Introduction - chair

Panel on Human-Machine cohabitation
Theme: Cognitive Humane-Machine Interaction at Work

Chair: Marie Sjölinder, RISE - Research Institute of Sweden, Sweden

Panelists:
Ryosuke Yamanishi, Kansai University, Japan: **Content-oriented Interaction: what is the specific features of the content we tackle**

Oky Dicky Ardiansyah Prima, Iwate Prefectural University, Japan : **3D Data Visualization Using a Spherical Display**

Yushan Pan, Norwegian University of Science and Technology, Norway: **The Principal Benefits Of Qualitative Research For Humane-machine Interaction At Work**

Marie Sjölinder, RISE, Sweden: **Human perception in VR-environments and in real life**

Focus of the panel

- Relationship between objective and subjective, real and perceived
- Representations and interpretation of data – humas and machines, qualitative and quantitative
- Interplay between humans and machines – data and representations

Topics for discussion

- Possibilities and limitations
- What can and cannot be represented by machines or a mathematical approach
- Future steps and possibilities

IARIA

Panel 4
Panel on ACHI
Theme: Cognitive Human-Machine Interaction at Work
(brain cognition, visual interaction, emotional communication, human-agents, accessibility, virtual reality, interest profiling, ergonomic design, digital assistants, etc.

Digital World 2020

Panelist Position

Content-oriented Interaction – what is the specific features of the content we tackle?

Ryosuke Yamanishi, Kansai University, Japan ryama@kansai-u.ac.jp

- Human naturally handles every content considering its features
- Generic interaction model should be useful though, the content is just a "data”
- Misunderstandings between human and the machine might happen because of emotion and cognition for the content
- Sometimes, even in human-human interaction at work, misunderstanding should raise almost all problems

→ The interaction should be designed for machine to catch the content features

→ Both human and the machine should see almost the same point of the content, which is not designed especially for each cognition of human and machine

→ Content-oriented interaction between human and machine

Panel 4
Panel on ACHI
Theme: Cognitive Humane-Machine Interaction at Work
(brain cognition, visual interaction, emotional communication, human-agents, accessibility, virtual reality, interest profiling, ergonomic design, digital assistants, etc.

Content-oriented Interaction:
what is the specific features of the content we tackle

Ryosuke “Leo” Yamanishi, Ph. D
Associate Professor at
Faculty of Informatics, Kansai University, Japan

ryama@kansai-u.ac.jp

THE OUTLINE

1. Definition the problems in interactions
   - How humans and machines respectively work
   - The difference between humans and machines
2. Reducing misunderstandings is a key
   - Features, points, and a kind of mind should be shared
   - Content-oriented approach
3. The example: our content-oriented approach
   - “Translation” is the target content

Interaction in humans and machines at work

What can be a problem in interactions?

- Misunderstanding between interaction partners may lead to almost all the problems even in human-human interactions
- Consensus should be formalized based on the same understandings
- At work, the interaction without any frustrations should improve our efficiency
  - The frustration is a kind of most frequent emotions we have at work

Humans and machines at work

- Nowadays we can NOT do work without machines
- Human-Machine interactions at work should be designed to be based on the same understandings as well as human-human interactions

How humans handle the contents?

- The picture of my lovely dog is more precious than others!
  - What is shown in the picture?
    - IT is a dog? No, HE is lovely "Kaito"
      - For me and my friends, the object in the picture is recognized as not just a golden retriever
  - The recognition depends on what we know and feel for the objects
- Even if almost the same pictures of him in my phone, I NEVER remove those
  - But, for the pictures of other dogs, I would probably remove the similar pictures

How the machine works in the general interaction model

Do the machines really watch pictures?

- Some models based on machine learning methods recognize the objects in the picture though…
  - The machines may not be able to distinguish the dog in the picture as "Kaito" even if it learns a lot of pictures of "Kaito" and other dogs

Just the distribution of valuables concerning the objects is learned with the annotated labels

- Even the features of the content we tackle should be different for each content
  - Feelings, memories and other emotions are not taken in the general learning
- Sometimes, we need not only image features to recognize the visual contents

Misunderstandings between humans and machines

Humans may overtrust the current AI performance

- Some of them say "AI can understand what we feel"
  - It is partially correct, but not entirely correct
  - If the machines behave wrongly against how humans expected, humans may be strongly disappointed
- Humans can not know how the AI works and understands the data in the process
  - The attention for the data should be different between humans and machines

This problem may happen even in human-human communication too

- Most problems in the communication should be based on the misunderstandings between the partners

How to resolve the misunderstandings

Imperfect information should be reduced as much as possible

- Consumpting what we feel and think is one of the ways to prevent the misunderstandings
  - What is focused on should be clear
- The features we see are sometimes different for each individual, culture, field, and background.

What is important in the content should be taken into consideration

- What can be the features should depend on the content
- The more misunderstandings are reduced,
  the more the interaction becomes efficient

Content-oriented approach

- Content and its surroundings are the keys to make the interaction effectively and reduce the misunderstandings in the interaction
  - In human-machine interaction, we humans have to know how the machine feels for the content
  - Even in human-human interaction, the content-oriented mind is effective
- Now, the content-oriented minds, methods, and approaches should be important more than before
  - We have to catch the emotion or something like that towards content remotely
    - We can not vaguely understand the partners' cognition; it is hard to use nonverbal information such as facial expressions and gestures

Our target content: “translation”

The example we have already tackled with the content-oriented approach: the translation

- The translation has become more and more typical example in the human-machine interaction at work
- Some translation systems show a good result in communication though, something is different from a human translator

How humans behave with a human translator?

- We may change our expressions or words by reflecting on the translator’s response
  - We may not repeat the exact same script that the translator can not understand

What can be the features of humans and machines in the translation?

In case humans use the machine translator, we sometimes repeat the exact same script

- Even though the speech is not recognized well
- Even though they can guess the script is wrongly translated

They behave as if our script is perfect for the translation

What can the machine show like human translators?

- Usually, the machine shows only the results of the recognition and the translation
- It can show the likelihood for the speech recognition and the translation
  - Those are like the facial expressions or asking back from the translator in the human-human interaction

Our approach: letting us know what the machine feels

- In the content "translation," not only the result but also the process should be shown to the users
  - The demo is here: https://youtu.be/u4H7jakiur0

The collaborative work by J. White, R. Yamanishi, K. Matsumura, and Y. Nishihara

- The feedback helps humans to change and re-think the word and expressions like we ask human translators
  - The speech input to the system is evaluated inside
  - The face is changed based on the confidence of recognition and translation
  - Humans may be able to get more appropriate words and expressions for translation
    - How to use the machine translators should be acquired
    - Human would behave like those which are for human translators

Conclusion

The same understanding is so important in human-human and human-machine interaction

- How the individual feels and thinks should be shared
- How the machine works should be presented to the users

How to reduce the misunderstandings: content-oriented approach

- The content-oriented approach helps us to understand what is important based on the same "mind"
- For efficient interaction at work, we have to model what we should provide through the interaction

Panellist Position

The Principal Benefits Of Qualitative Research For Humane-machine Interaction At Work

Yushan Pan, Norwegian University of Science and Technology, yushan.pan@ntnu.no

- Computer-supported cooperative work
- Social computing
- Design science in information systems

Panellist Position

The Principal Benefits Of Qualitative Research For Humane-machine Interaction At Work

Yushan Pan, Norwegian University of Science and Technology, yushan.pan@ntnu.no

Types of learning

Procedure learning (Card et al., 1986)

1. Cognitive processor is ‘programmed’ with procedural knowledge acquired from learning.
2. At first procedures are declarative knowledge from problem solving (trial and error) and explicit instructions (comprehend instructive material – very common).
3. With practice, converted into procedural knowledge one can routinely executed to achieve goal – a routine skill.
4. With extensive practice, a skill becomes automated – you can perform procedure ‘without thinking’.

Human learning

‘*There exists a qualitative leap in [adults] learning process from the* ***rule-governed*** *user of* ***analytical rationality*** *in* ***beginners*** *to the fluid performance of tacit skills in what Pierre Boudieu (1977) calls virtuosos and Hubert and Stuart Dreyfus (1986) true human experts. We may not that* ***most people are experts*** *in a number of everyday social, technical, and intellectual skills like giving a gift, …while only few reach level of true expertise for more specialized skills like playing chess, … or flying a fighter jet. Common to all experts, … they operate on the basis of* ***intimated knowledge*** *of several thousand concrete cases in* ***their areas of expertise****. … Such knowledge and expertise also lie at the centre of case study [qualitative] as a research and teaching method; …, still: as* ***a method of learning****.* ((Dreyfus and Dreyfus 1986) quoted by *Flyvbjerg 2006)’*

The relationship between experimental work and technology

The development of experimental work has been intricately interwoven with the development of technology (Tiles and Oberdiek 1995)

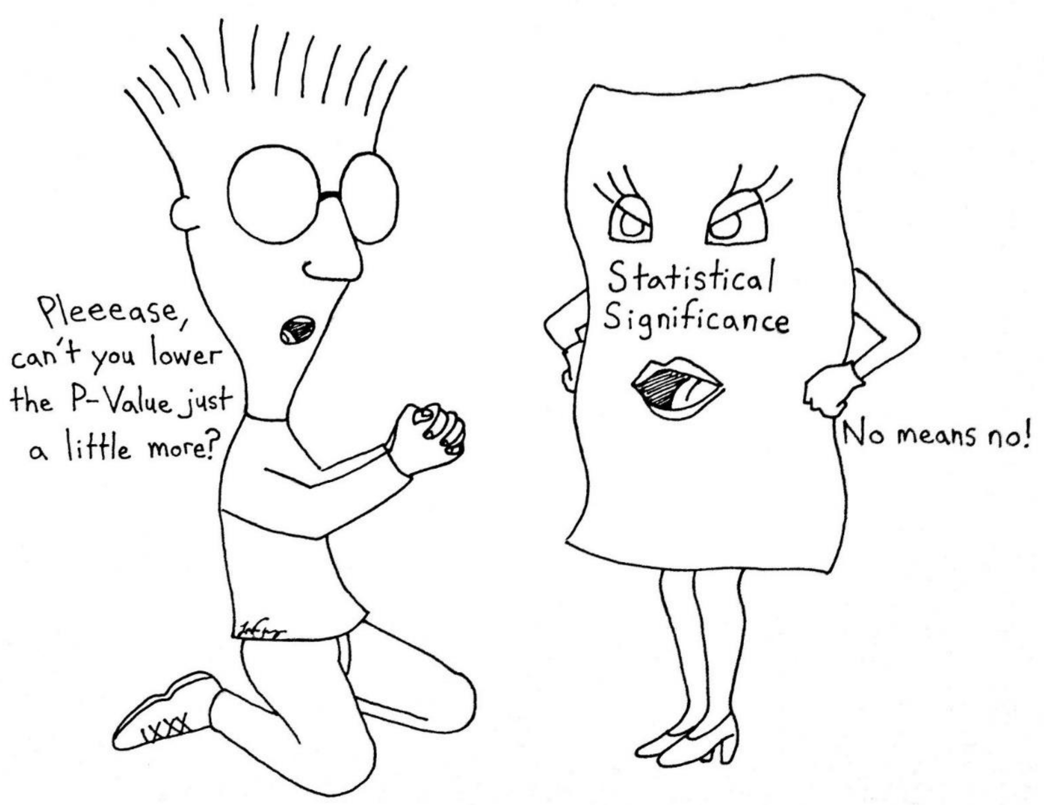
Focus on *instruments and equipment* employed in experimental practice – *nature and function of scientific instrumentation.*

- Investigation of *scientific instruments* is a rich resource of insights for experimental work (Gooding, Pinch and Schaffer 1989, Heidelberger and Steinle 1998).
- Instruments ***represent a property by measuring its value*** (a device that registers pause, heart rate), Instruments create phenomena that do not exist in nature (a laser) (Radder 2009)
- Instruments that closely ***imitate natural processes*** in the lab (***Eye-tracker***).

However, some devices are not usually called instruments, but they are equally crucial to a successful performance of work in workplace

- alarm clock and calculator (Another story from my fieldwork)

The problem with p-values



http://1.bp.blogspot.com/-m1ucuiGYiw0/UNaULkQ0IYI/AAAAAAAAARk/OTDIBDI7Ydo/s1600/photo+(2).JPG

Induction vs. Deduction (p-values)

True probability may depend on *study power* and *bias*, *the number of other studies* on the same question, and, importantly, *the ratio of true to no relationship* among the relationships probed in each scientific field (Ioannidis 2005)

The reason is clear: WE GOT PROBABILITY WRONG! (Colquhoun 2016)

PPV = ([1-β]*R)/ ([1- β]*R+α) (Ioannidis 2005)

Safe analysis in experimental work (HCI and CS)

- *The experiment should put under* ***duress an idea*** *about how the world works to see if the idea is* ***able to predict what will happen****. … An actual experiment must narrow down* ***many things*** *to produce something that is* ***only a test a single aspect of the idea****. No experiment can test the whole idea (Mayo 1996). So recognizing that, it is always better to* ***have several experiments****. There is no need to expect one experiment to deliver everything. Furthermore, if each* ***separate experiment*** *offers* ***clear and unambiguous results*** *then this* ***is a lot better*** *than one big experiment that has a complex design and an even more complex analysis.*
- *(Cairns 2013)*

Dilemmas

1. How do you know what to measure after each experiment?

Construct validity – ***you know what you are measuring*** (Cairns 2013).

But, if I do not know? Where to get this knowledge?

2. **If equally crucial artifacts are overlooked, how do you know they will not influence the experiment?**

Internal validity – if an experiment is set up to ***severely test*** whether influence X really does affect outcome Y then it needs to be clear any systematic changes in Y are wholly due to the change of X (Cairns 2013)

Because I do not know what is missing? How could I know it will influence the experiment?

3. The external and ecological validity

It depends on **the decision of experimental design**. Also, readers will **decide themselves** if the experiment is relevant to the real world.

If the above two can not be answered, then the purpose of experimental work is what?

Turn to the social

Qualitative research is **a process of naturalistic inquiry** that seeks ***in-depth understanding*** of social phenomena within ***their natural setting***.

It focuses on the '***why***' rather than the 'what' of social phenomena and relies on the ***direct experiences*** of human beings as meaning-making agents in ***their every day lives***.

Rather than by logical and statistical procedures, qualitative researchers use ***multiple systems of inquiry*** for the study of ***human phenomena*** including case study, ethnography, grounded theory, phenomenology and so on.

Focus areas are individuals, societies and cultures, and language and communication.

Knowledge is *subjective* rather than objective.

Re-visiting dilemmas - fieldwork

The use of qualitative research that unfolds the importance of cooperative work practices relevance vital to the simulator design

1. How do you know what to measure after each experiment? → But, if I do not know? Where to get this knowledge? → Qualitative technique helps address where is the problem! → Observation and interview help make work practices visible

2. If equally crucial artifacts are overlooked, how do you know they will not influence the experiment? → How to know? → Qualitative results also opens that there are several elements influence human-machine interaction at work

3. The external and ecological validity → The purpose of experimental work → If it is valid for this case, it is valid for all cases (Flyvbjerg 2006)

Presentation of panelist

Human perception in VR-environments and in real life

Marie Sjölinder, RISE – Research Institutes of Sweden, marie.sjolinder@ri.se

- Ph.D psychology, senior researcher
- User interaction
- E-health and welfare technology

Panellist Position

Human perception in VR-environments and in real life

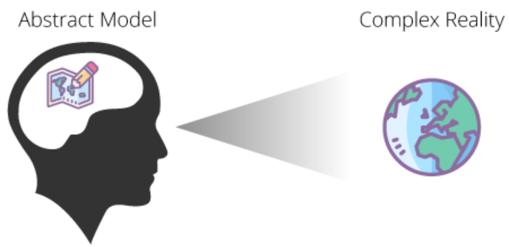
Marie Sjölinder, RISE, marie.sjolinder@ri.se

Differences between VR learning and learning in real life

- VR offers the possibility to create objects and environments that give the illusion of being real, making it possible to interact with objects, move about, and explore environments with a strong sense of presence
- VR does not offer any haptic feedback besides vibration (if not additional devices are used)
- Limitations when it comes to using VR for purposes such as interactive training tasks where the physical performance demands considerations such as weight, task complexity and fine motor movements

Mental models

- Mental models used during interaction with a real environment might not be applicable (Johnson-Laird, 1983; Staggers & Norcio, 1993)
- Scholars have defined what is required of a visual representation in order to be accepted as a representation of an object or milieu (Tversky, 2013; Bae & Watson 2013; Eppler 2013)



Visual representation

- Necessary to consider that a visualization is a representation, where something is presented with a specific intention that might not necessarily be understood by a user in the way intended (Rose, 2016)
- A question of agreement between the visualization and the user, which can be described as "the beholders share" (Gombrich, 1982)
- The users need to contribute with their imagination and must accept the conditions for interaction with a given visual

Sound and vision

- Visuals constitute the largest proportion of all information we collect in order to gain situational awareness. An estimated 40% of the brain's cortex is used to process visual information (Lennie, 1998)
- However, sound is also extremely important, and can often mean the difference between an operational and dysfunctional VR application
- We are more likely to benefit from audio information at a lower level of consciousness, while we become aware of visual events in a more obvious way. Sound puts us on standby, preparing us to act in new situations— before we even know it (Lipscomb, 2013)

Interplay between perceptual cues

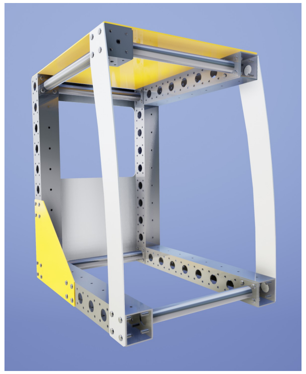
- Interplay between different perceptual cues are well known
- Study by David Katz 1920 where he compared touch with vision and came to the conclusion that, “When we touch some common object, the tactile impression is always permeated with visual experience” (Katz, 1989, p.156)

VR environments and real life experiences

- VR is an altogether visual and aural medium, where it is possible to act and interact with the environment and its objects in ways that other mediums cannot provide
- It is possible to lift, move, push, rotate, and throw imaginary objects, but the only confirmation the user receives that the interaction is taking place is what they can see and hear
- Even though VR is considered to be realistic, it has its limitations and the representations do not totally replace reality
- When conducting tasks in a VR environment, the interplay between different perceptual cues needs to be understood based on the specific roles that apply in the VR environment
- Interaction with objects needs to be adapted to the presuppositions in this environment
- Lack of research when it comes to understanding how mental conventionalization influences users' understanding and use of VR environments in relation to the real environment
- Need to develop a framework for interaction based on human perception, but within a context based on the rules within the VR environment (Rubio-Tamayo et al. 2017)

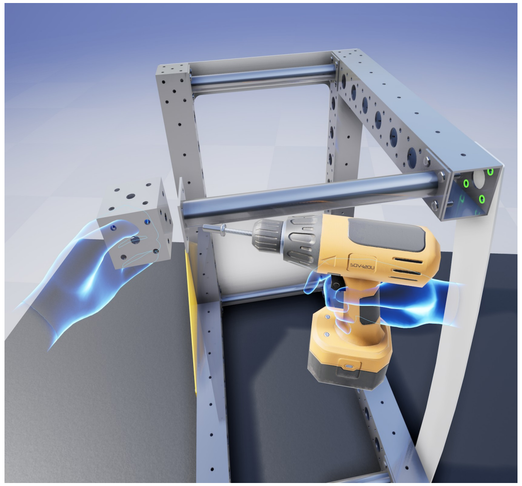
Testbed for assembly tasks

- In order to explore how VR can be used as a learning tool for assembly and planning of a production line
- Development of a testbed environment that consisted of assembly pieces that constituted a simplified and scaled-down version of the cab of an Excavator made by Volvo Construction Equipment
- The representations in the VR testbed were to be regarded as engineering models.
- A VR model, like an engineering drawing, constitutes a language in itself, containing conventions and symbols that represent very specific details and expected actions (Fergeson, 2001)



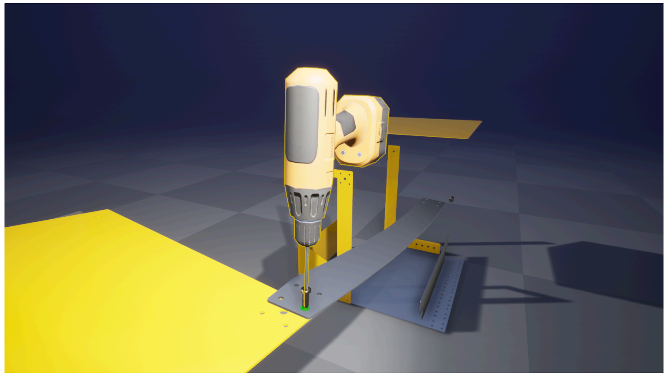
Pilot study

- As part of the engineering program students learned about assembly by using a manufactory lab consisting of three different simplified and scaled-down models of construction equipment cabs
- The testbed was a VR version of this manufactory lab
- The students were expected to figure out the best way to assemble the cabs and determine the time for assembly different parts
- After using the VR environment the students were interviewed



Insights

- Lack of weight and gravity challenged existing perception of interaction with objects - conflict with physical rules from the real world
- Objects moved and it was hard to get a grip of them – hard to learn tasks using fine motor movements
- Assembling tasks were faster in a VR-environment and it was not possible to estimate real times for different steps
- Perceptual cues can be supported and changed (for example using sound that represent increasing resistance/weight)
- VR can be used for learning sequences, but not (today) for initial learning of tasks that require fine motor skills and “tacit knowledge” that is hard to describe



Future work

- Further investigate important perceptual features to convey in the VR-environment
- Perceptual features that needs to be similar between VR environments and real life (to be able to learn, experience etc)
- Perceptual features that can be compensated for – how and what will work?
- Differences between between different domains, applications and usage purposes

Panellist Position

3D Data Visualization Using a Spherical Display

Prima Oky Dicky Ardiansyah
Iwate Prefectural University, Department of Software and Information Science, Japan.
prima@iwate-pu.ac.jp

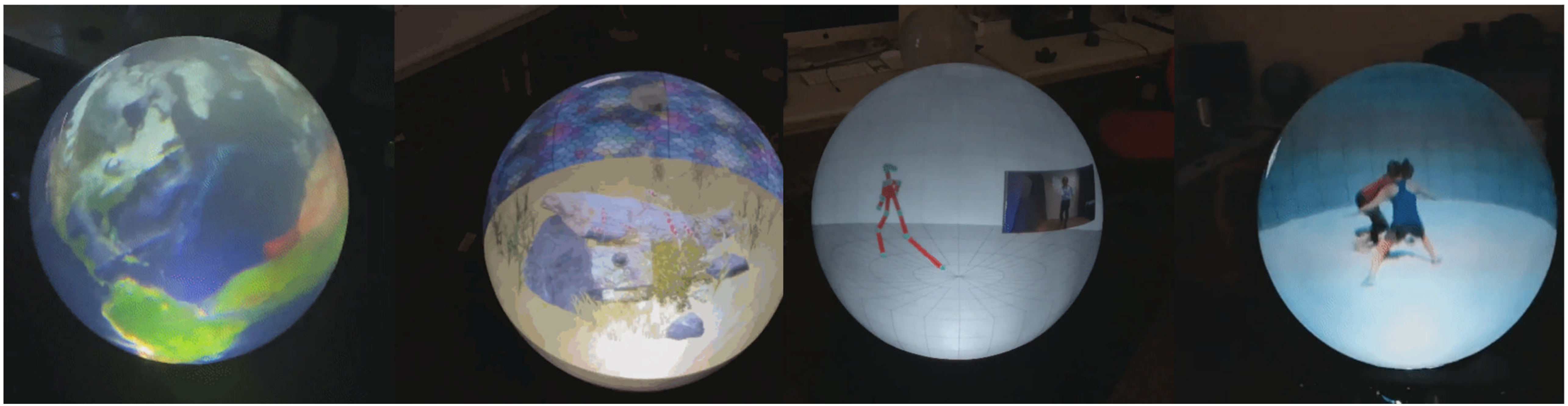
- 3D Eye Tracking
- 3D Human Pose Estimation
- Behavior Analysis
- Virtual Reality
- 3D Perspective-Corrected Display

Iwate Prefectural University

岩手県立大学
ソフトウェア情報学部
Faculty of Software and Information Science

Our lab has been developing some computer-vision-related devices and applications, such as a 3D eye tracker based on convergence eye movements, automated behavior analysis during conversations, and a 3D perspective-corrected spherical display.

We will share our experience in visualizing the 3D data into a spherical display.



3D Data Visualization Using a Spherical Display

Oky Dicky Ardiansyah Prima

Email: prima@iwate-pu.ac.jp
Department of Software and Information Science
Iwate Prefectural University, Japan.

Our lab has been developing some computer-vision-related devices and applications, such as a 3D eye tracker based on convergence eye movements, automated behavior analysis during conversation, 3D motion analysis using a single camera, and a 3D perspective-corrected spherical display. We will share our experience in visualizing the 3D data into a spherical display.

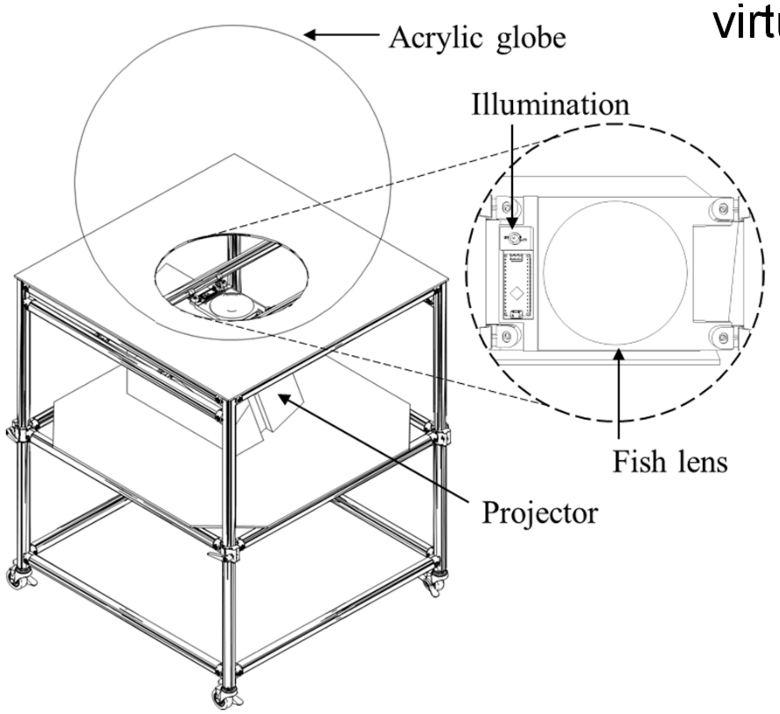
Introduction:: A 3D Perspective-Corrected Spherical Display

Eye contact with a virtual character [2]

It was in 2018, when I took a sabbatical to the HCT-lab[1] at the University of British Columbia, where I had an opportunity to work on a 3D perspective corrected spherical display. After I got back to Japan, I started building an affordable spherical display that has the same function as the HCT-lab's.

Features implemented in our spherical display:

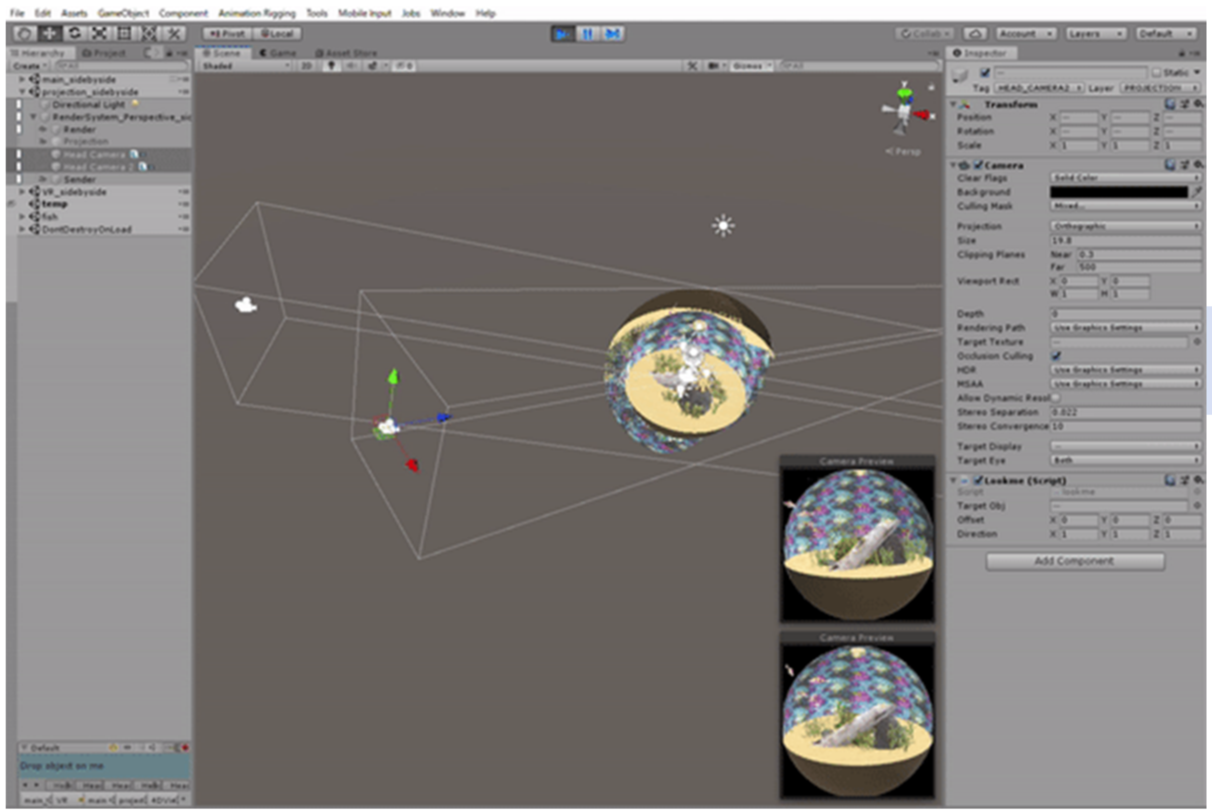
- Stereoscopic 3D rendering (single viewpoint)
- Motion-parallax 3D rendering (multi-viewpoint)
- Touch interface
- SDK for the Unity 3D



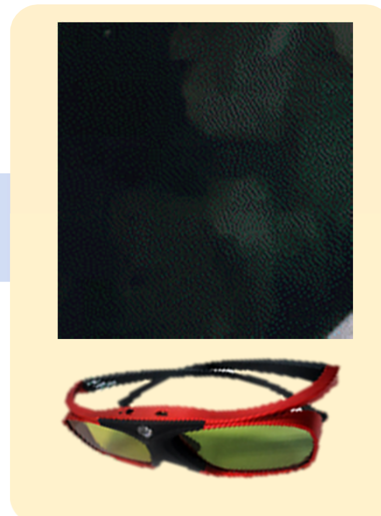
Our spherical display [3]

Introduction:: A 3D Perspective-Corrected Spherical Display

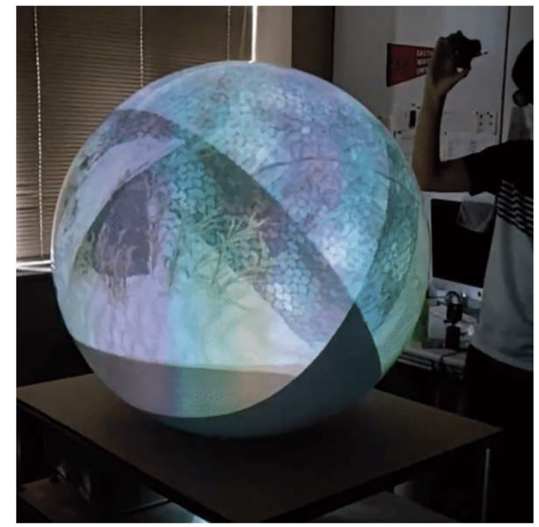
3D images are rendered in real-time according to the users' viewpoints, which are tracked by a motion capture device.



3D rendering with the Unity 3D SDK (two viewpoints)



Shutter glasses

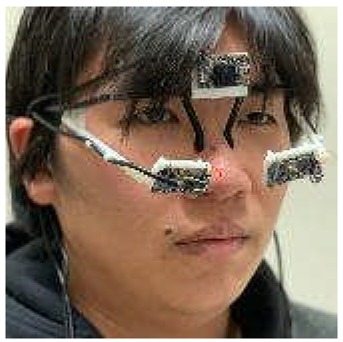


Two viewpoints

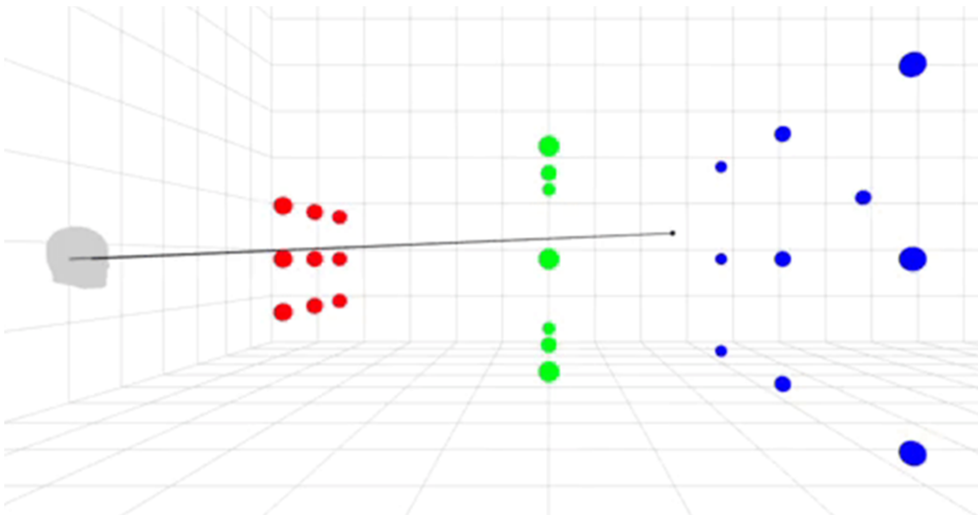
Iwate Prefectural University

Applications:: 3D eye tracker based on convergence eye movements

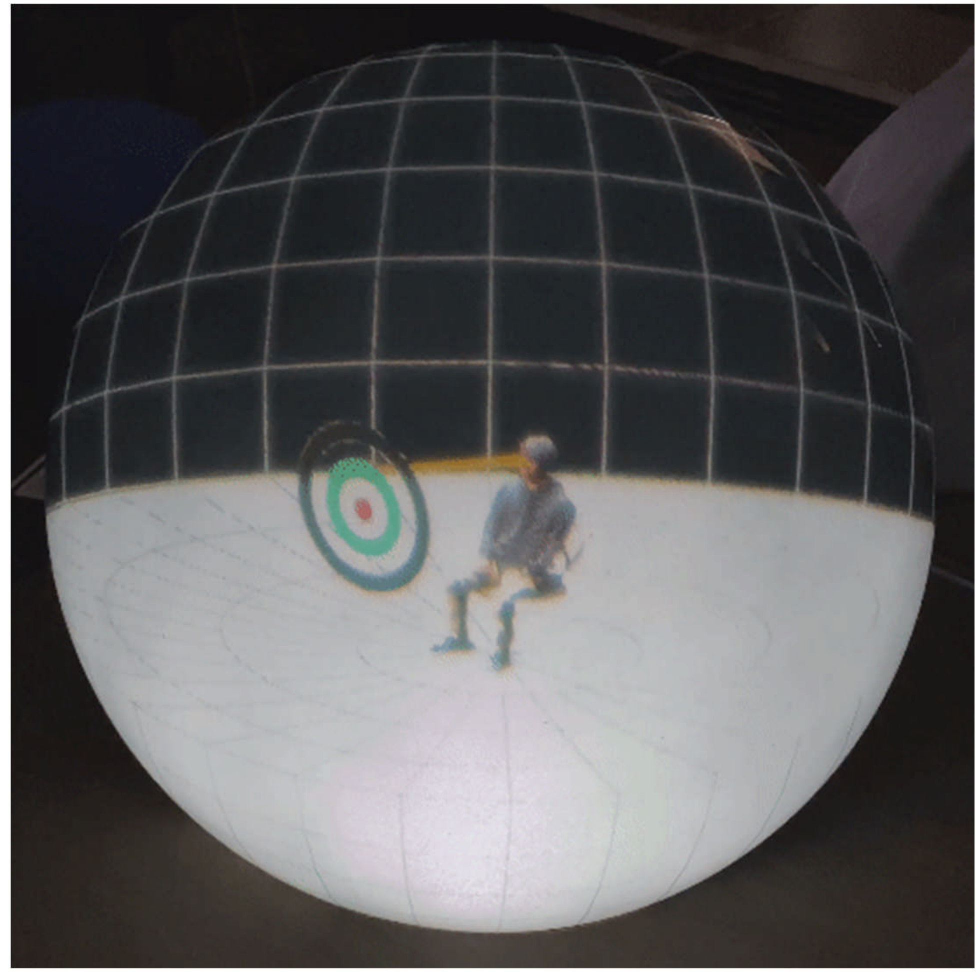
Our glass-type 3D eye tracker [4] consists of three webcams: one camera for capturing the scene and one camera for capturing the pupils of both eyes. The user's 3D eye-scanpaths can be visualized from various angles on the spherical display.



Our 3D eye tracker



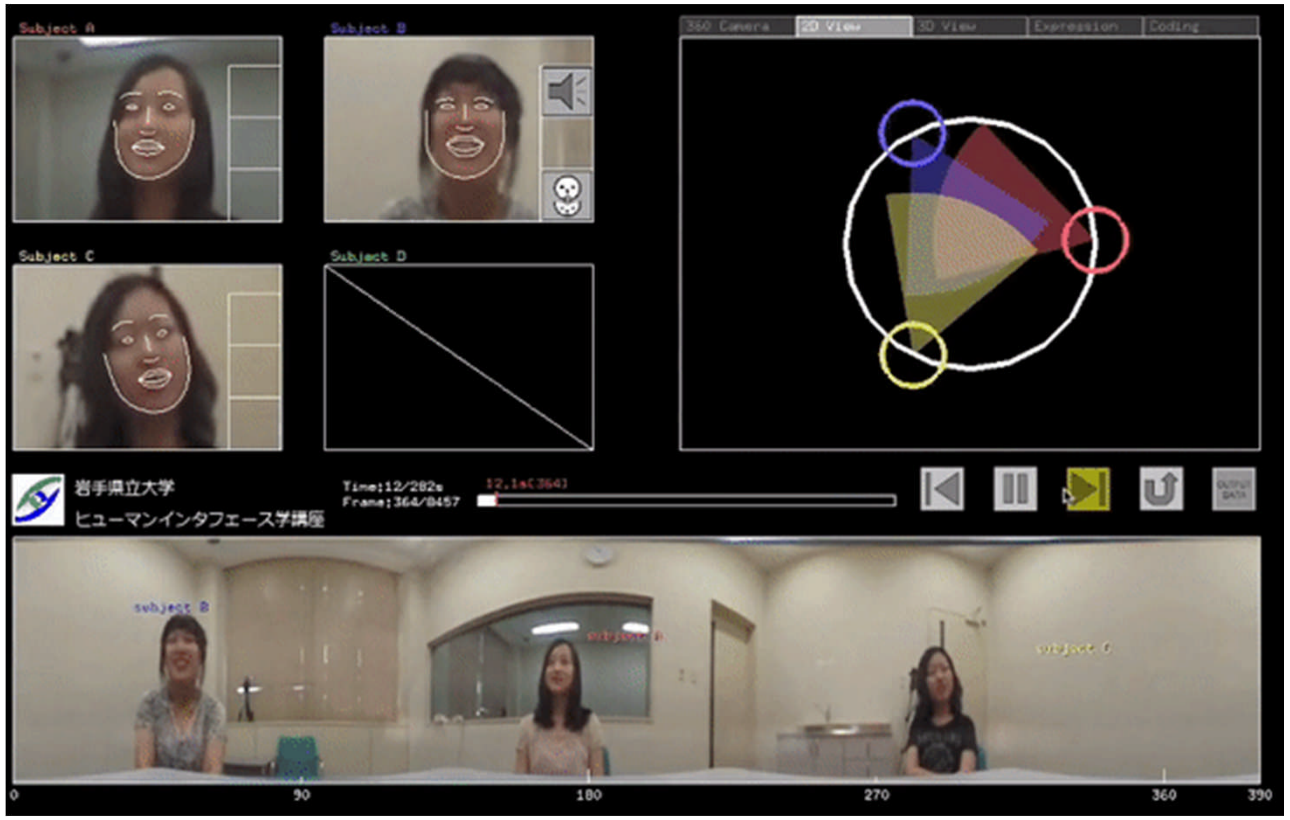
3D scanpath of a user looking at a moving target



3D scanpath visualization in a 3D spherical display

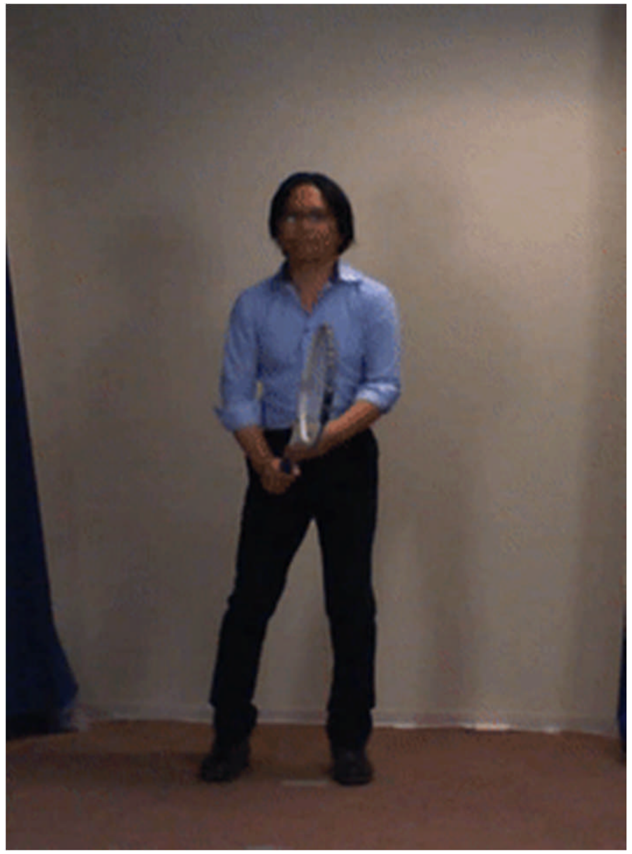
Applications:: Automated behavior analysis during conversation

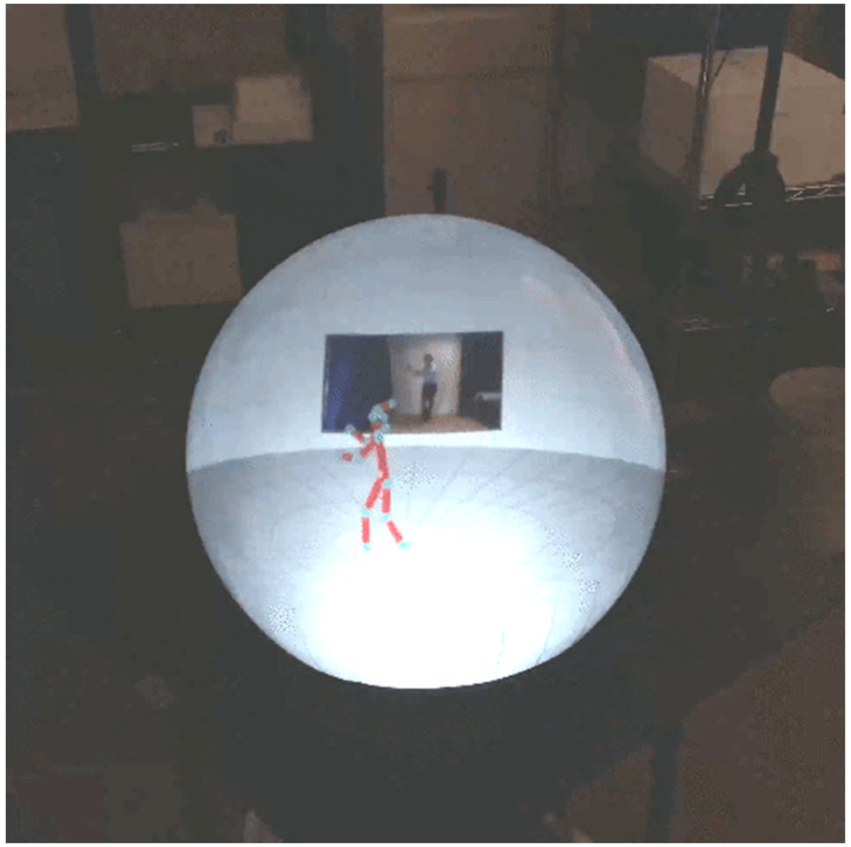
We developed a software application program [5] to automatically analyze behaviors of the participants including utterances, facial expressions (happy or neutral), head nodding, and head pose using a single omnidirectional camera.



Applications:: 3D motion analysis using a single camera

We visualized 3D human skeletons based on 3D human body pose estimation from a single vision camera. The results demonstrated that the vision camera has an advantage over the Microsoft Kinect for estimating semi-occluded joint positions [6].



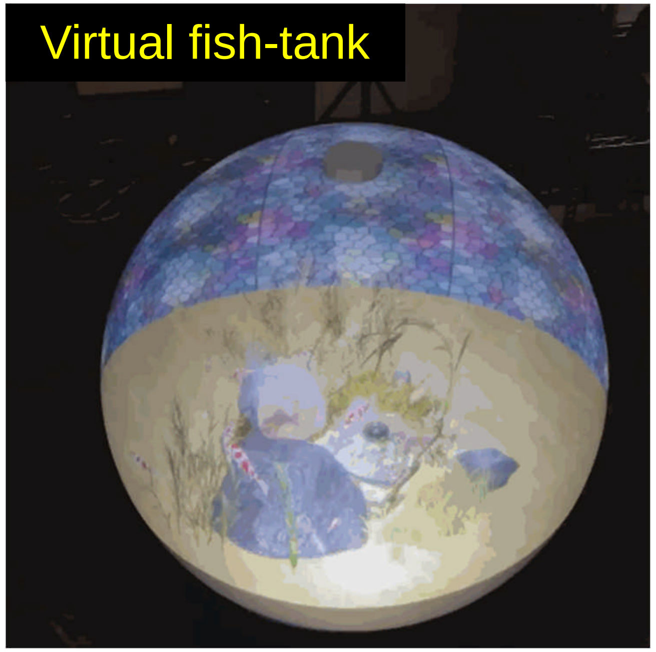


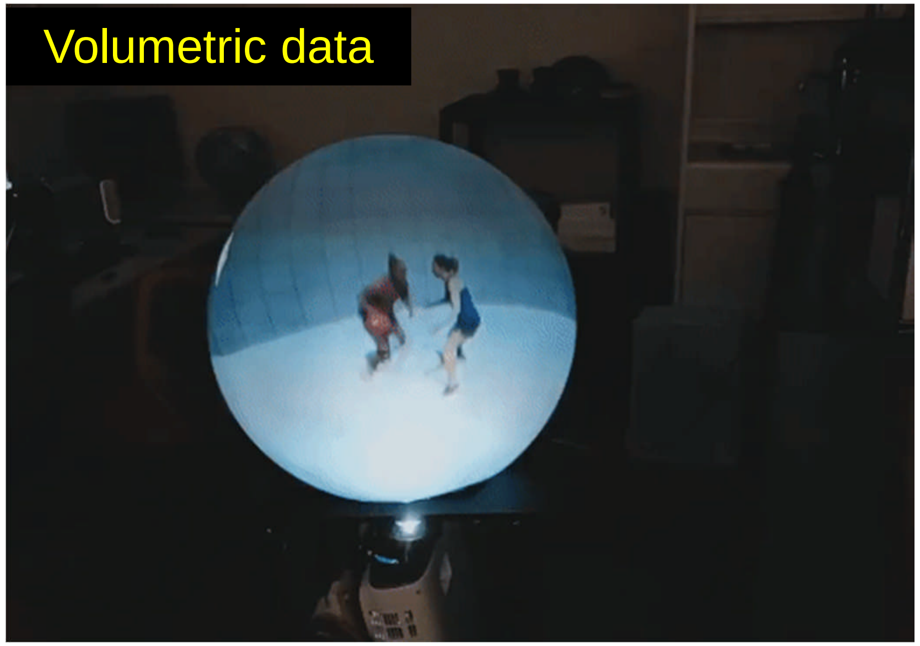
Iwate Prefectural University

Concluding Remarks

3D spherical display has the potential to visualize a variety of 3D contents, such as volumetric data, motion capture data, virtual fish-tank, and 3D digital globe. The multi-view feature allows users to observe the same object from various angles at the same time.

In the future, we will enhance its feature to enable a virtual surgical .





References

1. Human Communication Technology laboratory, https://hct.ece.ubc.ca/
2. G. Hagemann, Q. Zhou, I. Stavness, O. D. A. Prima, and S. S. Fels, “Here’s looking at you: A Spherical FTVR Display for Realistic Eye-Contact,” ISS 2018 - Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces, 357–362. 2018. https://doi.org/10.1145/3279778.3281456
3. R. Takahashi, K. Hotta, O. D. A. Prima, and H. Ito, “A Perspective-Corrected Stylus Pen for 3D Interaction,” The 13th International Conference on Advances in Computer-Human Interactions (ACHI2020), Special Track on Human Computer Interaction and Behavior Analysis Using Vision Cameras, Valencia, Spain, 2020.
4. K. Kato, O.D.A. Prima, H. Ito, “3D Eye Tracking for Visual Imagery Measurements,” In: Stephanidis C. (eds) HCI International 2019 - Posters. HCII 2019. Communications in Computer and Information Science, vol 1033. Springer, Cham., 2019. https://doi.org/10.1007/978-3-030-23528-4_32
5. Y. Jaana, O.D.A. Prima, T. Imabuchi, H. Ito, and K. Hosogoe, “The development of automated behavior analysis Software,”Proc. SPIE 9443, Sixth International Conference on Graphic and Image Processing (ICGIP), 2014.
6. O. D. A. Prima et al., “Evaluation of Joint Range of Motion Measured by Vision Cameras,” International Journal on Advances in Life Sciences, issn 1942-2660, 11, 3&4, 128-137, 2019.