	N/A
3) Medical Ultrasound Imaging. [5]	HTC VIVE	Unity 3D C#	Visualisation Toolkit for Artificial and Medical Volumetric Image Data [5]

Additionally, 90% of Samsung Gear apps and 53% of Oculus Rift games use Unity and C# for development. While the papers do not go into detail as to why they chose the HTC Vive, initial research suggests that the HTC Vive is capable of full room tracking using external trackers, providing enhanced motion tracking and more accurate tracking as opposed to the Oculus Rift. In terms of visual ability, the HTC Vive provides users with a complete resolution of 2,160 x 1,200 (1,080 x 1,200 per eye), and a field of view of 110 degrees, which combined with a 90Hz refresh rate allows for an improved immersive experience inside the HTC Vive [18].

B. Methods and Approaches

A crucial part of the research is how the users interacted with the system. Providing users with a series of objectives that enable researchers to assess the users level of ability and user experience of the application. The following is a list of conventional methods used for this purpose.

1) Süncksen et al [8]: In the x-ray imaging application, the aim is for the user to correctly reproduce radiographic images, with the constraints of the game being that users complete the challenge in an efficient amount of time, with a minimal amount of patient radiation dosage. The user is provided with text-based instructions and imagery to show the expected output and what the users must do to achieve this. Additionally, the user is scored based on three variables: accuracy of the radiographic image, amount of time taken, and patient dosage, with points being provided based on the user's score, while also comparing the users score to an expert's score. Furthermore, to improve the development of spatial awareness, a non-medical mode exists in which users must identify objects hidden in a box, correctly, to improve their skills in c-arm navigation.

2) Harrington et al [12]: For this project, researchers developed an application that places the user at the centre of a traumatic situation in which they need to make critical decisions and diagnoses pivotal to the patient's survival, RSCI Medical [19]. The application has the user following a patient through the early stages of hospital arrival, listening to the doctors and nurses provide information about the situation. Users then choose an option regarding patient care, which contributes to a score at the end, based on whether the patient survived. It requires the user to critically think and lets them experience the quick paced and stressful environment where it is imperative to take on multiple sources of information at once in order to succeed in making the right decisions.

3) Byl et al [5]: This paper on ultrasound imaging takes an unconventional approach to gamifying the training of ultrasound imagery by taking users out of a medical environment and into an industrial setting, with the objective being to conduct ultrasound scans on packages containing objects to determine that the contents were packaged in the correct box in the correct way. This scenario allows the user to gain experience conducting ultrasounds and analysing what is seen on screen, allowing them to determine whether what they are seeing is correct. The simulation uses a point system that increments as the player makes correct guesses, as well as how long it took for users to make decisions, and the total number of correct answers. Additionally, a leader board is provided to compare high scores to other users as an incentive to perform better and obtain a higher score.

C. Evaluation and User Experience

As a part of the user experience, researchers can record qualitative and quantitative feedback from users regarding the user's experience of an application, the functionality, and determine the applications overall effectiveness.

1) Süncksen et al [8]: While utilising the application, users had several tasks to accomplish; this exposed the user to multiple aspects of the system, which would allow for a variety of responses from the users. Users provided feedback in the form of a user experience questionnaire (UEQ) [14]. Users were asked to provide quantitative feedback regarding the usability and effectiveness of the x-ray imaging application. 65.85% (27) of users agreed wholeheartedly that the system is adequate for medical training, 29.27% (12) of users agreed and provided additional feedback on potential improvements, 2.44% (1) of users mostly agreed that the application was useful, lastly, 2.44% (1) of users disagreed with the notion of using this application for medical training.

 ISBN: 978-1-61208-724-5

2) Harrington et al [12]: The application evaluated the use of VR technology for the training of Doctor's critical thinking and ability to diagnose patients in a stressful environment. The user experience questionnaire used a Likert scale between 0 and 7; scores were considered negative if they were below 3, with scores above five classed as positive. User feedback identified that the RCSI Training Simulator received positive results with an average score of 5.09 out of 7 regarding Immersion and Realism of the VR simulation, 5. Regarding the method of learning, it was rated 5.7 out of 7 as a useful teaching tool, with 58% of candidates claiming their belief that there currently are not enough patient management simulators available.

3) Byl et al [5]: Evaluation of the novel application for medical ultrasound imaging was conducted using User Experience Questionnaires which provided quantitative responses regarding aspects of the system using a Likert scale between -3 and 3, with results between -0.8 and 0.8 being neutral. Nine users tested the application: 6 male and three male participants, with five already possessing experience with VR technology. Quantitative responses from the UEQ looked at six factors: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. Results showed that the Attractiveness, Efficiency, Stimulation, and Novelty categories all scored good results over 0.8. Regarding Perspicuity and Dependability, users believed improvements could be made to improve information dissemination, a modification of object transforms to accommodate changes in users height, and audio cues to support their stimulation.

IV. CONCLUSION

The literature review conducted in this paper indicates that the use of VR technology is indeed beneficial to the learning experience of medical professionals. The feedback from users regarding the projects revealed that most users believe VR to be a useful tool for medical training and assessment. The training applications reviewed benefitted from a content-rich and realistic working environment for the user's training. The applications developed also utilised a point system, which would keep score of the user's progress throughout the tasks and be indicative of their performance, which provides accurate measurement for supervisors to refer to when reviewing the user's progression when training. Furthermore, a common occurrence in these projects was users initially struggling to understand how to accomplish tasks within the simulation, this combined with the low number of users that have experienced VR, indicates a lack of familiarity with VR technology and immersive applications.

The lack of familiarity about VR in the medical community could be worth investigating, potentially identifying methods in which VR could further enhance the quality of quality as identified by the reviewed projects. Furthermore, the use of gamification in medical training as a metric for progress review could be reviewed further to develop a greater understanding of the impact this could have on monitoring progression within medical training platforms.

REFERENCES

[1] Virtual Reality Society (2011) *Virtual Reality Learning Environments* [online] Available at: https://www.vrs.org.uk/virtual-reality-education/learning-environments.html [Accessed: 18th February 2019]

[2] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart, "The CAVE: audio visual experience automatic virtual environment," Communications of the ACM, [online] vol 356, pp.64–72. Available at: http://portal.acm.org/citation.cfm?doid=129888.129892, 1992. [Accessed 18th Feb. 2019].

[3] A. Febretti, "CAVE2: A Hybrid Reality Environment for Immersive Simulation and Information Analysis," Proceedings of SPIE - The International Society for Optical Engineering, vol 8649, pp.01-13. [Accessed 30th May. 2019].

[4] M. Maytin, T. P. Daily, and R. G. Carillo, "Virtual reality lead extraction as a method for training new physicians: A pilot study," PACE - Pacing and Clinical Electrophysiology, vol 383, pp.319–325, 2015. [Accessed 25th Feb. 2019].

[5] B. Byl, M. Süncksen, and M. Teistler, "A serious virtual reality game to train spatial cognition for medical ultrasound imaging," 2018 IEEE 6th International Conference on Serious Games and Applications for Health, SeGAH 2018, pp.1–4, 2018. [Accessed 19th Feb. 2019]

[6] C. K. Lam, K. Sundaraj, and M. N. Sulaiman, "Computer-based virtual reality simulator for phacoemulsification cataract surgery training," Virtual Reality, vol 184, pp.281–293, 2014. [Accessed 18th Feb. 2019].

[7] L. Li, et al, "Application of virtual reality technology in clinical medicine," American journal of translational research, [online] vol 99, pp.3867–3880. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28979666, 2017. [Accessed 19th Feb. 2019].

[8] M. Süncksen, et al, "Gamification and virtual reality for teaching mobile x-ray imaging," 2018 IEEE 6th International Conference on Serious Games and Applications for Health, SeGAH 2018, pp.1–7, 2018. [Accessed 19th Feb. 2019].

[9] K. Yang and J. Jie, J, "The Designing of Training Simulation System Based on Unity 3D," In: 2011 Fourth International Conference on Intelligent Computation Technology and Automation. [online] IEEE, pp.976–978. Available at: http://ieeexplore.ieee.org/document/5750762/, 2011. [Accessed 30 May 2019].

[10] M. Eichelberg, et al, "Ten years of medical imaging standardization and prototypical implementation: the DICOM standard and the OFFIS DICOM toolkit (DCMTK)," Proc.SPIE, 5371. Retrieved from https://doi.org/10.1117/12.534853 [Accessed: 29 May. 2019]

[11] J. Egger, et al, "HTC Vive MeVisLab integration via OpenVR for medical applications," PLOS ONE, [online] 123, p.e0173972. Available at: https://dx.plos.org/10.1371/journal.pone.0173972 [Accessed 30 May 2019].

[12] C. M. Harrington, et al, "Development and evaluation of a trauma decision-making simulator in Oculus virtual reality," American Journal of Surgery, [online] vol 2151, pp.42–47. Available at: https://doi.org/10.1016/j.amjsurg.2017.02.011. [Accessed 18th Feb. 2019].

[13] J. Kim, C. Y. L. Chung, S. Nakamura, S. Palmisano, and S. K. Khuu, "The Oculus Rift: a cost-effective tool for studying visual-vestibular interactions in self-motion perception," Frontiers in Psychology, [online] vol 6, p.248. Available at:

 ISBN: 978-1-61208-724-5

http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.00248/abstract [Accessed 30 May 2019].

[14] M. Schrepp, et al, “Design and evaluation of a short version of the user experience questionnaire (UEQ-S),” IJIMAI - International Journal of Interactive Multimedia and Artificial Intelligence, vol 3, pp.103-108, 2017. [Accessed 29th May. 2019].

The Rational Dilation Wavelet Transform: A Flexible Tool for Perception-inspired Signal and Image Processing

Vittoria Bruni
Dept. of SBAI
University of Rome La Sapienza
Rome, Italy
e-mail: vittoria.bruni@uniroma1.it

Domenico Vitulano
Institute for Calculus Applications (IAC)
CNR
Rome, Italy
e-mail: d.vitulano@iac.cnr.it

***Abstract*—The aim of this work is to investigate the properties of an adaptive multiscale transform for defining perception-based methods for signal and image processing. Particular attention is devoted to the possibility of partitioning the frequency plane in a flexible way, which depends on both human perception and task purpose. Examples concerning image enhancement and timbre recognition will be presented and discussed.**

Keywords-contrast sensitiviy; rational dilation wavelet transform; MEL cepstrum; adaptive scale selection.

I. INTRODUCTION

In the last years, there has been a huge research work concerning time-frequency transforms, since many problems of signal and image processing can be successfully solved by expanding the signal in a proper basis where signal features are emphasized. In general, this property is indicated with the term sparsity, i.e., the signal is represented as a series expansion in a proper basis where only few coefficients are non zero. Unfortunately, there is not a unique optimal basis for each kind of signal or for each kind of problem; moreover, sometimes there is not a unique optimal basis for a given signal, since it depends on its spatial/spectral components. That is why the family of transforms/bases is wide. They differ according to the class of functions they are able to compactly represent, the existence of a computable inverse transform and a fast algorithm for their discrete implementation.

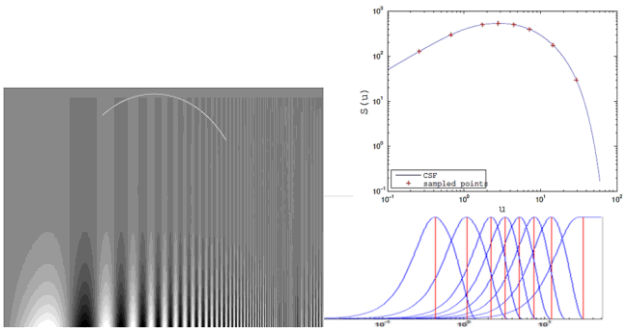


Figure 1. Left) Campbell-Robson map. The white curve is the contrast sensitivity function (CSF). Right) CSF sampling grid (top); corresponding frequency partition (bottom).

On the other hand, in several applications there is the need of reproducing the different spectral components of a signal or image which not always obey a logarithmic law. Let us consider, for example, visual perception. The Campbell Robson image in Figure 1 represents a sinusoidal stimulus having changing frequency and contrast --- contrast is constant at each row and decreases from bottom to top; frequency increases from left to right. However, the stimulus is not perceived in the same manner in the whole image. At a fixed distance, the sinusoidal shape is perceived just below the white line and it is more evident in the middle of the image. By changing the observation distance, the curve shifts, changes its amplitude, shrinks or dilates. The curve that separates the perceptually homogeneous region (top) from the non homogeneous part (bottom) is the Contrast Sensitivity Function (CSF) [1]. It would be then desirable to construct a CSF that is better adapted to the content of the analyzed image in order to simulate frequency axis partitioning to eye sensitivity, i.e., more dense close to CSF maximum. Hence, a multiscale transform that changes its frequency resolution according to CSF shape has to be defined: higher resolution is required at frequencies to which human eye is more sensitive while less resolution is allowed far from them. Similarly, in audio processing, Mel scale is the one that better simulates the ability of the auditory system to distinguish two similar sounds [2]. Even in this case, Mel scale does not correspond to a logarithmic partition of the frequency axis.

The paper is organized as follows: Section II briefly revises the rational-dilation wavelet transform (RADWT) [3] properties. Section III shows two representative examples, while the last section draws the conclusions and provides hints for future work.

II. THE RATIONAL-DILATION WAVELET TRANSFORM

RADWT [3] is an overcomplete discrete wavelet transform in which the dilation factor can be set between 1 and 2 to perform a more gradual scaling between consecutive subbands, as depicted in Figure 2. It is a powerful tool for signal analysis and our purposes for the following reasons:

- it allows a tunable scale factor (known as Q factor). It means that it gives a time-scale representation of the signal where the scale parameter changes according to a factor that is smaller than two. Hence, the high

 ISBN: 978-1-61208-724-5

frequencies of the signal are analysed with a finer resolution than in the dyadic case;

- it is implemented through a filter bank by using just a couple of filters (low-pass and high pass) that satisfy the perfect reconstruction condition (Figure 2);
- it involves downsampling and upsampling operations (Figure 2) even though it has some redundancy with respect to the dyadic case. The redundancy depends on the Q factor, i.e., the frequency resolution chosen for the analysis;
- it allows a high flexibility in the construction of the involved filters since it is possible to select not only the Q factor and/or the redundancy, but also the filter decay in the transition band;
- it also has a straightforward extension to 2D, making it also suitable for image analysis.

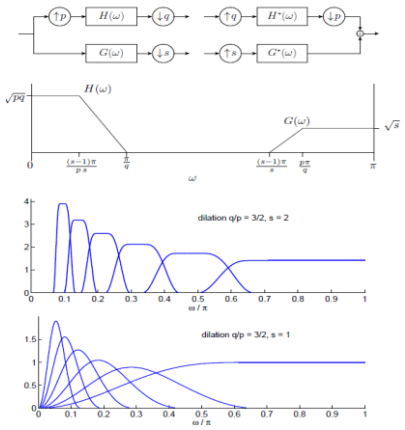


Figure 2. Top) RADWT implementation. Bottom) Frequency partition provided by two different Q factors.

Based on these properties, the dilation factor can be adjusted at each scale while preserving perfect reconstruction property. As a result, RADWT provides a useful tool for defining a multiscale transform that changes its frequency resolution according to human perception.

III. EXAMPLES

In the field of audio processing, the Mel-cepstrum transform [2] combines two elements: the logarithmically spaced Mel scale, modeled on the human auditory system; and the cepstrum transform, which allows separation between excitation and resonances. The aim of our research is to mimic the Mel scale using RADWT to obtain the Mel frequency cepstral coefficients (MFCC). The main idea is to find the parameters p_j, q_j, s_j, such that the support of the high pass filter G_j at level j is close to the support of the j-th Mel band. The result is shown in Figure 3, where the energy distribution in the adaptive RADWT of three different instruments is also shown. The energy distribution clearly characterizes the instrument and then it represents a feature that can be successfully used in timbre recognition.

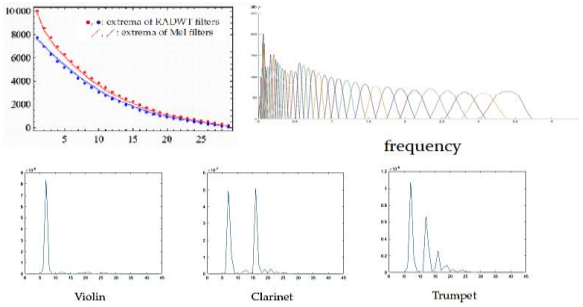


Figure 3. Top) RADWT-based frequency partition according to MEL filters. Bottom) RADWT energy distribution of three different instruments.

With regard to visual perception, the adaptive transform derives from an adaptive sampling of the contrast sensitivity function related to the analysed image. CSF sampling is based on the conjecture in the field of visibility, which asserts that CSF is the envelope of the contrast sensitivity of the cortical cells that take part to the visual perception process [1]. Hence, it is possible to design a filter bank covering all the frequency axis but having bandwidths that are adapted to the curve shape: tighter around the point of maximum visibility and wider elsewhere, as shown in Figure 1.right. The distance between consecutive points is fixed such that the interpolation error is less than a prefixed tolerance. Once the transform is fixed, expansion coefficients can be processed according to the task, as for example, denoising, deblurring, fusion, etc. For example, for degraded images, coefficients can be modified in order to map the CSF of the degraded image to the one of the ideal not degraded image, preserving or emphasizing frequencies close to the maximum of the curve. Preliminary results showed that whenever this mapping is used as a preprocessing step of conventional restoration methods, the final result improves in terms of objective measures, as Peak Signal to Noise Ratio (PSNR), and, especially in terms of image quality assessment measures, as Structural SIMilarity Index (SSIM).

TABLE I. DENOISING USING RADWT-BASED CSF MAPPING AS PREPROCESSING STEP: IMPROVEMENT WITH RESPECT TO STANDARD DWT

Metric	**Denoising results (Lena image)**		
	Noisy image	***Denoised after DWT-based CSF mapping***	***Denoised after RADWT-based CSF mapping***
PSNR	20.22	28.31	**28.40**
SSIM	0.599	0.788	**0.812**

IV. CONCLUSIONS

In this paper the properties of the rational-dilation wavelet transform (RADWT) have been revised and the advantages in using it in some perception-based models have been discussed. Specifically, contrast sensitivity function sampling and Mel-like frequency axis partition have been considered and the benefit given by RADWT has been demonstrated respectively in some classical image processing problems and for audio processing purposes. Future research will be devoted to define automatic and fast

 ISBN: 978-1-61208-724-5

methods for the definition of the best level-dependent dilation parameters for the analysed data and/or the task.

REFERENCES

[1] P. Barten, "Formula for the contrast sensitivity of the human eye," Proc. of SPIE El. Imaging, vol. 5294, 2004.

[2] L. R. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition," Prentice Hall, 1993

[3] I. Bayram and I. W. Selesnick, "Frequency-domain design of overcomplete rational-dilation wavelet transforms," IEEE Trans. Signal Process., vol. 57, pp. 2957–2972, Aug. 2009.

 ISBN: 978-1-61208-724-5

Model-Based 3D Visual Tracking of Rigid Bodies using Distance Transform

Marios Loizou

Department of Computer Science
University of Cyprus
1 Panepistimiou Avenue, Aglantzia
Nicosia, Cyprus
Email: mloizo11@cs.ucy.ac.cy

Paris Kaimakis

School of Sciences
University of Central Lancashire Cyprus
12 - 14 University Avenue, Pyla
Larnaka, Cyprus
Email: pkaimakis@uclan.ac.uk

***Abstract*—The core idea of model-based 3D tracking is that of continuously estimating the pose parameters of a 3D object throughout a sequence of images, e.g., a video feed. Here, we present an edge-based method for achieving 3D object tracking, via Gauss-Newton optimization. We rely on natural features observations, like edges, for the detection of interest points and by using the 3D pose of the object in the previous frame, we correctly estimate its new 3D position and orientation, in real-time. There is also a C++ implementation of the visual tracking system, with the use of the OpenCV library, which can be found in our GitHub repository (https://github.com/marios2019/Visual_Tracking).**

Keywords–Object 3D tracking; Model-based; Gauss-Newton optimization; Distance Transform

I. Introduction and Related Work

Object 3D tracking is used in a variety of Computer Vision applications, like Augmented Reality [1] where virtual objects are super-imposed to the scene and in Robot Object Manipulation [2] [3] where the target object is manipulated with the use of a mechanical device. Also, 3D tracking enables cultural heritage reconstruction applications, where usually through a mobile device the user can reanimate and view ancient architecture. In all of these cases, the goal is to estimate the 3D pose (position and orientation) of the object with respect to the observer.

There are many approaches to 3D tracking, depending on the targeted application and the means that are being used [4] [5]. Techniques like [6] and [7], fall into the marker-based tracking category, were they make use of point and planar markers, that are carefully placed in the scene by the user. Because of their pattern uniqueness, they can be identified as image features, which lead to 2D-3D correspondences with high precision. The latter provides reliable measurements for pose estimation. Despite their good performance, marker-based tracking techniques require engineering the environment, which sometimes the application's end-users dislike and sometimes is impossible, e.g., outdoor environments. By contrast, 3D tracking by detection techniques, are based on natural features that can be detected in the scene. Works like [8], construct a database from Scale-invariant Feature Transform (SIFT) features [9], that are detected from images with different viewpoints of the object. Multi-view correspondences can be found and the 3D positions of the features are recovered using Structure-from-Motion (SfM) algorithms. At runtime, SIFT features are extracted for each video frame, which yield to 2D-3D correspondences. Camera pose can be estimated using algorithms like Random Sample Consensus (RANSAC) [10]. Also recent advances in Deep Learning has given rise to techniques like [11] and [12] for simultaneously detecting and tracking multiple objects, although they require huge datasets for training. In this work we are focused on non-learning methods, as they do not required collecting, analysing and preprocessing huge amount of data.

Most of the work that is been done belongs to feature-based 3D tracking category, where camera pose estimation, just like tracking by detection, relies on natural features, like edges or corners. Furthermore, techniques that belong to this category provide a strong prior knowledge of the camera pose for each new frame, which aids the pose estimation task. This yields to a jitter-free camera pose between consecutive frames, unlike tracking by detection methods, where camera pose is recovered in each frame indepentedly. Edge-based methods like Real-time Attitude and Position Determination (RAPiD) [13] or [14], sample the edges of the model into 3D control points, which they are rendered and impose onto the image, along with model. Each control point is matched with a point that lies on a detected edge, by searching along the normal of the edge that the control point belongs to. Given enough control points, pose estimation is achieved, by minimizing the sum of squares of the perpendicular distances, using a least squares approach.

Our approach on model-based 3D visual tracking belongs to the feature-based 3D tracking category, specifically in the edge-based methods, as we rely on measurements being made along an edge, to find the displacement between the virtual and real object. Unlike RAPiD methods, the measurements are not being made on 3D edges of the virtual model, but on their 2D projections on the image. Additionally, we do not search along the normal direction of each edge, to find correspondences between control points and points on the detected edges. Instead, we detect features - edges from the original image and calculate the Distance Transform (DT) [15] [16] of the image, that is formed by the detected edges. The distance between the edge measurements and the detected edges is calculated using the image produced from DT. Finally, by using the Gauss-Newton algorithm, we minimize the distance between the edge measurements of the virtual model and the detected edges from the real object.

 ISBN: 978-1-61208-724-5

Section II provides a formulation of the 3D visual tracking problem. In Section III we present the architecture and of our visual tracking system and the experimental results on simulated and real data are shown in Section IV. Finally, in Section V we discuss about the merits of our method and provide some cues for further improvement.

II. FORMULATION

We treat the 3D visual tracking problem, as a procedure of estimating the camera pose (3D position and 3D orientation of the camera) relative to the object, i.e., estimate the extrinsic parameters of the real camera that led to the projection of the object onto the image. With the use of a known model of the object, we construct a virtual camera that projects and imposes the model to the real image. By finding the distance between the projection of the virtual object (model) and the real object (object in the real scene, that we want to track), we are able to estimate the extrinsic parameters of the virtual camera that minimize this distance, with the use of the Gauss-Newton iterative algorithm. If the distance of the projection of the two objects is nearly zero (global minimum), the extrinsic parameters of both cameras match, and the pose of the real camera is adequately estimated.

A. Camera Model

For our purposes, we use the full camera model, which is described by the projection matrix P, that projects each world point of a 3D scene to the image plane of the camera. The projection matrix is constructed by the multiplication of the intrinsic and extrinsic parameters matrices of the camera. We use the following intrinsics matrix, which maps a 3D point in camera coordinates (x_c, y_c, z_c) to a 2D point in pixel coordinates (u, v),

$$\boldsymbol{K} = \begin{bmatrix} f_p & 0 & u_0 & 0 \\ 0 & f_p & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 4} \tag{1}$$

where f_p is the focal length in pixel units and (u_0, v_0) is the principal point i.e., the point of intersection between the camera's optical axis and the image plane. We do not take into account any lens distortion that may occur by defects in lens design and manufacturing or by the nature of the lens. In this approach the intrinsic parameters are known and fixed, so the camera we use is calibrated. In a future implementation of our algorithm as a mobile phone application, camera calibration could easily be incorporated during installation or first use on the end-user's phone. During this stage, the end-user can be guided through the process of calibration, following the procedure explained in [17].

The extrinsic parameters matrix holds the position and orientation of the camera in world coordinates and it maps a 3D point in world coordinates to a 3D point in camera coordinates. This matrix is defined as follows,

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0}^\top & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4} \tag{2}$$

where $\boldsymbol{R} = \boldsymbol{R}(\theta_y)\boldsymbol{R}(\theta_z)\boldsymbol{R}(\theta_x) \in \mathbb{R}^{3\times 3}$ is a 3D rotation matrix represented in Euler angles and $\boldsymbol{t} = [t_x, t_y, t_z]^\top \in \mathbb{R}^3$ is a position vector. The matrix in (2) describes the position and orientation of the object relative to the observer - camera. In our approach, we want to estimate the camera pose, i.e., the position and orientation of the observer - camera relative to the object. To achieve this, we calculate the inverse of the aforementioned matrix as,

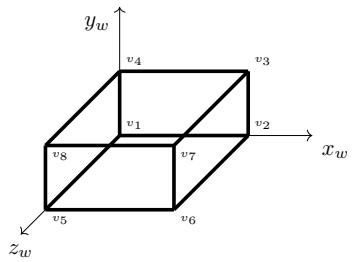


Figure 1. Model in world coordinates.

$$\boldsymbol{E}_c = \begin{bmatrix} \boldsymbol{R}^\top & -\boldsymbol{R}^\top \boldsymbol{t} \\ \boldsymbol{0}^\top & 1 \end{bmatrix} \tag{3}$$

B. Cuboid Object

We are using a cuboid object (box) for tracking, which is depicted in Figure 1. The model is composed of a set of eight vertices $\mathcal{V} = \{\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^3\}, |\mathcal{V}| = 8$ and a set of six surfaces $\mathcal{S} = \{\boldsymbol{s} : \boldsymbol{s} \subset \mathcal{V}\}, |\mathcal{S}| = 6, |\boldsymbol{s}_i| = 4, 1 \leq i \leq |\mathcal{S}|$. Because there are no other objects in the scene, the model's local coordinates are equal to the scene's global coordinates. With the use of the known intrinsic parameters $\boldsymbol{K}$ and an initial hand-picked estimation of the real camera's extrinsic parameters $\boldsymbol{E}^{(0)}$, we construct a virtual camera, that is being defined by the following projection matrix,

$$\begin{aligned} \boldsymbol{P} &= \boldsymbol{K}(f_p, u_0, v_0)\boldsymbol{E}_c^{(0)}(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z) \\ &= \boldsymbol{K}(f_p, u_0, v_0)\boldsymbol{E}_c(\boldsymbol{x}^{(0)}) \end{aligned} \tag{4}$$

where $\boldsymbol{x}^{(t)}$ is a vector containing the extrinsic parameters at time t.

Each vertex $\boldsymbol{v}_i$ of the cuboid is projected to the image plane of the virtual camera as,

$$\tilde{\boldsymbol{p}}_{\boldsymbol{i}} = \boldsymbol{P}\tilde{\boldsymbol{v}}_{\boldsymbol{i}} \tag{5}$$

where $\tilde{\boldsymbol{p}}_i = [x_p^{(i)}, y_p^{(i)}, z_p^{(i)}]^\top$ and $\tilde{\boldsymbol{v}}_i = [x_w^{(i)}, y_w^{(i)}, z_w^{(i)}, 1]^\top$ are expressed in pixel and world homogeneous coordinates, respectively. The conversion to cartesian pixel coordinates is done as follows,

$$\boldsymbol{p}_i = \begin{bmatrix} \frac{x_p^{(i)}}{z_p^{(i)}} & , & \frac{y_p^{(i)}}{z_p^{(i)}} \end{bmatrix}^\top = \begin{bmatrix} u_i \\ v_i \end{bmatrix} \tag{6}$$

C. Parameter Estimation

As we have mentioned in Section II-A, we model our camera in such a way, that the intrinsic parameters are fixed. In this case, the camera is said to be calibrated i.e., we have a prior knowledge of the camera's intrinsic parameters. This reduces the degrees of freedom (DoFs) for parameter estimation to only six; the extrinsic parameters of the camera (three DoFs for camera position and three DoFs for camera orientation).

The extrinsic parameters $\hat{\boldsymbol{x}}$ can be estimated as the minimization of the squared sum of the reprojection errors between $\boldsymbol{v}_i$ and $\boldsymbol{p}_i$,

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \sum_i \parallel \boldsymbol{P}(\boldsymbol{x})\tilde{\boldsymbol{v}}_i - \boldsymbol{p}_i \parallel^2 \tag{7}$$

 ISBN: 978-1-61208-724-5

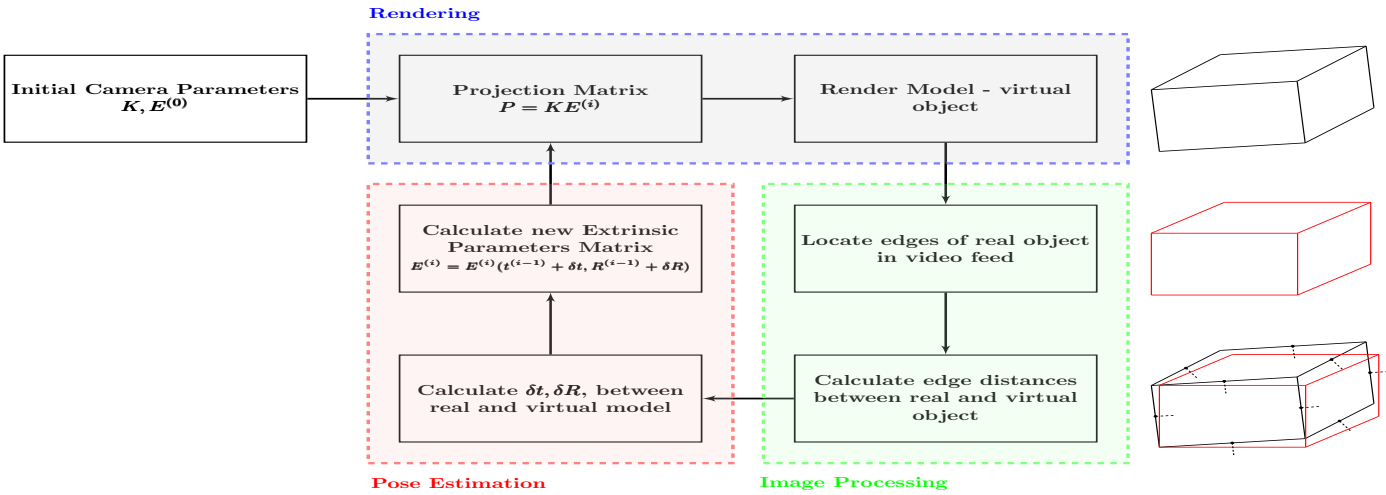


Figure 2. Visual Tracker work-flow, illustration of the main building blocks a) Rendering (blue), b) Image Processing (green) and c) Pose Estimation (red).

where $\boldsymbol{P}$ and $\boldsymbol{x}$ are the projection matrix and the extrinsic parameters, as shown in (4). The pose estimation problem can be solved by minimising the sum of residual errors, $r_i = \| \boldsymbol{P}(\boldsymbol{x})\tilde{\boldsymbol{v}}_i - \boldsymbol{p}_i \|$. Equation (7) can also be written in vector form as,

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \| f(\boldsymbol{x}) - \boldsymbol{b} \|^2 \tag{8}$$

where again $\boldsymbol{x}$ is a vector that contains the extrinsic parameters, $\boldsymbol{b}$ is vector containing some type of measurements made on the image (detected features) and $f(\cdot)$ is a function that relates vectors $\boldsymbol{x}$ and $\boldsymbol{b}$ (projection matrix as seen in (4)).

Function $f(\cdot)$ is usually of a non-linear nature, due to the perspective projection transformation. So in our case we use the Gauss-Newton optimization algorithm, which is a non-linear least squares technique [18]. With the use of prior knowledge of an initial state $\boldsymbol{x}_0$, the residual error between consecutive states within a time frame τ, is minimized as

$$\begin{aligned} \boldsymbol{x}_{k+1}^{(\tau)} &= \boldsymbol{x}_k^{(\tau)} + \delta_{\boldsymbol{x}} \\ &= \boldsymbol{x}_k^{(\tau)} - \boldsymbol{J}_k^{\dagger(\tau)} \boldsymbol{\epsilon}_k^{(\tau)} \end{aligned} \tag{9}$$

where $\delta_{\boldsymbol{x}} = -\boldsymbol{J}_k^{\dagger(\tau)} \boldsymbol{\epsilon}_k^{(\tau)}$ is the minimization step defined by the Gauss-Newton algorithm, $\boldsymbol{J}_k^{\dagger(\tau)}$ is the pseudo-inverse of $\boldsymbol{J}_k^{(\tau)}$, the Jacobian of $f(\boldsymbol{x}_k^{(\tau)})$ and $\boldsymbol{\epsilon}_k^{(\tau)} = f(\boldsymbol{x}_k^{(\tau)}) - \boldsymbol{b}^{(\tau)}$ is the residual error at iteration k.

III. Visual Tracking System

Our method consists of three main building blocks: The first building block, Rendering, is responsible for the correct projection of the known virtual 3D object to the image plane. The second building block, Image Processing, consists of image processing methods for scanning every frame of the video feed from the camera and identifying 2D image features, which are likely to describe the object of the scene. In addition, it measures the distance between the projection of the virtual object and the extracted image features. The last building block, Pose Estimation, uses a non linear fitting method (Gauss-Newton), for accurate estimation of the 3D position and orientation of the camera, so that the projection of the virtual object matches the projection of the real object. Figure 2 illustrates the general procedure of our 3D visual tracker, which consists of the aforementioned building blocks.

A. Rendering

The rendering procedure is responsible for rendering the known model of the object we would like to track. By using the virtual model as shown in Figure 1, we project each of its vertices $\boldsymbol{v}_i$ to the image plane (5). The projection matrix $\boldsymbol{P}$ is constructed as shown in (4), which models our virtual camera. Each vertex is projected to a 2D point $\boldsymbol{p}_i$ expressed in pixel coordinates. Subsequently, visibility culling techniques are applied to the projected object, to determine if it is visible form the virtual camera.

The first technique is called *front camera visibility*, where for each vertex $\boldsymbol{v}_i$ we calculate its distance d relative to the camera position $\boldsymbol{t}$. If any distance $\boldsymbol{d}$ is smaller than f_m (camera's focal length in metric units), then this vertex is not visible and in this case, we cull the whole object. Secondly, *back face culling* is used to determine which surfaces of the model are facing the camera. For each surface, the angle θ is calculated between the camera's look vector, from the camera's position $\boldsymbol{t}$ and a vertex of the surface, and the normal of the surface. If the angle is smaller than $90°$, then the surface is facing the camera and it should be rendered. Finally, *edge clipping* is applied, where each edge of the imposed model, is clipped along the borders of the image.

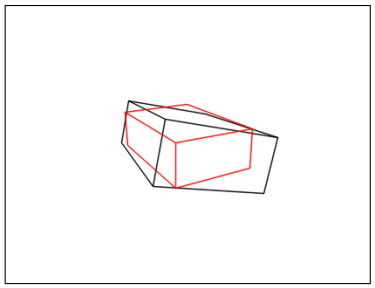
The output of the rendering procedure, for simulated and real data, are illustrated in Figure 3.

B. Image Processing

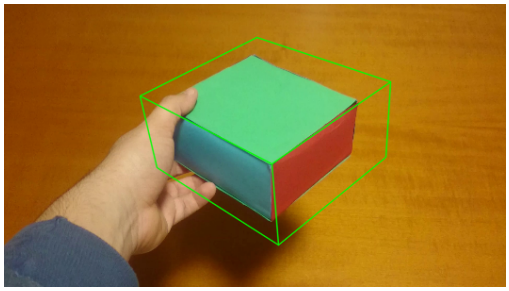
The second main building block of our visual tracking system, Image Processing, is responsible for dividing each projected edge of the virtual models into *control points* and calculating the minimum distances between them and the image features extracted from the real image. These control points help in measuring the spatial difference between the virtual and real projected object and give an estimate of how much the extrinsic parameters of the virtual camera have to be altered, so the edges of the virtual object match the edges of its real counterpart.

For each visible projected edge of the virtual model, we form its direction vector as $\boldsymbol{o}_i = \boldsymbol{p}_i^{(2)} - \boldsymbol{p}_i^{(1)}, i = 1, \cdots, |\mathcal{N}|$, where $\mathcal{N}$ is the set of all the visible projected edges, $|\mathcal{N}| = N$ is their corresponding number and $\boldsymbol{p}_i^{(1)}, \boldsymbol{p}_i^{(2)}$ are the projected

 ISBN: 978-1-61208-724-5

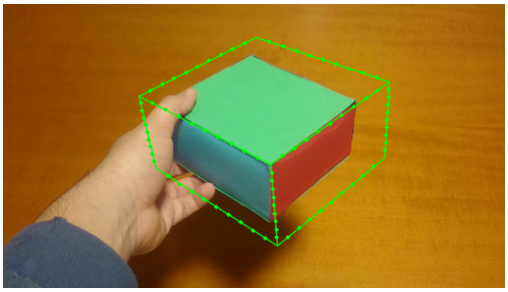


(a)

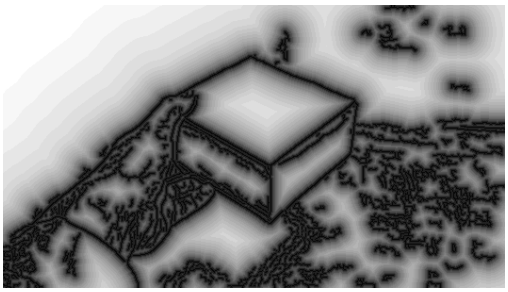


(b)

Figure 3. (a) The red and black cuboid represent our simulated data and the virtual model, respectively, (b) A frame from a real video-feed and the imposed virtual model (green cuboid).



(a)



(b)

Figure 4. (a) Control points for each edge of virtual object, (b) distance transform.

vertices of the cuboid and the endpoints of the i-th edge. The control points are then calculated as,

$$\boldsymbol{m}_{ij} = \boldsymbol{p}_i^{(1)} + j\frac{\boldsymbol{o}_i}{M-1} \tag{10}$$

where $\boldsymbol{m}_{ij}$ is the j-th control point of the i-th edge and M are the number of control points for each edge. We omit control points at the endpoints of the edge so $j = 1, \cdots, M-1$ and the correct number of control points are $M-2$. Figure 4a illustrates the calculated control points for the real data scenario.

The next step of the image processing procedure is to extract features from the real image. The type of features that we are detecting are edges, which appear as intensity discontinuities on the image. For the purposes of our method, we used the Canny edge detection algorithm [19]. Because the real image of the simulated data is rendered by us, the only information that appears on it are the projected edges so there is no need for edge detection. This is not the case for real data, as every frame of the image contains a lot of information with a high quantity of noise. The use of Canny edge detection algorithm in this case is mandatory. We also want to remove as much noise we can, so the two thresholds for the hysteresis procedure of the algorithm, are set to high values, to make sure that edges that represent noise or small and weak edges of the scene are discarded.

The final step of image processing is about finding and measuring the distances between the control points and their closest image features, i.e. the edges that have been extracted in the previous step. To achieve this, we apply the Distance Transform to the image being produced by the Canny algorithm. This results to a new image, where the value of each point is the Euclidean distance between that point and its nearest image feature, as seen in Figure 4b.

Formalizing distance transform in the context of our method, we can write,

$$d_{ij} = DT(\boldsymbol{\mathcal{F}})[\boldsymbol{m}_{ij}] \equiv \min_{\boldsymbol{f}} dist(\boldsymbol{m}_{ij}, \boldsymbol{f}) \tag{11}$$

where d_{ij} is the distance measurement for each $\boldsymbol{m}_{ij}$ (control point), $\mathcal{F} = \{\boldsymbol{f}_1, \cdots, \boldsymbol{f}_E\}, \boldsymbol{f}_e \in \Re^2$ is the set of all edge pixels $\boldsymbol{f}_e$ that have been extracted, $|\mathcal{F}| = E$ is the number of all edge pixels and $dist(\boldsymbol{m}_{ij}, \boldsymbol{f}) = \| \boldsymbol{m}_{ij} - \boldsymbol{f}_e \|_2$ is the Euclidean distance between $\boldsymbol{m}_{ij}$ and $\boldsymbol{f}_e$.

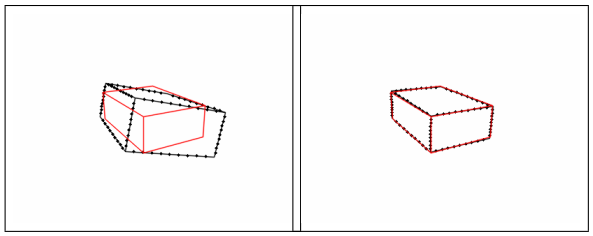
The distance d_{ij} for each control point $\boldsymbol{m}_{ij}$ can be found by treating the distance transform image as a lookup table, i.e., just lookup the value from the pixel of the distance transform image with the same position as the $\boldsymbol{m}_{ij}$ pixel position.

The main advantage of using DT is that is easy to express the distance of each edge pixel $\boldsymbol{f}_e$ and control point $\boldsymbol{m}_{ij}$ in closed form w.r.t. extrinsic parameters of the camera. So using this in the context of the Gauss-Newton algorithm, as explained in Section III-C, is straightforward.

C. Pose Estimation

The last building block, Pose Estimation, is the main procedure of the 3D visual tracker. In this phase, based on the measurements that have been made to the real image, we try to minimize the difference between the extrinsic parameters of the virtual and real camera. To achieve this, we use the Gauss-Newton algorithm, as it has been explained in Section II-C. In our case, the state vector $\boldsymbol{x} = [x_1, x_2, x_3, x_4, x_5, x_6] \equiv [t_x, t_y, t_z, \theta_x, \theta_y, \theta_z]$, contains the extrinsic parameters of the virtual camera and $f(\boldsymbol{x})$ is the function that maps the distances that are found for the control points $\boldsymbol{m}_{ij}$, w.r.t. the extrinsic

 ISBN: 978-1-61208-724-5



(a) (b)

Figure 5. Fitting of model cube (black) to real cube (red), after 5 iterations.

parameters. The residual error ϵ equals to,

$$\boldsymbol{\epsilon}_k = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{NM} \end{bmatrix} = \boldsymbol{d}_k = \begin{bmatrix} d_{1,1} \\ d_{1,2} \\ \vdots \\ d_{N,M} \end{bmatrix} \tag{12}$$

where $\boldsymbol{d}_k$ is a vector that contains the distances for all control points, for the k-st iteration of the Gauss-Newton algorithm. To iteratively minimise the residual errors we use (9). Here, the objective function that we are trying to minimize is $\boldsymbol{f}(\boldsymbol{x}) = \sum_{i=1}^{NM} r_i(\boldsymbol{x}) = \sum_{i=1}^{N} \sum_{j=1}^{M-1} d_{ij}(\boldsymbol{x})$, where $r_i(\boldsymbol{x}) = d_{ij}(\boldsymbol{x})$ is the Euclidean distance between control point $\boldsymbol{m}_{ij}$ and edge pixel $\boldsymbol{f}_k$, as a function of $\boldsymbol{x}$. Therefore, the Jacobian of $\boldsymbol{f}(\boldsymbol{x})$ for iteration k equals to,

$$\boldsymbol{J}^{(k)} = \nabla f(\boldsymbol{x}) = \left[\frac{\partial \boldsymbol{d}(\boldsymbol{x})}{\partial x_1}, \cdots, \frac{\partial \boldsymbol{d}(\boldsymbol{x})}{\partial x_S} \right] \in \Re^{MN \times 6} \tag{13}$$

where x_s is one of the 6 parameters of the state vector $\boldsymbol{x}$ and $\boldsymbol{d}(\boldsymbol{x})$ is the function that maps the measured distances from the distance transform, with the extrinsic parameters, i.e., the pose of the camera.

By minimising this quantity, the distance between the control points and image features will also be minimized. At this point, the real object's edges match the virtual object's edges, in which case the pose parameters can be inferred.

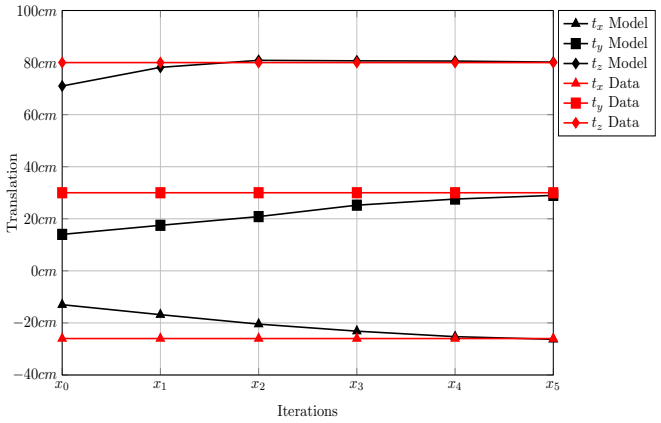
IV. Results

In this section, we present some experiments that we have made with our visual tracker both on simulated and real data, so we can determine it's strengths and weaknesses.

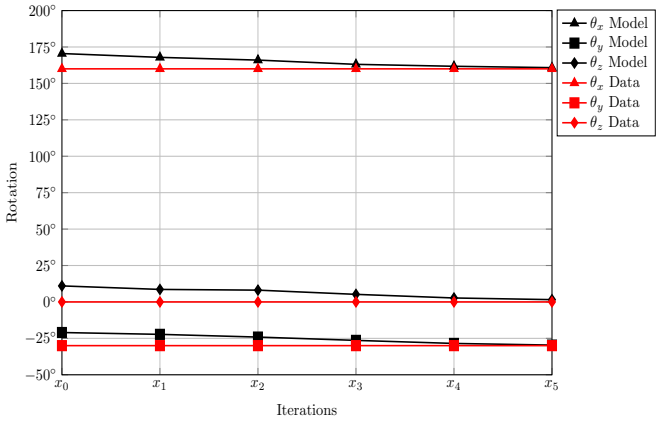
A. Simulated Data Experiments

For our experiments on simulated data, we have rendered another cuboid with the use of a computer generated (CG) camera. The intrinsic parameters for both cameras are set to the same values. The CG camera extrinsic parameters are set to $\boldsymbol{x}_{\text{CG}} = [t_x, t_y, t_z, \theta_x, \theta_y, \theta_z] = [-26\text{cm}, 30\text{cm}, 80\text{cm}, 160°, -30°, 0°]$ and we use them as ground truth. The virtual camera extrinsic parameters are set to $\boldsymbol{x}_{\text{virtual}} = [-13\text{cm}, 14\text{cm}, 71\text{cm}, 170.5°, -21°, 11°]$. Finally, we render on the same image, using both cameras, the cuboid depicted on Figure 1.

As we can see from Figure 5, after 5 iterations, the pose of the virtual object matches the simulated object's pose. The final virtual camera extrinsic parameters we get after the end of the pose estimation procedure are $\boldsymbol{x}_{\text{virtual}} = [-26.3cm, 29cm, 80.2cm, 160.8°, -29.6°, 1.6°]$, which are really close to the extrinsic parameters of the CG camera. As



(a)



(b)

Figure 6. Convergence between virtual and simulated camera extrinsic parameters, a) cameras position, b) cameras orientation.

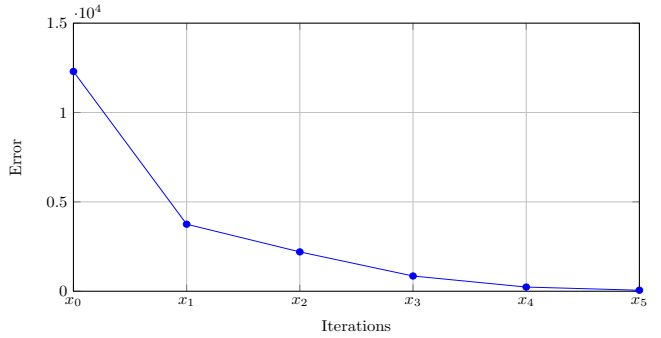
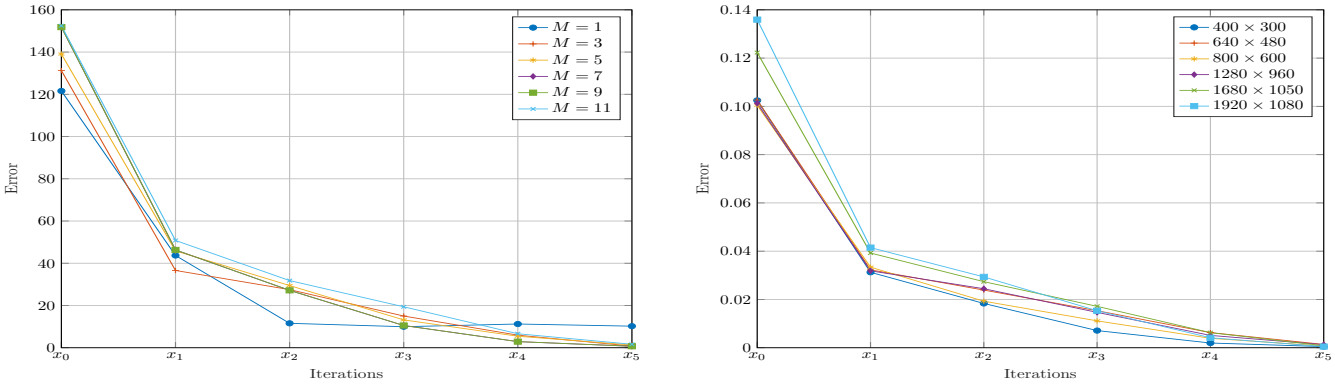


Figure 7. Error between simulated and virtual cuboid for each iteration.

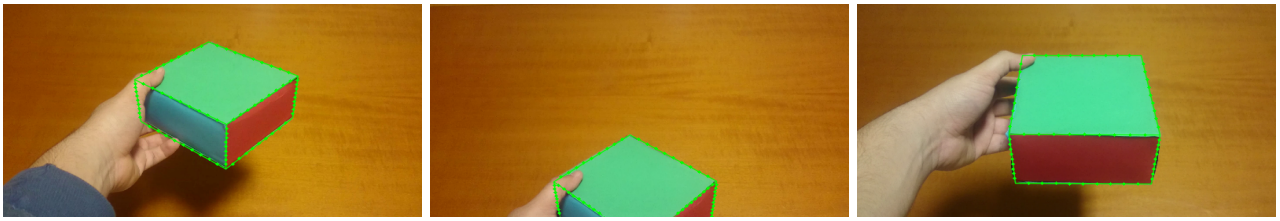
shown in Figure 6, with each iteration k, each parameter of the state vector $\boldsymbol{x}_{\text{virtual}}$ gradually converges to the the ground truth parameters. We can actually notice the gradient descent step, as for each time we get closer to the minimum, the next step tends to be smaller, because the gradient magnitude decreases.

The error between the two cuboids is calculated as the squared sum of the distances d, $E_k = \boldsymbol{d}_k \cdot \boldsymbol{d}_k = \sum_{i=1}^{MN} d_i^{2\ (k)}$,



(a) (b)

Figure 8. Error evaluation a) for various number of control points, b) for various image resolutions



(a) (b) (c)

Figure 9. a) Fitting of virtual cube on a real data image, b) Edge clipping along the borders of the image, c) Fitting of virtual cube with two surfaces visible.

for each iteration k, which is the square of the objective function $f(\boldsymbol{x})$ calculated at the extrinsic parameters $\boldsymbol{x}_k$. From Figure 7, it is clear that the error drops exponentially, another clue that the pose estimation was correct. Again, the exponential rate of the error has the same traits as the convergence of the extrinsic parameters.

As seen in Figure 8a, in all cases error converges to 0, except the case where $M = 1$ because a small amount of measurements are extracted from the frame. Furthermore, if $M \geq 11$ despite the fact that the error after 5 iterations is close to 0, the time that is needed for each frame for the fitting procedure starts to exceed the acceptable limits ($33ms$ time cap). This is even worse in the case where we are using real data, as the number of iterations is doubled to 10. Here the error is normalized w.r.t. number of the control points for all visible edges $\hat{E}_k = E_k/MN$. Also in Figure 8b , we show that our tracking method works well for various number of resolutions and aspect ratios. In all cases, after 5 iterations the squared distance between the virtual and the simulated object converges to 0. Here the error is normalized with the image area $\hat{E}_k = E_k/(Width \times Height)$.

With frames of 400×300 we ensure that our data contain sufficient information for our models to extract the correct pose, while not posing a significant computational overhead. Likewise, we choose to operate with $M = 9$ in an effort to ensure that a good majority of the elements in $\boldsymbol{d}_k$ of (12) indeed holds the distances between truly corresponding points in the model and data.

B. Real data experiments

For our real data experiments, we have used a box of cuboid shape. In this scenario, we altered the intrinsic parameters of the virtual camera, so they match the intrinsics of the real camera. The fitting of the cuboid in Figure 9a is achieved after the 3 first frames of the video. This happens because, even after we doubled the number for iterations for each frame ($k = 10$), the image gradient of the distance transform for each frame is calculated with the use of the Sobel operator, which smooths the produced image. This smoothing results to smaller Δ steps at each iteration. Nevertheless, we want to keep it that way, so it does not affect the overall accuracy of our visual tracker and the first 3 frames work as an initialization stage.

Further down the same video feed, the real object is clipped along the four edges of the image plane, in Figure 9b. Even at these conditions our tracker manages to successfully track the real object. Of course, there is a limit to the portion of the object that is being clipped, which is about half of the object. After that point, the pose estimation of the tracker becomes unstable and at some point it completely loses the correct pose of the real camera. In general, when we want to track some object on a real scene, we usually set the pose of the real camera to be looking straight at the object and the object being in the middle of the image plane. So, conditions where half or more of the object is clipped along some edge of the image, are not expected in real applications, such as Augmented Reality.

The final experiment was conducted in order to determine how many surfaces of the cuboid should be visible, for the visual tracker to correctly estimate the camera's pose. We have concluded that at least two surfaces should be always visible so the estimation of the state vector $\boldsymbol{x}$ is stable. In fact, at least

 ISBN: 978-1-61208-724-5

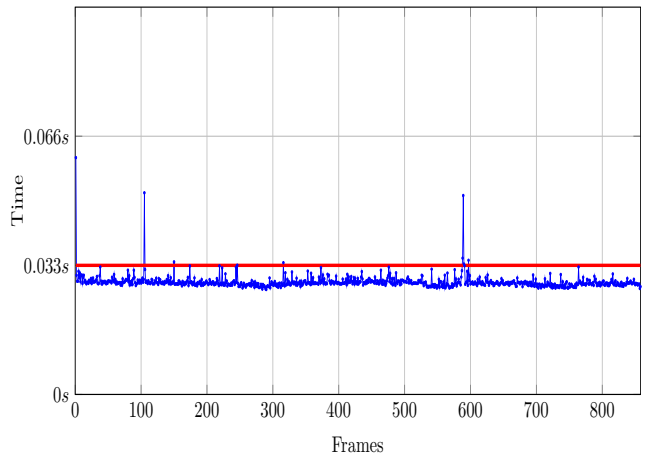


Figure 10. Elapsed time for the visual tracker process on each frame. Red line is the 33 milliseconds cap, for real time performance.

6 rows of $\boldsymbol{J}$ have to be linearly independent from each other, to be possible to calculate correctly the pseudo-inverse of $\boldsymbol{J}$. The $\boldsymbol{m}_{ij}$'s along each edge are linearly dependable, so we need at least 6 visible edges to surpass the previous constraint. For at least 6 edges to be visible, we need at least 2 surfaces of the object to be visible. This is the case in Figure 9c, where the cuboid is correctly tracked.

Finally, our visual tracker can run in real-time, i.e., process each frame of the video feed and estimate the camera's pose in under 33 milliseconds, which is the time interval between two consecutive frames in a video feed of 30 frames per second, as shown in Figure 10. This is due to the fact that, even if three surfaces of the cuboid are visible, we have to render 9 edges of the model. Each edge contains $M = 9$ control points and the overall number of control points are 81. So, for each frame we have to compute the camera's pose, using only 81 distances for the non-linear fitting iterations, which is a small number of data to process for modern workstations. This leads to the fact that we can impose correctly a virtual object that matches the real one, without adding any extra delay to the video's sequence. The experiments were conducted on a 3.1GHz dual-core processor with 8GB of RAM.

V. Conclusion and Future Work

In this paper, we have presented an approach to model-based 3D visual tracking with the use of Distance Transform, which gives us an estimation of how far or close the virtual projected object is, relative to the real object we want to track. With the use of this estimation, we were able to measure the difference between the extrinsic parameters of the virtual and the real camera. Then we proceeded to the minimization of this difference using the Gauss-Newton algorithm, which led to the estimation of the real camera's pose in real-time, for each frame of a video feed.

One main problem of our method is that it does not handle very well cases when the object we want to track is partially occluded. To overcome this problem, we can explore techniques that find outlier measurements from the image, like RANSAC [10], and thus be able to remove features that are being extracted and that belong to the occluding object. Another approach is to find correspondences between the 2D (non-occluded) projections of the real object and the 2D or 3D control points of the virtual object, so the pose estimation process is based solely on them.

The initial extrinsic parameters of the camera are hand picked, so that the projected virtual object bounds the real one. An investigation on automatic initialization of the initial parameters will also be conducted. Finally, the renderer of our implementation needs to be expanded, so it can handle more complex objects, non-convex objects. To achieve this, the rendering procedure needs to support *z-buffering* techniques.

Although we use a simplistic model (cuboid) for our tracking experiments, our method can be expanded to more complex and non-convex objects, according to the targeted application.

With the aforementioned improvements implemented, it will be possible to compare our method with techniques like [13] or [14]. To the best of our knowledge, there is not any public source code for these two techniques, so their implementation is left for future work.

References

[1] J. Barandiaran and D. Borro, "Edge-based markerless 3d tracking of rigid objects," in 17th International Conference on Artificial Reality and Telexistence (ICAT 2007), Esbjerg, Jylland, Denmark, Nov 2007, pp. 282–283.

[2] M. L. et al., "Fast object localization and pose estimation in heavy clutter for robotic bin picking. the international journal of robotics research, 31(8), 951-973," International Journal of Robotic Research - IJRR, vol. 31, 07 2012, pp. 951–973.

[3] D. Berenson and S. S. Srinivasa, "Grasp synthesis in cluttered environments for dexterous hands," in Humanoids 2008 - 8th IEEE-RAS International Conference on Humanoid Robots, Daejeon, South Korea, Dec 2008, pp. 189–196.

[4] E. Marchand, H. Uchiyama, and F. Spindler, "Pose estimation for augmented reality: A hands-on survey," IEEE Transactions on Visualization and Computer Graphics, vol. 22, 01 2016, pp. 2633–2651.

[5] V. Lepetit and P. Fua, "Monocular model-based 3d tracking of rigid objects," Found. Trends. Comput. Graph. Vis., vol. 1, no. 1, Jan. 2005, pp. 1–89. [Online]. Available: http://dx.doi.org/10.1561/0600000001

[6] W. A. Hoff, K. Nguyen, and T. Lyon, "Computer vision-based registration techniques for augmented reality," in Intelligent Robots and Computer Vision XV, 1996, pp. 538–548.

[7] D. K. et al., "Real-time vision-based camera tracking for augmented reality applications," in Proceedings of the ACM Symposium on Virtual Reality Software and Technology, ser. VRST '97. Lausanne, Switzerland: ACM, 1997, pp. 87–94. [Online]. Available: http://doi.acm.org/10.1145/261135.261152

[8] I. Skrypnyk and D. G. Lowe, "Scene modelling, recognition and tracking with invariant image features," in Third IEEE and ACM International Symposium on Mixed and Augmented Reality, Arlington, VA, USA, Nov 2004, pp. 110–119.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, Nov 2004, pp. 91–110. [Online]. Available: https://doi.org/10.1023/B:VISI.0000029664.99615.94

[10] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM, vol. 24, no. 6, Jun. 1981, pp. 381–395. [Online]. Available: http://doi.acm.org/10.1145/358669.358692

[11] W. L. et al., "SSD: Single shot multibox detector," in European Conference on Computer Vision (ECCV), vol. 9905, Amsterdam, The Netherlands, 10 2016, pp. 21–37.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016, pp. 779–788.

 ISBN: 978-1-61208-724-5

[13] C. Harris and C. Stennett, “RAPiD - a video rate object tracker,” in In Proc. British Machine Vision Conference BMVC ’90, Oxford, UK, 1990, pp. 73–78.

[14] T. Drummond and R. Cipolla, “Real-time tracking of complex structures with on-line camera calibration,” in In Proc. British Machine Vision Conference BMVC ’99, Nottingham, UK, 1999, pp. 574–583.

[15] A. Fitzgibbon, “Robust registration of 2d and 3d point sets,” Image and Vision Computing, vol. 21, 04 2002, pp. 1145–1153.

[16] D. Huttenlocher, “Cs664 computer vision - 7. distance transforms,” 2008, [retrieved: May, 2019]. [Online]. Available: https://www.cs.cornell.edu/courses/cs664/2008sp/handouts/cs664-7-dtrans.pdf

[17] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, Nov 2000, pp. 1330–1334.

[18] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, “Bundle adjustment - a modern synthesis,” in Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, ser. ICCV ’99. London, UK, UK: Springer-Verlag, 2000, pp. 298–372. [Online]. Available: http://dl.acm.org/citation.cfm?id=646271.685629

[19] J. Canny, “A computational approach to edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, Nov 1986, pp. 679–698.

 ISBN: 978-1-61208-724-5

The Role of Complexity in Visual Perception: Some Results and Perspectives

Vittoria Bruni
Dept. of SBAI
University of Rome La Sapienza
Rome, Italy
e-mail: vittoria.bruni@uniroma1.it

Domenico Vitulano
Institute for Calculus Applications (IAC)
CNR
Rome, Italy
e-mail: d.vitulano@iac.cnr.it

***Abstract*— In the last years, there has been a change of perspective in approaching some information processing problems: from functional to human perception perspective. In this work, some information theoretic concepts concerning information complexity will be revised and used as formal model for human visual perception based approaches that have been used for solving some image and video processing problems, as for example restoration, detection, tracking and visual quality assessment.**

Keywords-Kolmogorov complexity; asymptotic equipartition property;just noticeable detection threshold; human perception.

I. INTRODUCTION

The vision process has much in common with compression and more in general with the concept of information complexity. The idea behind compression is to reduce the statistical redundancy of the data, for example for storage purposes. In particular, the compressor is an algorithm whose aim is to produce a string of bits whose length is less than the string of bits that are necessary for representing the original data. On the other hand, the Kolmogorov complexity [1] of a signal is the length of the shortest computer program that outputs it. Several neurological studies proved that in the observation of a scene (early vision) very few points are used by human observers to understand scene content [2]. These points are foveated, i.e., human attention decreases as one moves away from them, and they are the ones showing independence between local luminance mean and contrast. In other words, few fixation points are necessary to code scene information and to understand (learn) its content. Vision process can be then modeled as an encoder/decoder system, where human eye is the decoder (final receiver of image information) or the transmission channel. As a result, information theory concepts can be used for coding visual information in order to

- design algorithms for objective evaluation of quality in a way that is consistent with subjective human evaluation;
- develop automatic algorithms adapted for optimal perceptual quality.

This strategy provides a new perspective on image content representation where image pixels are no longer seen as simple probabilistic data but should be accounted for by Human Visual System (HVS) limits and rules.

Next section provides some examples while the last section draws the conclusions.

II. SOME NOTES ON VISUAL COMPLEXITY

Human perception can guide digital restoration according to a new paradigm: reducing the visual contrast of image anomalies till they are masked by surrounding information. As a result, according to the task, image degradation is not removed but its contribution is hidden in the image according to the visual contrast masking effect; in other words, it represents negligible information and does not contribute to signal complexity/quality [3]. Figure 1 shows an example: noise is equally distributed in the image but it is not perceived in the textured region, as correctly measured by the Structural SIMilarity index (SSIM); therefore, denoising can be applied only in regions where noise is visible in order to avoid the introduction of visible artifacts due to oversmoothing or misalignments --- the latter are caused by a not correct motion estimation in case of video denoising. The same concepts have been successfully applied for tuning and balancing quantization errors in image compression.

In this context, looking at the scene inspection as a random walk, image anomalies capture human eye attention at first sight as they are perceived as foreign objects in the scene, independently of scene complexity. As a result, the automatic detection of image anomalies, as well as a moving target, is allowed by looking at them as those resulting visually different from the remaining image content. In other words, the anomaly represents a “surprise” and then the code of visualized information increases whenever the “surprise” occurs. More in general, human perception offers new ways for the representation of image content; for example, in a hierarchical way: from the most visible to the less one (saliency maps); or using perception-based sampling rules (fixation points). It seems based on a measure of object complexity in the sense of Kolmogorov complexity, i.e., the probability of finding that object in the nature. It is obvious that this kind of interpretation generalizes the common concept of surprise which is measured by the Shannon entropy. Therefore, the challenge is to define novel mathematical tools able to directly account for these concepts. The minimum description length represents a useful tool in this sense as it provides a formal equivalence between coding and learning [4]. It is based on the Occam razor concept [5] for which the simplest solution is more likely to be the best, not necessarily the real one. However, some refinements are required as it represents, in some

 ISBN: 978-1-61208-724-5

sense, a computable interpretation of Kolmogorov complexity which, in contrast, is not computable even though more close to the real scenario. Classical information theory concepts can be then used not only for coding but also for learning: simple and few information is enough for describing a more complex one. In this setting, feature extraction is a way of coding and if the characteristics of the decoder (human eye) are embedded in the model, feature extraction becomes a way of coding based on human vision. An example is shown in Figure 2 where fixation points are used for assessing image quality, or where human perception is used for segmenting dermoscopy images.

III. Conclusions and Future Work

The simulation of human visual system represents the key issue for developing novel and effective solutions for some classical visual information-based processing problems. In this paper, the double role of complexity, in the sense of information compression, has been discussed. On the one hand, complexity provides a theoretical tool useful for formal modeling. Conversely, the mechanisms regulating human perception can offer the way of defining novel paradigms for information coding. The goal of future research is then to provide formal models and mathematical methods able to represent and process such visual information. In particular, since compact representation is the key issue in several applications, the definition of a visual perception-inspired multiscale transform represents an interesting and promising challenge.

References

[1] R. Cilibrasi and P. M. B. Vitányi, "Clustering by compression," IEEE Transactions onInformation Theory, vol. 51 no. 4, pp. 1523-1545, 2005.

[2] R. A. Frazor and W. S. Geisler, "Local luminance and contrast in natural images," Vision Research, vol. 46, pp. 1585-1598, 2006.

[3] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," IEEE Trans. Image Process., vol. 14, no. 12, pp. 2117–2128, 2005

[4] V. Bruni, E. Rossi, and D. Vitulano, "On the equivalence between Jensen-Shannon divergence and Michelson contrast," IEEE Trans. on Information Theory, vol. 58, no. 7, pp. 4278-4288, 2012

[5] P. D. Grünwald, "A tutorial introduction to the minimum description length principle," Advances in Minimum Description Length: Theory and Applications, I. J. Myung, P. D. Grünwald, and M. A. Pitt, Eds. Cambridge, MA: MIT Press, 2004.

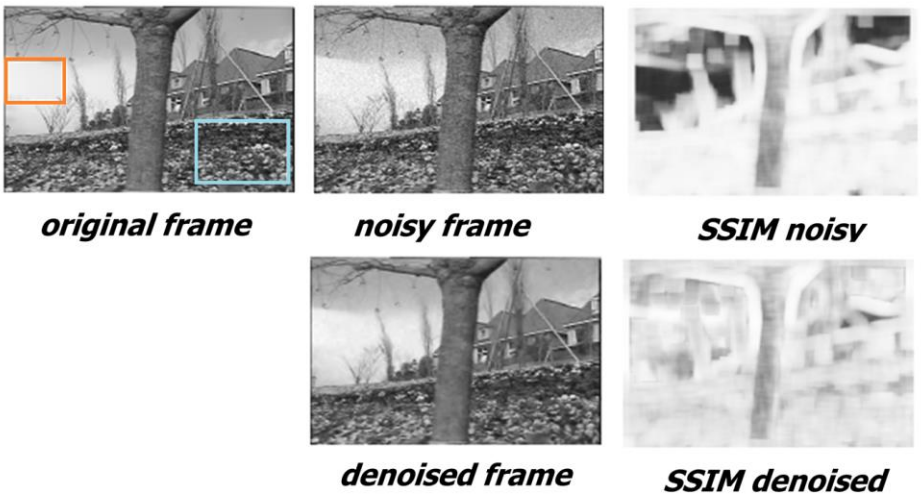


Figure 1. Top) Noise is differently perceived in the image: more visible in the red box than in the cyan box. Accordingly, SSIM in the red box is darker than in the cyan box. Bottom) Denoising applied to the noisy image increases the visual quality in the red box but decreases the one in the cyan box. Accordingly, SSIM of the denoised image in the cyan box is darker than the one of the noisy one in the same box.

 ISBN: 978-1-61208-724-5

Figure 2. Left) Blurred image. Middle) Red boxes are the points necessary for assessing image quality in a way which is consistent with HVS. Right) Dermoscopic image and its human perception based-quantized version using the least number of quantization bins.

 ISBN: 978-1-61208-724-5

VMPepper: How to Use a Social Humanoid Robot for Interactive Voice Messaging

Paola Barra
Dept. of Computer Science
University of Salerno
Salerno, Italy
email:pbarra@unisa.it

Carmen Bisogni
Dept. of Computer Science
University of Salerno
Salerno, Italy
email:cbisogni@unisa.it

Riccardo Distasi
Dept. of Computer Science
University of Salerno
Salerno, Italy
email:ricdis@unisa.it

Antonio Rapuano
Dept. of Computer Science
University of Salerno
Salerno, Italy
email:arapuano@unisa.it

***Abstract*—VMPepper is an encrypted voice mail system managed by Pepper, the humanoid robot. Pepper ships with face recognition software that has been used in voice mail system design. The robot has a voice recording module as well. A human that has already been recognized can record or listen to messages to or from another user. Sensitive data such the audio message itself, sender and recipient are encrypted during transmission. A set of users has been selected to evaluate the system. During the experiments, Pepper could move freely, initiating interaction on recognizing a user. The interactions completed successfully in 100% of the cases.**

***Keywords*—** *Humanoid Robot; Pepper; Face Recognition; Voice Mail; Cryptography; Cloud.*

I. Introduction

Humanoid robots are evolving and becoming increasingly simple to use and program. At the same time, sensors are getting less and less invasive. This progress makes it possible to design helper applications that interact with humans in a more casual and relaxed way. The particular humanoid robot used in this project is Pepper, designed and developed by SoftBank Robotics Corp. and Aldebaran Robotics SAS [2]–[4]. It is depicted in Figure 1.

Pepper is not the first robot produced by Robotic SoftBank: there is a direct predecessor named NAO, which also runs the NAOqi operating system. The simplest way to develop custom applications is the box programming environment Choreographe, available on all of Aldebaran's products. Snippets of code are pasted into text boxes that get activated under specific circumstances. Choreographe's SDK can interface with several programming languages; the choice for this project was Python 2.7.

Both Pepper and NAO are suitable for interaction with humans [5], [17]. In particular, Pepper presents itself with a childlike appearance due to its height of 120 cm, its large eyes and other soft facial features. This eases humans into a more spontaneous and cooperative interaction. Pepper, in fact, has already been used for experiments and shown to facilitate people's existence, as reported in [1]. Pepper's main specifications are shown in Table I.

The paper is organized as follows. The next section outlines the state of the art regarding Pepper and other humanoid robots used in a variety of applications, as well as non-humanoid robots specifically offering answering machine services. Section III describes the method used for the present proposal, while Section IV illustrates the experiments involving interaction with the robot. Finally, Section V presents our conclusions and possible future developments.

II. Previous Work

Previous research work using Pepper and other humanoid robots has been centered on three main areas: interaction with the environment, medical applications, and social interaction including education. In all three cases, the sensor component is the main driver

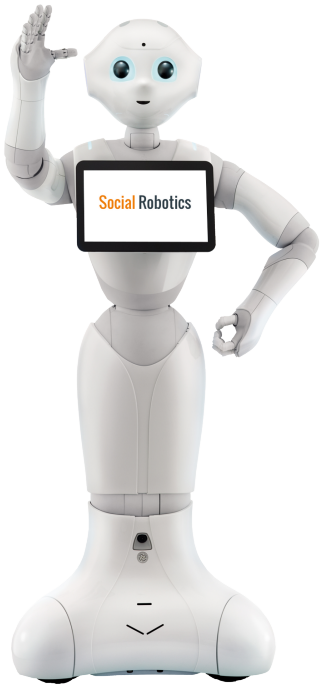


Fig. 1: The humanoid robot Pepper

TABLE I: SPECIFICATIONS

Hardware and Connections	Details
Size (H x D x W)	121 x 425 x 485 [mm]
Weight	28kg
Battery	Li-ion 30.0Ah / 795Wh
	3D sensor × 1, touch sensor × 3
Sensors (trunk)	Gyroscope sensor × 1
Sensors (hand)	Touch sensor × 2
Sensors (leg)	Ultrasonic sensor × 2, laser sensor × 6, bumper sensor × 3, gyroscope sensor × 1
DOF	20
Display	10.1 inch touch screen
OS	NAOqi OS
Network	Wireless and wired interfaces
Speed	Max. 3 km/h

 ISBN: 978-1-61208-724-5

for innovative use.

1) Interaction with the environment: Allowing Pepper to walk inside a closed building is a challenge because GPS sensors cannot be used. The most natural solution is try to make the most of the information provided by other sensors such as cameras and proximity sensors. In [6], a 3D map is created so the robot is able to move independently, avoiding obstacles. In [7], the indoor trajectory that the robot will follow is calculated a priori. These papers show that it is possible to create "intelligent space" with just visual sensors. On the other hand, the content in [8] and [9] covers more technical details, in particular a study of the maximum inclination that the robot motors can withstand. Management of physical contacts with the robot is also investigated through the use of proprioceptive sensors.

2) Medical care: Humanoid robots can be useful in the medical field, too: for example to manage anxious patients as shown in [10]. Another possible application is as a medical assistant: the robot is used for daily patient data collection [11], or to help patients respect their prescription drug schedule [12]. By charging the robot with work previously entrusted to man, human staff gain time while patients manage to deal with robots in the simplest tasks.

3) Social interaction and education: Social interaction with Pepper is widely used to offer services and entertainment. When using Pepper in human interaction, it is necessary to use sensors to collect feedback and human emotions. A robot programmed for social entertainment typically interacts based on its own perception of human emotions, and can respond to such emotional stimuli by showing emotions itself, as shown in [15]. In other experiments, a similar approach was followed to interact with children, and Pepper turned out to be an effective interactive educator [16].

Applications closer to our voice mail system are provided by well known commercial voice assistants such as Alexa, Siri, Cortana, and the like. A detailed description of the services offered by these systems can be found in [18]. The interaction in these cases is exclusively vocal, so a point by point comparison with Pepper is not possible. However, the services offered by a typical voice assistant are quite similar: making phone calls and sending or reading text messages and emails. The hardware in these commercial systems does not include a range of sensors as wide as Pepper's, so face recognition is out of the question. On the other hand, the variety of motors, motion sensors and cameras offered by Pepper enable free range movement and face recognition to be an integral part of the services offered.

III. Method

The proposed system is an interactive service. The general design goal is that the users should be able to interact with the robot as if it were an intelligent answering machine. After recognizing the user, the robot records a message for a specific recipient, who should also be registered with the system; when the robot meets a recipient in its pending message list, it asks them if they want to listen to the message or record a new one. Data privacy is based on facial recognition. In fact, facial recognition is performed twice: when a user approaches the robot for the first time, and when a message is to be recorded.

Voice Mail Services offered by the robot are restricted to authorized users. Therefore, a registration phase is required. The registration process needs help from a human operator. The operator starts recording face features with the "Learn Face" box of Choreographe. The extracted features are stored in robot memory, and they are recalled during the face recognition step, an example of which is illustrated in Figure 2.

A. Interacting with Pepper

Users that want to interact with the robot must approach it in order to get recognized and therefore authenticated as a registered user. When the robot recognizes a user, it acknowledges them by saying "Hello *Name*", as shown in Figure 3.

Fig. 2: Pepper learns a face

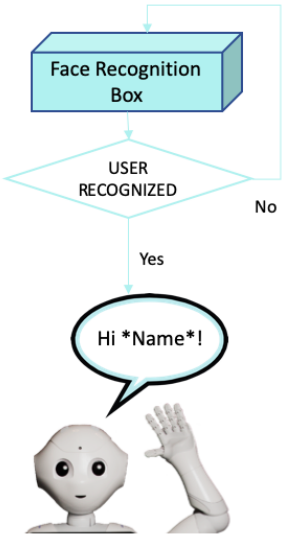


Fig. 3: Pepper recognizes a user

The user can access the voice mail services by uttering phrases with keywords "Leave" or "Listen".

If a phrase with the keyword "Leave" is pronounced, the interaction proceeds as follows.

1) The robot asks "Who is the message recipient?".
2) If the name the user says is not stored in memory, the robot will reply "Im sorry, your friend is not registered".
3) Otherwise, the recipient is well defined, so Pepper notifies the sender that recording is starting.
4) Once recording is done, Pepper tells the sender that the message will be delivered—that is, replayed—to the recipient as soon as the occasion arises—that is, as soon as Pepper meets the recipient and completes facial recognition.
5) The process returns to the face recognition step.

If a phrase with the keyword "Listen" is pronounced, the interaction proceeds as follows.

1) The robot asks the user "Who is the message sender?".
2) If there is no match between sender and recipient in the stored data, the robot says "Sorry, there are no messages for you from him/her".
3) If there is a match, Pepper performs face recognition again. The face recognition at this step is to enforce basic privacy/ security: if the would-be recipient moves away from Pepper,

 ISBN: 978-1-61208-724-5

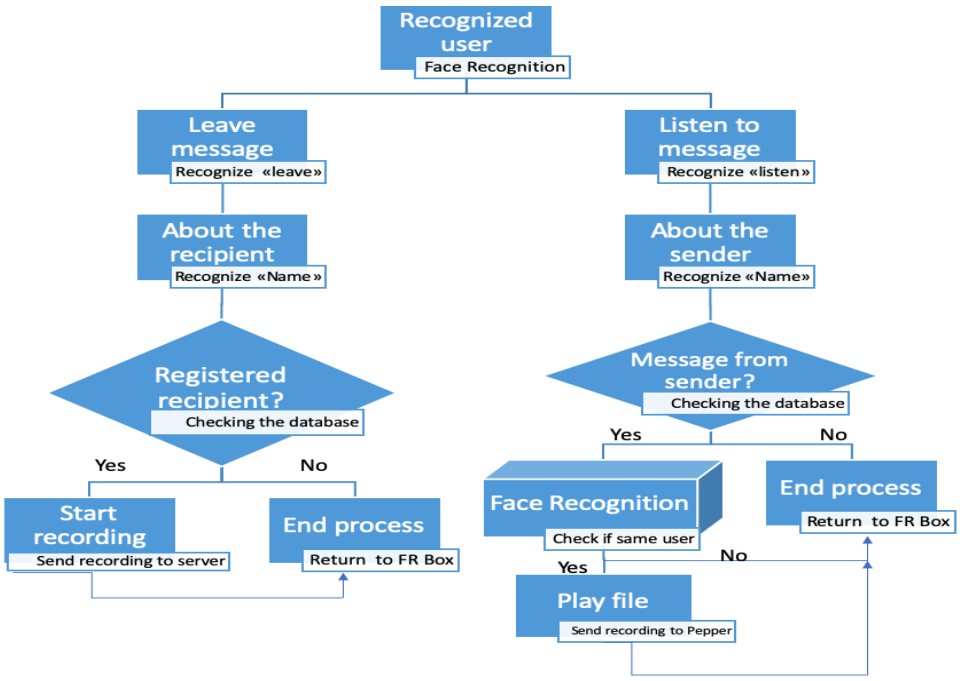


Fig. 4: Proposed workflow

or if some other user tries to take over the interaction, the message will not be played.

4) If face recognition succeeds, the message is played.
5) The process returns to the face recognition step.

Both procedures are visually summarized in Figure 4.

B. Pepper-Server data flow

The voice mail method system is made by two subsystems, both of which rely on NAOqi's library to exploit Pepper's capabilities. A remote server is used to perform part of the tasks.

- The *Record* subsystem, related to the "Leave a message" module, stored and executed on Pepper.
- The *Replay* subsystem, which performs audio file upload and runs the "Listen to a message" module, stored and executed on the server.

The first subsystem runs the face recognition module. After the interacting user has been recognized, Pepper will listen for the keywords "Leave a message" or "Listen to a message".

When the "Leave a message" module is activated, the Record subsystem, running locally on Pepper's operating system, asks for the recipient of the message. If the recipient is recognized as a registered user, a voice message is recorded. After that, the Record subsystem sends an encrypted HTTP request to the Replay subsystem, running remotely. The parameters of this HTTP request are the sender's ID, the recipient's ID, and the audio file.

When the request is processed successfully, the Record system will delete the audio file from Pepper's local memory. The Replay subsystem runs on the server and listens for HTTP requests. As soon as a request arrives, it stores the audio file on a cloud storage service (Google Drive in the first prototype) and adds a new item into an associative array. The item contains the sender's ID, the recipient's ID, and a link to the audio file on cloud storage.

When Pepper recognizes the "Listen" keyword, it sends a message to the Replay subsystem, which in turn checks for the listener's and the sender's registration. If both are in the list, Pepper performs face recognition on the listener. Finally, Pepper's "open WebView" module is run with the cloud storage audio link as a parameter, so the audio is actually played out. Processes are shown in Figure 5.

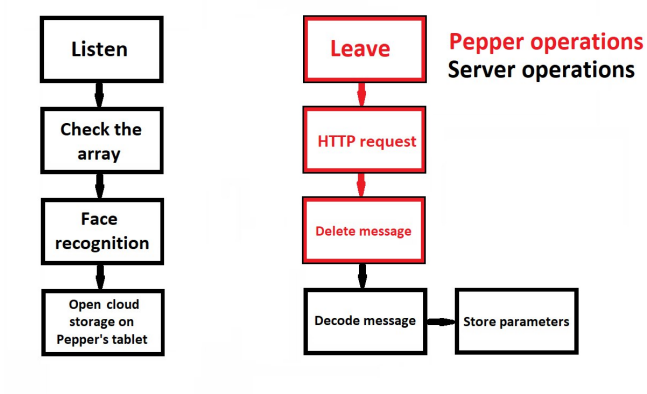


Fig. 5: Pepper-Server local-remote data flow

IV. Experiments

In human-robot interaction, the qualitative indexes try to assess the user's feelings during the interaction, as well as the robot's ability to conclude the interaction according to user expectations.

Pepper has an operating parameter "autonomous life" that can be switched on or off. The parameter dictates if the robot can react to external stimuli besides those managed by its current custom programmed behavior. Therefore, when autonomous life is off the robot cannot be distracted by sounds in the background—or even by its own limb movements. When autonomous life is on, the robot reacts to the external stimuli according to its factory settings. Therefore, with autonomous life on, the robot is able to reply to user questions, turn its head to follow a locked-in user face, or turn its body to point the camera in a suitable direction: as an example, if

 ISBN: 978-1-61208-724-5

a sound is heard, Pepper will try to aim the camera to the apparent source of the sound.

Leaving autonomous life set to on means that Pepper's behavior is more natural. The autonomous movements make for better and more fluid interactivity. However, there is the tradeoff of external interference that could make interactions harder to program and to carry out in practice. It is not even necessary to have a malicious agent trying to disrupt operation: random external events might be quite enough.

For this reason, two sets of experiments were performed, with autonomous life respectively off and on.

A. Autonomous life off

The user pool consisted of five people of age between 23 and 30. Their faces were recognized as described at the beginning of Section III, and they were added to the registered user set. They were provided a brief verbal explanation about Pepper's services and voice commands.

Each user in this block of experiment interacted with Pepper 10 times, for a total of fifty social interactions. We set a maximum message duration of 10 seconds, so that they could leave meaningful messages without having to wait for longer timeouts and without the need to program an "end message" aural signal.

The results are quite encouraging. Pepper carried out face recognition and correctly recorded and forwarded the message in 100% of cases. Furthermore, the users found the workflow of method very natural and friendly.

B. Autonomous life on

With autonomous life turned on, the experiments followed a different protocol.

- We set the robot free to move in a room in any direction, performing all possible translations and rotations.
- When the robot sees a user that it can recognize, it stops.
- It says hello to the user and starts the flow as described above.

In this operating mode, we found that it is difficult for Pepper to recognize users if they are moving. This was expected. However, if the users exploit Pepper's autonomous life to divert its attention to themselves by making suitable sounds, the robot turns its head and recognition happens easily. The flow then proceeds in the same way as in the "autonomous life off" mode.

V. Conclusions and Future Work

We observed that the use of a humanoid social robot as a voice messenger is quite natural and enjoyable for the user. The ability of a humanoid robot to move autonomously in large spaces makes this application useful for a large population of users. Interaction is efficient and without significant glitches, particularly with the face recognition box and message recording. This may enable extended applications, particularly in combination with other services. As an example, if a mapping of the entire building or complex is available, the robot may be able to bring the message to the recipients directly rather than waiting to meet them.

References

[1] Amit Kumar Pandey and Rodolphe Gelin, Pepper: "A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind", IEEE Robotics & Automation Magazine, Volume: 25 , Issue: 3 , Sept. 2018, pp. 40 - 48.
DOI: 10.1109/MRA.2018.2833157

[2] CNN. Meet Pepper, the emotional robot. Retrieved February 24, 2015, from https://edition.cnn.com/2014/06/06/tech/innovation/pepper-robot-emotions/index.html

[3] TIME. Meet Pepper, the Robot Who Can Read Your Emotions. Retrieved February 24, 2015, from https://time.com/2845040/robotemotions-pepper-softbank/

[4] IEEE SPECTRUM. How Aldebaran Robotics Built Its Friendly Humanoid Robot, Pepper. Retrieved February 24, 2015, from https://spectrum.ieee.org/robotics/home-robots/how-aldebaran-robotics-built-its-friendly-humanoid-robot-pepper

[5] Arkadiusz Gardecki and Michal Podpora, "Experience from the operation of the Pepper humanoid robots", 2017 Progress in Applied Electrical Engineering (PAEE), 15 August 2017.
DOI: 10.1109/PAEE.2017.8008994

[6] Eiji Kaneko and Nobuyuki Umezu, "Rapid Construction of Coarse Indoor Map for Mobile Robots", 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE 2017), 21 December 2017.
DOI: 10.1109/GCCE.2017.8229367

[7] Dorota Belanová, Marián Mach, Peter Sinčák and Kaori Yoshida, "Path Planning on Robot Based on D* Lite Algorithm", 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), August 2018.
DOI: 10.1109/DISA.2018.8490605

[8] Jory Lafaye, Cyrille Collette and Pierre-Brice Wieber, "Model predictive control for tilt recovery of an omnidirectional wheeled humanoid robot", 2015 IEEE International Conference on Robotics and Automation (ICRA), 02 July 2015.
DOI: 10.1109/ICRA.2015.7139914

[9] Anastasia Bolotnikova, Sbastien Courtois and Abderrahmane Kheddar, "Contact Observer for Humanoid Robot Pepper based on Tracking Joint Position Discrepancies", 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 08 November 2018.
DOI: 10.1109/ROMAN.2018.8525774

[10] Sachie Yamada, Tatsuya Nomura and Takayuki Kanda, "Healthcare Support by a Humanoid Robot", 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 25 March 2019.
DOI: 10.1109/HRI.2019.8673072

[11] Daisy van der Putte, Roel Boumans, Mark Neerincx, Marcel Olde Rikkert and Marleen de Mul, "A Social Robot for Autonomous Health Data Acquisition Among Hospitalized Patients: An Exploratory Field Study", 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 25 March 2019.
DOI: 10.1109/HRI.2019.8673280

[12] Keitaro Ishiguro, Saki Minamino, Jun Kawahara and Yukie Majima, "Development of a Robot Intervention Program in Medication Instruction at a Pharmacy", 2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI), 18 April 2019.
DOI: 10.1109/IIAI-AAI.2018.00198

[13] Chiao-Yu Yang, Ming-Jen Lu, Shih-Huan Tseng and Li-Chen Fu, "A companion robot for daily care of elders based on homeostasis", 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 13 November 2017.
DOI: 10.23919/SICE.2017.8105748

[14] Thi Le Quyen Dang, Nguyen Tan Viet Tuyen, Sungmoon Jeong and Nak Young Chong, "Encoding cultures in robot emotion representation", 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 14 December 2017.
DOI: 10.1109/ROMAN.2017.8172356

[15] Wen-Feng Shih, Keitaro Naruse and Shih-Hung Wu, Implement human-robot interaction via robot's emotion model, 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), 15 January 2018.
DOI: 10.1109/ICAwST.2017.8256522

[16] Fumihide Tanaka, Kyosuke Isshiki, Fumiki Takahashi, Manabu Uekusa, Rumiko Sei and Kaname Hayashi, "Pepper learns together with children: Development of an educational application", 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), 28 December 2015.
DOI: 10.1109/HUMANOIDS.2015.7363546

[17] Softbank Robotics Documentation Retrieved 26 October 2018 from http://doc.aldebaran.com/2-5/index.html

[18] Matthew B. Hoy, "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants", Medical Reference Services Quarterly, Volume 37, 2018 - Issue 1, pp 81-88.
DOI:10.1080/02763869.2018.1404391

 ISBN: 978-1-61208-724-5