Your Memory Palace in the Metaverse with AI

Immanuel Koh and Ashley Chen
Singapore University of Technology and Design
immanuel_koh@sutd.edu.sg, ashley_chen@sutd.edu.sg

Abstract

The metaverse is recognized as an immersive environment where individuals congregate as avatars to perform activities in imaginary places, similar to how they would in reality. Although promising, it has been unsuccessful in retaining novelty with current renditions of the metaverse looking unappealing and devoid of human touch, lacking in genius loci. The objective of this paper is to construct a framework for the metaverse that is built directly from the human experience to become deeply personalized and meaningful, providing users with a unique and immersive experience. This paper proposes a framework for the metaverse that leverages wearable technology and artificial intelligence (AI) to generate virtual spaces based on the conscious and subconscious experiences of individuals.

Introduction

The term "metaverse", popularised by Mark Zuckerberg's Meta has led tech companies, game developers, software engineers and architects on a race to pioneer the conceptualisation of this new digital world. The metaverse is recognized as an immersive environment where individuals congregate as avatars to perform activities in imaginary places, similar to how they would in reality. The push by companies into the virtual world, although promising and exciting, has been unsuccessful in retaining novelty with current renditions of the metaverse looking unappealing and devoid of human touch, lacking in genius loci (Tassi 2022). Apart from visuals, the biggest difficulty yet to be overcome is that of building a world that resonates with a population of different backgrounds and experiences.

This paper aims to construct a framework for the metaverse, in response to the increasing exposition on the potential of AI to visualise and describe perceptions. The sensory inputs obtained from wearable technology could potentially materialize and express human memory and perception in the digital realm. The impression of AI from that of a far cybernetic future has proven to be an important daily tool with open-to-all AI engines such as ChatGPT and Midjourney. The adoption of AI systems and algorithms coupled with direct human inputs could be used to quantify human experiences, resulting in a more successful curation of the metaverse as spaces become more targeted and relevant to provide a unique user experience. The metaverse can then act as an ever growing archive of memories and perception distinctly synthesised by its user. The paper will explore a potential structure for the metaverse using AI, brainwaves, eye tracking and photogrammetry technology.

Methodology

Typically utilised in the medical field to detect or diagnose conditions affecting the brain, EEGs are used to record brain activity by converting them into real-time electrical signals for analysis. A study done in 2015 demonstrated the potential of EEG signals in understanding emotion by highlighting specific EEG markers associated with different emotional states (Hiyoshi et al. 2015). Furthermore, these discrete emotions can be further distilled into categories of pleasure and displeasure (Barret et al. 2007) to be investigated through frontal asymmetry. The unique impact of the imbalance in frontal alpha power was first identified in research examining indicators of personality. Individuals exhibiting higher levels of left-frontal alpha displayed a tendency to process information in a predominantly positive manner such as feeling joy or delight, whereas right-lateralization indicated an inclination towards a more negative processing approach such as fear and disgust (Coan et al. 2003). This technology has been increasingly valuable, especially in creative industries where perception is subjective and difficult to quantify.

Eye tracking technology ranges from light and portable sunglasses to screen-based trackers that can be mounted onto a laptop or computer like a webcam. They utilise sensors to detect eye movement such as gaze patterns, fixation time, points of interest, pupil dilation or constriction and convert them into readable data streams (Farnsworth 2022).

Lastly, photogrammetry is a technique that can be used to extract 3D information from 2D photographs. The process involves taking photographs from different angles. The overlapping images are converted from still images into 3D digital models through a series of algorithms. These models can be exported as various file types (Aber, Marzolff, Ries 2010).
Based on current advancements in wearable technologies like Neuralink by Elon Musk (Capoot 2022), it is expected that wearable technologies will follow a similar trajectory and become increasingly portable, eventually being adopted for daily use, much like smartwatches are today. In the future, the life of a user is recorded and extracted as core memories based on their significance as indicated by the wearables. The eye tracker registers points of interest and generates gazemap signatures to uncover how intensely a particular object or space is looked at. Spaces and objects of interest are identified and matched to their corresponding EEG spectrogram which can describe emotional valence to substantiate the eye tracking data. Scenes along with their gazemaps and spectrograms are fed into a GAN (Goodfellow et al. 2014) to learn the user's perception, becoming an artificial brain capable of regenerating memory in a spatial manner. New memories and their corresponding weighted perception are added into a database and fed into the AI model daily, ensuring that the model is constantly being updated to learn and adapt. Thus, the memories of the user can subconsciously be extracted from the real world and are reformed as 3D assets in the metaverse.

Experiment

An Enobio-8 EEG and Smart-Eye AI-X screen based eye tracker and a right handed participant were used for this experiment. A series of 500 images, each shown for a duration of 3 seconds were compiled into a video. The images chosen were of interior spaces such as living rooms, bedrooms etc. to resemble memories of home, a space in which one typically spends most of his/her time. The screen based eye tracker was mounted to the laptop and calibrated to match the gaze of the participant and the EEG cap was placed on the participant with electrodes gelled to the regions F3/F4 (frontal), C3/C4 (central) to measure the alpha frequency band (8-12 Hz) to measure frontal assymetry (Hageman et al. 2002; Zhao, Zhang and Ge, 2018). Alpha frequency brainwaves from F3/F4 and C3/C4 are often used for the measurement of frontal asymmetry because of their placement on the frontal and central lobe which is associated with emotional and cognitive processing (Imotions 2017).

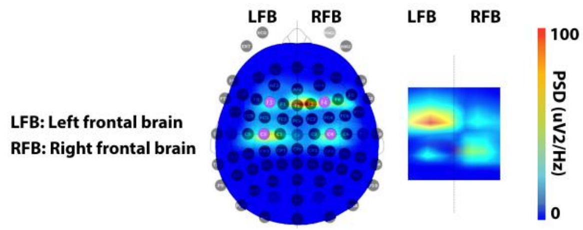

Figure 1: EEG electrode placement for alpha frontal asymmetry and spectrogram generated from power spectral density (PSD) measured by each node

Each node's value represent power spectral density (PSD) which describes how the power of a signal is distributed across different frequencies (measured in uV2/Hz) and are depicted using a color scale ranging from red (highest) to blue (lowest) (Figure 1). These values are then mapped directly to their corresponding locations on the scalp, forming a spectrogram.The participant was instructed to watch the video on a laptop while wearing the EEG cap.
The gazemaps and spectrograms were matched with their corresponding images and utilized as inputs to train two image-to-image conditional generative adversarial network models (Isola et al. 2018). The first model, AI-1, takes in 500 interior images along with 500 associated gazemaps, while the second model, AI-2, takes in 500 interior images paired with 500 corresponding spectrograms. AI-1 primarily captures the user's visual perception, whereas AI-2 focuses on capturing the perceptual and subconscious aspects of the user's mind.
Both models were trained to 500 epochs and predicted images showing different interior spaces were chosen for evaluation. For the bathroom interior, the generated image from AI-1 shows a window the size of a door, opening to lush green outdoors at the midpoint of the scene, similar to the ground truth (GT) (Figure 2). The greenery continues to bleed inwards and blends with the interiors. The low ceiling with wooden buttress from the ground truth (GT) has also shifted to become a higher ceiling exceeding the frame. Furthermore, there is a puddle on the floor as if to indicate a leaked faucet in the bathroom. Similarly, the output from AI-2 has glass block windows in the middle of the image that are typically used in bathrooms for privacy whilst letting daylight in. It can be observed that the outdoor greenery seeps into the scene and into a form resembling a bathtub with soap spilling over.

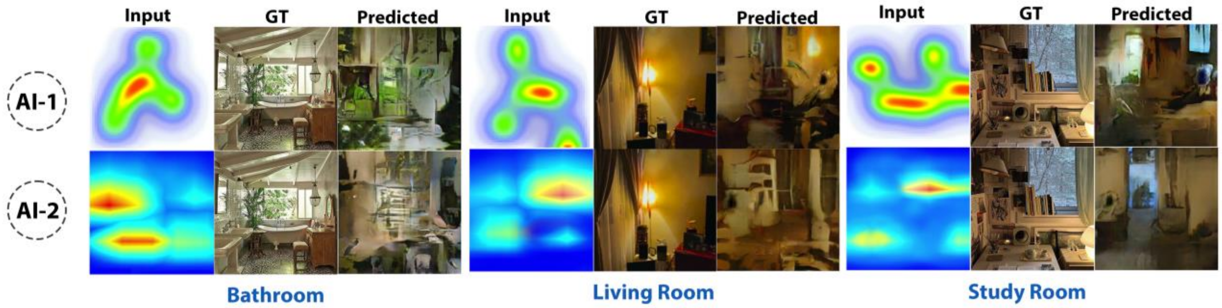

Figure 2: Predicted images based on bathroom, living room and study room interiors

The predicted image from AI-1 living room has a generated lamp of similar colour shade at the right corner of the room which illuminates the space. A window is formed on the adjacent wall and an armchair that resembles a portion of a piano is placed against it. The bottom corners of the image shows two more armchairs facing the window. One can infer that the scene is captured from the point of view of someone walking into the room towards the armchair. In AI-2 however, it is more difficult to distinguish pieces of furniture from the output. Instead, there is an emergence of wall mounted shelves of different shapes and sizes around a window with closed curtains. Although the lamp is absent from the scene, the space is illuminated with a similar light source that is out of the frame.

Lastly, the AI-1 study room prediction features two windows representing different times of day, morning and evening, in contrast to the ground truth (GT) which has only one large window during the evening. The generated study space appears from a side view, with the chair positioned on the left near the daytime window and the desk on the right beneath the evening window. A blue lamp-like object is placed on the desk, and wall-mounted shelves are present between the windows. On the other hand, the AI-2 prediction includes an evening window that appears open, with a large figurine situated on the windowsill. To the left of the window, there is a table featuring what appears to be a decorative item. Towards the right of the window, there is a desk placed against the wall with the back of the chair seen tucked into it, along with books on a wall-mounted shelf. The lighting in the room closely resembles that of the ground truth (GT).

The results generated by AI-1 and AI-2, while sharing certain similarities, introduce spatial elements and qualities derived from an individual's conscious and subconscious perception, thereby defamiliarizing the ground truth (GT) through design operations. These AI models form links between memory and emotion and are able to form a predicted perception, working as an artificial brain capable of perceiving the world from the perspective of the individual.

Figure 3: Point cloud living room designed according to predicted image from Figure 2

A potential metaverse space (Figure 3) was created based on the predicted image of the living room as generated in Figure 2. Furniture was scanned with an iPhone and converted into an image sequence to be uploaded into Polycam, an application available on mobile which converts the images into a 3d asset. For this experiment, point clouds were chosen as the medium to represent artificially generated memories. Their visual appeal conveys the essence of memories that are momentary perceptions, or fleeting experiences that leave a vague or incomplete impression in one's mind. This is due to its density and resolution that is highly dependent on the time and angle at which the space or object is looked at.

Future Work

The methodology presented in this paper demonstrates the feasibility of incorporating EEG, eye tracking, and photogrammetry technologies to capture and translate human experiences into the metaverse. As these wearable technologies become increasingly portable and widely adopted, they have the potential to enrich our understanding of human perception and memory. In the future, various data types beyond images can be extracted from wearable technologies that could be adapted as inputs for GANs. For example, it is possible to acquire data streams from the EEG, including information from all 8 electrodes, and gather data related to attention and facial analysis from the eye tracker. Increasing the number of participants and physiological inputs in future studies can also provide a broader understanding of human perceptions. Furthermore, human memory is complex and made up of sensations other than sight. It would be advantageous for the eye tracker to capture audio recordings encompassing speech and ambient sounds. These recordings could be utilized in natural language processing applications, such as speech-to-text conversion or text-to-3D using DreamFusion (Poole et al. 2022), adding another layer of perception onto generated spaces of the metaverse.

Evaluation

The integration of AI algorithms with wearable technology allows for the quantification of human experiences, enabling a more targeted and relevant curation of an ever-growing archive of memories and perceptions to form the metaverse. These synthesized digital assets can potentially reveal aspects of us that were forgotten or unknown, offering a new dimension of self-discovery and exploration.

Overall, it presents a new way of conserving memory which has wide-ranging applications across various industries. In the built industry, where heritage buildings or historical landmarks which are subject to deterioration and demolition can be preserved in its original state for future generations. In the healthcare industry, virtual reality environments have been proven to help ease the pain of burn-injured patients (Hoffman et al. 2011). These AI and wearables could potentially be used to simulate virtual environments associated with positive emotions from a patient's memory as treatment for their mental wellbeing and subsequently overall physical health. This becomes particularly beneficial for patients who are immobile, confined to their rooms, or dealing with conditions like dementia, as they can revisit or recreate significant memories, enhancing their overall quality of life.

References

Aber, J. S.; Marzolff, I.; Ries, J. B. 2010. Chapter 3 - Photogrammetry. *In Small-Format Aerial Photography*, edited by James S. Aber, Irene Marzolff, and Johannes B. Ries, 23–39. Amsterdam: Elsevier, doi.org/10.1016/B978-0-444-53260-2.10003-1.

Barrett, L. B.; Mesquita, B.; Ochsner, K. N.; and Gross, J. J. 2007. The Experience of Emotion. *Annual Review of Psychology 58*, no. 1 (2007):373–403. doi.org/10.1146/annurev.psych.58.110405.085709.

Capoot, A. 2022. Elon Musk Shows off Updates to His Brain Chips and Says He's Going to Install One in Himself When They Are Ready. https://www.cnbc.com/2022/12/01/elon-musks-neuralink-makes-big-claims-but-experts-are-skeptical-.html. Accessed: 2023-05-06.

Coan, J. A.; and Allen, J. J. B. 2003. Frontal EEG asymmetry and behavioral activation and inhibition systems. *Psychophysiology 40*: 106-114. DOI: 10.1111/1469-8986.00011.

Farnsworth, B. 2023. The iMotions Screen-Based Eye Tracking Module (Explained). https://imotions.com/blog/learning/product-guides/screen-based-eye-tracking-module/. Accessed: 2023-05-10

Goodfellow, I. J.; Pouget-Abadie, J. ; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Networks. arXiv:1406.2661.

Hiyoshi-Taniguchi, K.; Kawasaki, M.; Yokota, T.; Bakardjian,H.; Fukuyama, H.; Cichocki, A.; and Vialatte, F.B. 2013. EEG Correlates of Voice and Face Emotional Judgments in the Human Brain. *Cognitive Computation 7*, no. : 11–19. DOI:10.1007/s12559-013-9225-0.

Hoffman, H. G.; Chambers, G. T.; Meyer, W. J.; Arceneaux, L. L.; Russell, W. J.; Seibel, E. J.; Richards, T. L.; Sharar, S. R.; Patterson, D. R. 2015. Virtual Reality as an Adjunctive Non-Pharmacologic Analgesic for Acute Burn Pain during Medical Procedures. *Annals of Behavioral Medicine 41*, no. 2: 183–91. DOI: 10.1007/s12160-010-9248-7.

iMotions. 2017. Frontal Asymmetry 101 – How to Get Insights on Motivation and Emotions from EEG. https://imotions.com/blog/learning/best-practice/frontal-asymmetry-101-get-insights-motivation-emotions-eeg/. Accessed: 2023-05-13

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2018. Image-to-Image Translation with Conditional Adversarial Networks. arXiv:1611.07004.

Poole, B.; Jain, A.; Barron, J. T.; and Mildenhall, B. 2022 DreamFusion: Text-to-3D using 2D Diffusion. arXiv.2209.14988.

Tassi, P. 2022. Does Mark Zuckerberg Not Understand How Bad His Metaverse Looks? https://www.forbes.com/sites/paultassi/2022/08/17/does-mark-zuckerberg-not-understand-how-bad-his-metaverse-looks/?sh=7775497e37d4. Accessed: 2023-05-09

Zhao, G.; Zhang, Y.; and Ge, Y. 2018. Frontal EEG Asymmetry and Middle Line Power Difference in Discrete Emotions. *Frontiers in Behavioral Neuroscience 12 (2018)*. doi.org/10.3389/fnbeh.2018.00225.