Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.045768

ARTICLE

Enhancing Identity Protection in Metaverse-Based Psychological Counseling System

Jun Lee[1], Hanna Lee[2], Seong Chan Lee[2] and Hyun Kwon[3,*]

[1]Department of Game Software, Hoseo University, Asan-si, 31499, Korea

[2]Yatav Enter, Seoul, 04799, Korea

[3]Department of Artificial Intelligence and Data Science, Korea Military Academy, Seoul, 01805, Korea

*Corresponding Author: Hyun Kwon. Email: hkwon.cs@gmail.com

Received: 07 September 2023 Accepted: 20 November 2023 Published: 30 January 2024

ABSTRACT

Non-face-to-face psychological counseling systems rely on network technologies to anonymize information regarding client identity. However, these systems often face challenges concerning voice data leaks and the suboptimal communication of the client's non-verbal expressions, such as facial cues, to the counselor. This study proposes a metaverse-based psychological counseling system designed to enhance client identity protection while ensuring efficient information delivery to counselors during non-face-to-face counseling. The proposed system incorporates a voice modulation function that instantly modifies/masks the client's voice to safeguard their identity. Additionally, it employs real-time client facial expression recognition using an ensemble of decision trees to mirror the client's non-verbal expressions through their avatar in the metaverse environment. The system is adaptable for use on personal computers and smartphones, offering users the flexibility to access metaverse-based psychological counseling across diverse environments. The performance evaluation of the proposed system confirmed that the voice modulation and real-time facial expression replication consistently achieve an average speed of 48.32 frames per second or higher, even when tested on the least powerful smartphone configurations. Moreover, a total of 550 actual psychological counseling sessions were conducted, and the average satisfaction rating reached 4.46 on a 5-point scale. This indicates that clients experienced improved identity protection compared to conventional non-face-to-face metaverse counseling approaches. Additionally, the counselor successfully addressed the challenge of conveying non-verbal cues from clients who typically struggled with non-face-to-face psychological counseling. The proposed system holds significant potential for applications in interactive discussions and educational activities in the metaverse.

KEYWORDS

Metaverse; counseling system; face tracking; identity protection

1 Introduction

Psychological counseling typically involves clients meeting counselors to discuss their concerns and seek solutions [1,2]. In the traditional psychological counseling model, clients usually attend face-to-face sessions at the counselor's office. These sessions can be one-on-one interactions between

individual clients and counselors or group counseling sessions with multiple participants. Nevertheless, in traditional psychological counseling, clients often experience fear and reluctance when revealing their personal identity information during sessions. This is because it requires them to disclose their actual face and voice, which can be a source of discomfort for some clients.

Furthermore, starting with onset of the COVID-19 (Coronavirus disease 2019) pandemic, clients have shifted from traditional, in-person counseling to online, non-face-to-face counseling [3]. Non-face-to-face counseling is often conducted over phone or video-based programs, providing clients and counselors with a more accessible environment that eliminates time and space constraints. Moreover, clients experience less burden regarding the inadvertent leakage of their identity information compared to traditional face-to-face counseling, allowing them to more easily express thoughts and concerns that may be challenging to communicate directly to the counselor. Despite the advantages of non-face-to-face counseling, it still faces certain technical limitations. Firstly, from the client's perspective, there is a legitimate concern about potential leakage of identity information through their voice, as they are exposed when engaging non-face-to-face counseling. Secondly, even in face-to-face counseling, there are situations where clients may prefer not to reveal their face, leading to reluctance when asked to participate via video chat program [4,5]. Thirdly, counselors also encounter challenges. In non-face-to-face counseling, counselors heavily rely on the client's voice for communication, making it essential to perceive non-verbal cues such as facial expressions for a more accurate diagnosis and effective counseling. However, conventional non-face-to-face counseling systems struggle to address this aspect effectively.

The metaverse is a simulated environment that facilitates a broad range of activities, such as communication and economic interactions, within a three-dimensional (3D) virtual world, free from the physical constraints of time and space [6]. Avatars, which represent the users in the metaverse, are created user-created and find extensive utility across various fields, such as gaming, social networking, fashion, and fire-fighting training, allowing users to engage with others and explore the metaverse [7]. A metaverse-based counseling platform presents an alternative to traditional non-face-to-face psychological counseling platforms. It enabled users to interact with others through voice communication and chat while controlling their avatar's movements. This guarantees anonymity, allowing users to safeguard their personal identity.

Nonetheless, the conventional counseling platform in the metaverse environment has a drawback, wherein others can potentially identify clients through the unaltered transmission of voice information when voice communication is enabled. Additionally, the client's facial expression cannot be seen, making it difficult for the counselor to discern and analyze the client's non-verbal cues, despite the avatar being visible during the conversation.

To address these issues, this study proposes a metaverse-based psychological counseling system. This system employs real-time face recognition based on a histogram of gradient (HOG) to capture the client's facial expression and convey it to the counselor without revealing the client's identity. The proposed system allows for the real-time expression of the client's avatar in the metaverse environment. Additionally, the voice modulation method is utilized to protect the client's voice identity.

This study makes the following contributions:

- It proposes a metaverse-based psychological counseling system that utilizes identity protection technology through avatars and real-time voice modulation to safeguard information regarding client identity, while providing counseling services in the metaverse without revealing their face.
- It also employs real-time facial recognition technology for clients during non-face-to-face metaverse counseling on mobile devices to provide optimized information on their avatars.

The remainder of this paper is organized as follows: In Section 2, we present an overview of related literature concerning psychological counseling in the metaverse. Section 3 describes the details of the proposed metaverse-based psychological counseling system. In Sections 4 and 5, we discuss the experiments conducted and the results obtained, respectively. Finally, the conclusions are outlined in Section 6.

2 Related Work

2.1 Metaverse for Non-Face-to-Face Consultation and Identity Protection

Typically, counselors are required to hold professional qualifications and licenses, as they bear responsibility for ensuring the confidentiality of their clients [3,4]. However, several security issues can surface during the consultation process. These issues fall in two categories. The first is when the client's personal information, such as their name and face, gets revealed during counseling. The second scenario pertains to scenarios where the client's identity can be inferred based on their voice during the consultation process, even if they choose to conceal their basic identity information (name and face) [4]. Thus, preventive measures are necessary to safeguard the identity of clients since both cases can lead to serious ethical problems.

The term 'metaverse' combines elements pertaining to the virtual, the transcendental, and the universe in a 3D virtual world. It was first introduced in Neal Stephenson's science fiction novel "Snow Crash" in 1992 [6]. Although initially a conceptual idea, metaverse gained significant popularity following events in April 2020 when the Battle Royale game Fortnite hosted a concert featuring Travis Scott in his avatar [7]. This event sparked considerable enthusiasm in the metaverse era, with numerous celebrities releasing new songs, concerts, and products. This study focuses on the avatar as a key feature of the metaverse. Avatars are digital representations of users, allowing them to engage in games, entertainment, social interactions, and economic activities within the virtual world. Combining the metaverse environment with non-face-to-face counseling can be advantageous in mitigating the challenges faced in ensuring identity protection. Using avatars, clients who are hesitant to reveal their identity can conduct counseling sessions within the metaverse environment via accessing a voice communication platform. Some counseling services have adopted the metaverse platform to enable non-face-to-face counseling using avatars. Metaverse Seoul, a metaverse solution in Seoul, South Korea, provides administrative services and youth counseling (online access: https://metaverseseoul.kr/user). Voice modulation technology is widely used to protect user identity information in non-face-to-face communication environments. This technology changes the tone and pitch of a user's voice in real-time by analyzing their voice information [8]. Subsequently, while applying voice modulation on a user's speech can enhance the user's privacy and anonymity to some extent, it may not provide complete protection against being identified [9].

2.2 Facial Recognition and Expression Technology

Facial alignment technology involves detecting facial landmarks on an individual's face, identifying them based on these landmarks. One conventional computer vision approach for detecting facial landmarks entails locating both eyes in a given image and extracting their positional information. Subsequently, the face alignment process includes identifying the nose below both eyes and the mouth below the nose [10]. Despite its utility, the traditional feature point extraction method for face alignment often suffers from low accuracy in real-world environments. To improve accuracy, researchers have turned to deep-learning technology for tracking face alignment and recognizing facial features. As early as 2014, DeepFace incorporated deep-learning technology for facial recognition [11].

The model employed a pre-trained 3D face geometry model, followed by face alignment using affine transformation, and learned using a nine-layer locally connected convolutional network. Although this approach displayed high accuracy, its main drawback was the time-consuming network-building and learning processes. The Visual Geometry Group (VGG) proposed a deep network structure called VGGFace (or DeepFR) [12]. This structure utilizes 3×3 fill convolution filters, similar to VGG, to create a dataset structure that can achieve relatively high accuracy in face recognition. Nonetheless, while it performs well in recognizing static images, it encounters challenges in recognizing faces within the metaverse environments where users may assume varying poses and use real webcams, resulting in slow face recognition. Kazemi et al. presented a method to quickly and robustly extract a user's facial data through an ensemble on the Regression Tree [13]. While this method offers the advantage of faster recognition compared to other methods, it may have limitations in capturing a wide range of facial expressions. In addition, several techniques have been suggested for extracting various facial landmarks. However, the network structure can be too complicated for recognition on devices with low computing power, such as mobile devices.

Research into avatar facial expressions in avatars that closely resemble those of real people in virtual reality environments has the potential to enhance social interactions in virtual environments, enabling conversations with virtual entities that provide a conversational experience akin to real interactions. Thalmann et al. discovered that increasing the social presence of virtual avatars has a similar effect to virtual avatars interacting with real colleagues [14]. Virtual faces are dynamic and adaptable, allowing for interactive practice. Immersive VR studies utilizing implicit methods have demonstrated the ability to induce emotions [15–19]. However, conventional metaverse studies do not provide examples of real-time mapping of a user's actual facial expressions onto virtual avatars.

3 Proposed System

This study proposes a metaverse psychological counseling system, as shown in Fig. 1. Users can access the proposed system via both personal computers (PCs) and mobile devices. DBManager performs matching based solely on their identity information to ensure secure and accurate matching between the counselor and the client. Metaverse Counseling Matching does not have access to any other information except the user identity. The counselor and client can securely exchange passwords through RSA-based authentication to ensure confidentiality and use Metaverse Communication Manager to open a private counseling room. Subsequently, users may engage in counseling through their avatars. The proposed system recognizes the faces of the users who participate in metaverse counseling in real-time. Users who use a PC employ a webcam, while those using a mobile device utilize the camera attached to the device for face recognition and voice communication. The proposed system processes the user's voice input and facial image information through the Input Manager. Voice Modulation analyzes the user's voice information in real-time to determine tone and pitch. Based on this analysis, the proposed system classifies voice bands that can be modulated in high and low tones, applies the modulation, and transmits the counselor's voice message to the users. The proposed system extracts facial landmarks from the user's facial image data and matches the information to the facial expressions of the virtual avatar during Face Conversion.

The implementation of the metaverse system proposed in this study was carried out through a series of steps. Firstly, we utilized the powerful 3D game engine Unity 2021.3.27f1 LTS to facilitate seamless communication between 3D models and avatars within the metaverse counseling system (online access: https://unity.com/releases/editor/archive). Subsequently, we leveraged the capabilities of the Photon Engine, a cloud service, to enable real-time avatar movement, chatting, and voice

communication among users, thereby establishing a robust network infrastructure within the metaverse (online access: https://www.photonengine.com). Lastly, the facial recognition module was converted from the PyTorch-based trained file into an ONNX file, and the Unity Barracuda package was employed to seamlessly integrate it into real-time operations (online access: https://onnx.ai).

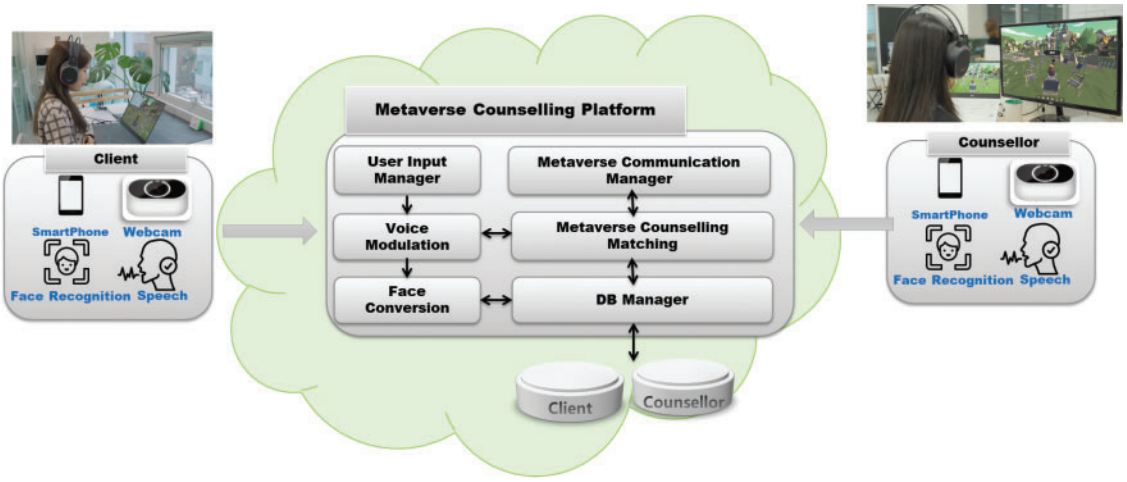

Figure 1: Overall structure of the proposed system

The flowchart illustrating the user input interaction within the system proposed in this study is depicted in Fig. 2. This proposed system employs a webcam to conduct human tracking, focusing on facial and voice data while the clients engage with the system. During this process, if the client's facial data is identified within the video feed, facial detection is executed by extracting the HOG features. This is followed by facial alignment based on the acquired data. Subsequently, using the information acquired from this procedure, the metaverse avatar promptly updates its facial expressions to mirror those of the client in real time.

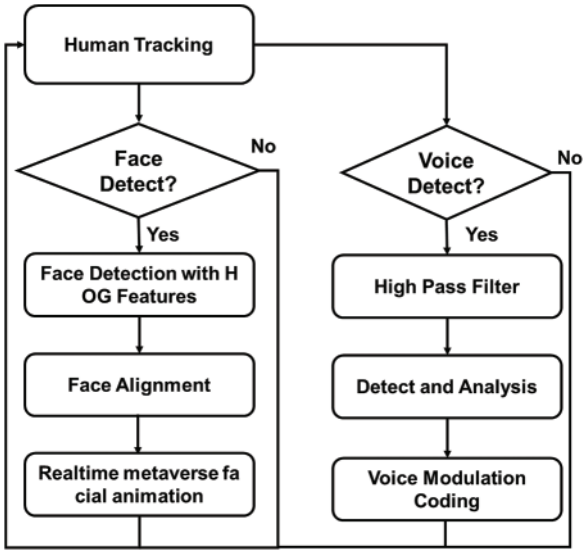

Figure 2: Flowchart of the interaction process

When the system detects the client's voice data, it undergoes voice signal processing to identify the pitch component after subjecting it to high-pass filtering. Subsequently, voice modulation is applied according to a predetermined voice pattern. The transformed video and voice data are then portrayed through the metaverse avatar, making it visible to other participants in the metaverse consultation, and the voice is reproduced in the modulated form. The algorithm for this process is outlined in List 1.

List 1: Interaction algorithm of the proposed system.

Human Tracking:

- If the face of the client is detected:
  - Perform face detection with HOG features
  - Extract face landmarks with face alignment
  - Animate facial expressions of metaverse avatar
- If the voice of the client is detected:
  - Filter the voice using a high pass filter
  - Detect the pitch of the voice and analyze it
  - Perform voice modulation with a specific voice pattern

A real-time metaverse face animation method involving two-stage face recognition and emotion recognition is proposed, as shown in Fig. 3. This method is designed to extract the user's facial data via a camera and replicate the avatar's facial expressions during metaverse counseling. In the first step of the proposed system, HOG features are extracted from an image captured through a camera. These extracted HOG features are then utilized for face landmark detection, and rapid recognition is achieved through an ensemble of regression trees to accurately track the face landmarks [13]. The recognized face landmark information is matched with a metaverse character, enabling real-time animation of the face landmarks. Additionally, pruning is applied to the constructed regression tree in the proposed system, as described in Eq. (1), aiming to reduce the data and facilitate regression tree construction on mobile devices [20].

$$R_1(j,s) = \{X|X_j < s\} \text{ and } R_2(j,s) = \{X|X_j \geq s\}$$

$$\sum_{i:x_i \in R_1(j,s)} \left(y_i - \hat{y}R_1\right)^2 + \sum_{i:x_i \in R_2(j,s)} \left(y_i - \hat{y}R_2\right)^2 \tag{1}$$

$$\sum_{m=1}^{|T|} \sum_{x_i \in R_m} \left(y_i - \check{y}_{R_m}\right)^2 + a|T|$$

Figure 3: Face fusion through face alignment and emotion recognition

A total of 64 facial landmark points are extracted. They are linked in real-time with the facial information of the user's avatar in the metaverse environment, as illustrated in Fig. 4. Specifically, the facial landmark information corresponding to eyebrows, eye couple, mouth, and jaw are individually blended in the metaverse environment corresponding to those of the avatar, as shown in Fig. 3. The proposed method defines the minimum and maximum ranges of the avatar's face that can be changed based on the position Pt, with P_{min} and P_{max} set to 0 and 1, respectively, as described in Eq. (2). If the two-dimensional position values of the facial landmark points recognized in real-time are represented

as F, with F1 = (x1, y1) and F2 = (x2, y2), the magnitude vector between the two points can be defined as |F|. When the distance between the two points is approximately 0, the points in that area are set to the minimum value, P_{min}, indicating that the user's eyes are closed, as illustrated in Fig. 3. Conversely, when the distance between the two points approaches 1, P is set to P_{max}, as this represents the most distinctive form when the user's mouth is open. Subsequently, blending is carried out along the corresponding points, enabling the real-time reflection of the primary components of the facial expression. The process of real-time animation using the Face Fusion Animator involves the seamless merging and manipulation of multiple facial models through Unity Engine's blend-shape. The resulting animation is achieved through the natural integration of various facial expressions, as shown in Fig. 4.

$$
\begin{aligned}
&F_1 = \{x_1, y_1\}, F_2 = \{x_2, y_2\} \\
&P_{mix} = 0, \ P_{max} = 1 \\
&|F| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \\
&IF\ |F| \to 0, P \approx P_{min} \\
&IF\ |F| \to 1, P \approx P_{max}
\end{aligned}
\tag{2}
$$

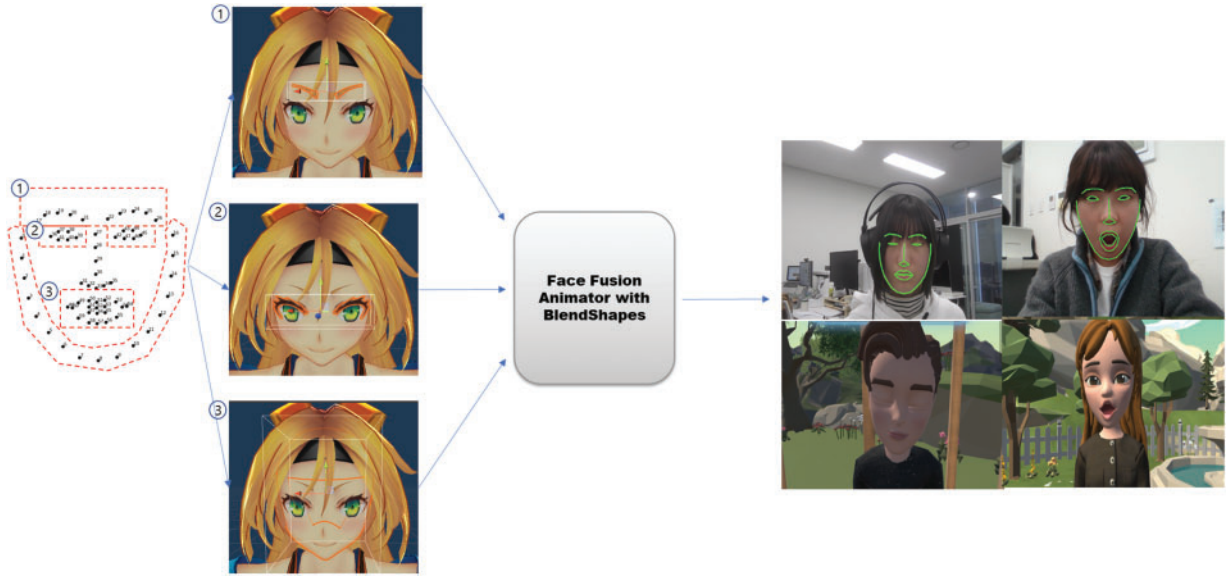

Figure 4: Face fusion through facial alignment

The protection of user identity is enhanced by implementing voice modulation in the proposed system. To prevent performance degradation when applying voice modulation to mobile device users, we first eliminate noise from the user's voice data and perform sampling, as shown in Fig. 5. Subsequently, the proposed system performs voice data processing through distinct modules specifically designed for pitch recognition and analysis, as well as linear predictive coding (LPC) analysis. The proposed system identifies and analyzes pitch information in the voice data within the pitch recognition and analysis module. The voice signals are patterned and sampled in the LPC module through vector quantization. Once completed, the proposed system modulates the voice by combining the previously defined voice modulation patterns with the analyzed pitch and sampled voice signal information.

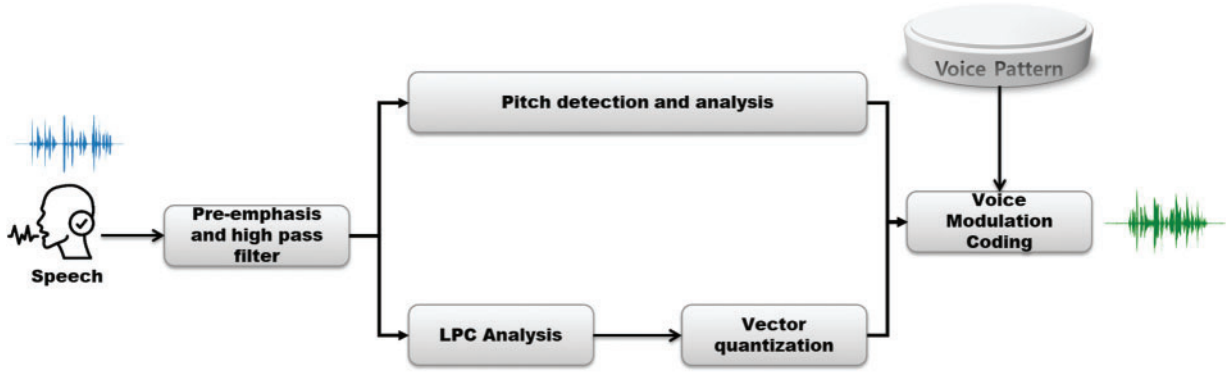

Figure 5: Voice modulation process in the proposed system

Then, four voice pattern pieces of information are read and matched before performing real-time voice modulation, which adjusts the tone and pitch of the user's voice. Finally, the resulting modulated voice is transmitted to the metaverse.

The process of conducting psychological counseling between the client and the counselor on the metaverse platform is illustrated in Fig. 6. The metaverse-based counseling can be accessed by users through PCs or mobile devices.

Figure 6: Secure counselling through the proposed metaverse platform

4 Performance Evaluations

We conducted a performance evaluation of the proposed metaverse psychological counseling system. Table 1 shows the specifications of the PC and mobile devices used in the experiment.

The proposed system underwent three performance evaluations. The first evaluation aimed to access the accuracy of facial landmarks recognition. This involved measuring the TN (True Negative) and FP (False Positive) values on images in the HELEN [21] face database. The second evaluation

measured the accuracy of real-time facial expressions of metaverse avatars based on the results of facial landmarks recognition and the corresponding application process. For this evaluation, 20 users were randomly paired and provided with the devices listed in Table 1 to engage in metaverse counseling for 30 min. Subsequently, the users evaluated the results through a questionnaire on how accurately their facial expressions were reflected in the metaverse avatar.

Table 1: Experimental environments

Environments	Value
PC	Windows 11 Pro, AMD Ryzen 7 5800X 8-Core Processor 3.80 GHz, 16 GB RAM, NVIDIA GeForce GTX 1650 Super
MacBook Pro	macOS Ventura, M2 (8-Core, 10 GPU), 16 GB RAM, 1 TB SSD, 13 inches Retina
GalaxyS10e	Android, Samsung Exynos 9820, 6 GB RAM, 2280 × 1080 display
GalaxyA53	Android, Samsung Exynos 1200, 6 GB RAM, 2400 × 1080 display
Galaxy Tab S7	Android, Qualcomm Snapdragon SM8250-AB Platform, 8 GB RAM, 2560 × 1600 display

The first experiment exhibited an accuracy of approximately 94% for the facial dataset. After the experiment, the average accuracy for users' facial recognition was 91%, slightly lower than the recognition results for facial landmarks, as the faces were not easily captured on the screen during metaverse counseling with mobile devices due to varying environments. In such situations, users perceived the facial recognition results to be somewhat poor. However, users generally found the facial recognition results satisfactory during the counseling process when they focused on camera views. Fig. 7 illustrates the face landmark recognition results and the corresponding metaverse avatar expressions of the 20 users who participated in the experiment.

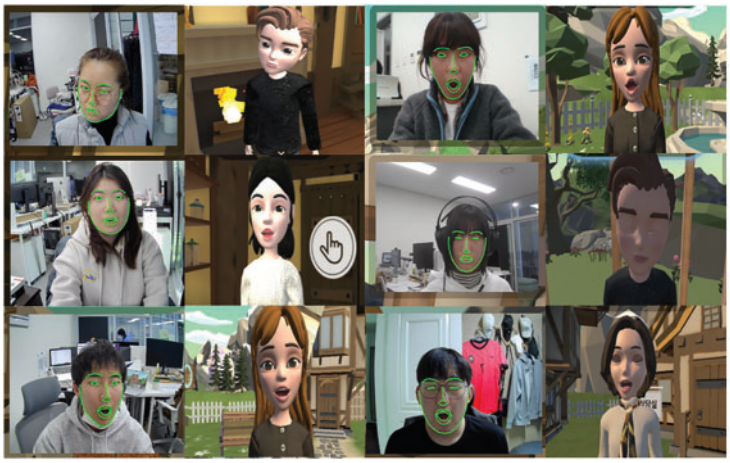

Figure 7: (Continued)

Figure 7: Results of facial expression in metaverse environment

During the 1:1 metaverse consultation in the second experiment, this paper measured three performance metrics of the proposed method. The first was the frame per second (FPS) value without facial recognition or voice modulation, the second was the FPS value with facial recognition but without voice modulation, and the third was the FPS value with both facial recognition and voice modulation. Participants in the experiment underwent a 30 min counseling session that was conducted in three stages, each lasting 10 min.

Table 2 presents the measurement results for the average FPS evaluation. The optimum performance was achieved when only using the metaverse while implementing real-time facial recognition and voice modulation, resulting in the lowest FPS performance compared to that in the other cases. However, the FPS value remained higher than the typical rendering speed of 30 FPS on mobile devices, even in the worst-performing case; thus, the real-time performance can be deemed sufficient. The results of the experiment indicate that even low-end devices such as GalaxyA53 and iPhone SE2, which

have relatively subpar device performance, are capable of producing stable performance. Table 3 lists the measurement results for the average memory size of the proposed system. The most common result for the average memory usage was real-time face recognition and voice modulation in the metaverse, which showed the highest average memory usage in all environments. However, the memory usage differed depending on the operating system environment in which the system proposed in this study is operated. Nonetheless, the average memory usage was lower in the PC operating system Windows 11 and Mac environments than in the mobile device operating system. The increase in memory usage on mobile devices, such as Android and iOS (iPhone Operating System), can be attributed to the characteristics of texture compression algorithms employed to render a 3D metaverse environment. By contrast, when real-time facial expression rendering and voice modulation were concurrently executed in the metaverse on Windows 11, the observed memory alteration was the most significant at 55.45 Mega Bytes (MB). The most minor difference, amounting to 1.73 MB, was recorded on the iPhone mini 12. The second experiment showed that real-time facial expression recognition, avatar expression technology, and voice modulation technology, as proposed in the metaverse environment in this study, remain dependable across various devices.

Table 2: Results of average FPS values

Environments	Metaverse only	Metaverse + Face landmarks	Metaverse + Face landmarks + Voice modulation
PC	657.36	587.43	555.51
Mac Book Pro	774.45	721.23	709.34
GalaxyS10e	63.4	58.74	55.21
GalaxyA53	57.12	50.23	48.32
Galaxy Tab S7	67.33	65.45	65.2
iPhone SE	64.7	62.2	61.3
iPhone SE2	54.7	53.2	50.1
iPhone mini 12	66.3	65.7	63.1

Table 3: Results of average memory size

Environments	Metaverse only	Metaverse + Face landmarks	Metaverse + Face landmarks + Voice modulation
PC	508.56 MB	509.91 MB	564.01 MB
Mac Book Pro	526.99 MB	528.35 MB	530.56 MB
GalaxyS10e	658.72 MB	659.18 MB	663.82 MB
GalaxyA53	659.15 MB	660.25 MB	662.35 MB
Galaxy Tab S7	660.41 MB	661.26 MB	663.82 MB
iPhone SE	600.56 MB	601.35 MB	603.25 MB
iPhone SE2	601.12 MB	602.54 MB	603.53 MB
iPhone mini 12	601.75 MB	602.7 MB	603.48 MB

The third experiment was conducted with 14 professional counselors holding psychological and counseling certificates to implement the non-face-to-face counseling process on the proposed metaverse system. The clients who participated in the metaverse counseling underwent 10 counseling sessions, each lasting an hour, resulting in a total of 550 sessions, typically conducted once a week. The clients utilized the metaverse counseling system proposed in this study, and strict confidentiality was maintained regarding their information. Following the completion of the experiment, interviews were conducted with the professional counselors who facilitated the metaverse counseling sessions to evaluate the identity protection and recognition of facial expressions in the proposed system in terms of user satisfaction. The professional counselors evaluated the proposed system's identity protection and facial recognition efficacy on a 5-point scale. The evaluation results are presented in Table 4. To ensure the survey's reliability, Cronbach's alpha value was calculated, yielding a result of 0.723, which exceeded the threshold of 0.6, indicating statistical significance.

Table 4: Results of the satisfaction survey on identity protection and facial recognition performance by professional counselors on a 5-point scale

Question	Value
1. Was the proposed system easy to use?	4.3
2. Is the metaverse counseling space satisfactory?	4.8
3. Did you comprehend and empathize with the client's narrative?	4.1
4. Did the client's facial expression animation help the counseling?	4.3
5. Did anonymity through voice modulation help metaverse counseling?	4.6
6. Do you intend to conduct metaverse counseling in the future?	4.67

5 Discussion

Additional user interviews were conducted with the professional counselors, revealing that they were generally satisfied with the program and the metaverse space. However, in the context of non-face-to-face counseling, counselors were advised to pay more attention to internet delays and variations in voice quality depending on the counseling location while listening to the clients' narratives. According to the majority of the counselors interviewed, the client's nonverbal expression could not be recognized during the consultation without using the client's facial expression animation with the proposed method. The real-time facial expressions of avatars using facial landmarks helped counselors recognize the nonverbal expressions of their clients. Specifically, the counselors mentioned that the proposed system has an advantage over telephone counseling and other metaverse platforms because it displays facial expressions. During the interviews, an interesting observation was made about the voice modulation function, which was widely used by the clients. All the counselors used their original voices during the consultations. Counselors believed that showing their real faces or letting clients hear their actual voices would build more trust with their clients. In contrast, clients who participated in the counseling session appreciated the anonymity provided by the non-face-to-face counseling in the metaverse, which enabled them to consult without revealing their real faces. The presence of voice modulation provided an added advantage, making clients feel more at ease when discussing personal issues that they may typically find difficult to communicate.

Interestingly, a few clients initially used the proposed voice modulation but stopped doing it during the course of 10 psychological counseling sessions. The counselors noted that as the metaverse

counseling progressed, the clients began to trust them more and feel healed, which led them to open up more and help address the clients' major issues. In response to questions about plans to continue metaverse counseling in the future, the interviewees generally expressed positive attitudes, with counselors who have difficulty securing offline counseling sites being particularly enthusiastic about continuing to participate. During the interview, feedback was provided regarding the mobile version of the system. Some counselors mentioned that the avatar was too small and requested that it be displayed in a larger format. Others suggested that the system should be improved to better recognize non-verbal expressions by incorporating more diverse facial expressions and better motion recognition. Additionally, a few counselors recommended optimizing and improving its stability due to the system's reliance on the network environment. The applicability of the proposed method can be extended to various security issues [22–26].

In a case study, a client who initially declined offline counseling due to concerns about revealing personal identity information opted for metaverse counseling. For 10 counseling sessions within the metaverse, the client successfully addressed the challenges they were facing with the counselor's assistance. Moreover, the client expressed high satisfaction with the system proposed in this study, particularly appreciating the voice modulation feature that minimized exposure. A professional counselor who worked with the client noted that the specialized metaverse psychological counseling application elicited a more positive response from the client than traditional non-face-to-face counseling methods.

Extensive efforts were made to comprehensively evaluate the proposed metaverse-based psychological counseling platform and other psychological counseling platforms employing video chat. The client's involvement in the video chat consultation process was facilitated through professional psychological counselors as mentioned in Section 4. Unfortunately, the non-face-to-face consultation sessions via video chat could not be executed due to a lack of applicants. Feedback from clients who participated in the metaverse counseling indicated the tremendous pressure they experienced when it came to reveal their faces during video chat sessions. As a result, we actively sought our clients' perspectives regarding using video chat for voice counseling, eliminating the need to reveal their facial identity. It was suggested that conversing while observing avatars in a metaverse environment offers enhanced convenience, even if facial concealment is maintained. Additionally, interviews were conducted with four professional counselors who actively participated in experiment 2 and possessed experience in utilizing various metaverse platforms in their counseling practice. The objective was to evaluate the effectiveness of the proposed system in facilitating nonverbal representation counseling. The insights obtained from these interviews revealed that alternative metaverse platforms fail to convey the non-verbal expressions of clients, thereby necessitating sole reliance on the user's voice during the counseling process. This limitation posed significant challenges in accurately perceiving the client's condition. By contrast, the proposed system enables the expression of non-verbal cues such as eye blinking, mouth movements, and facial orientation, albeit with potential for further improvement. Consequently, the counselors evaluated the proposed system as more advantageous compared to other platforms in accurately recognizing the essential elements for effective counseling.

The overall administration and management of the proposed metaverse system were entrusted to Yatav Enter, serving as the principal entity responsible for its implementation. The experiment conducted within this study involved a meticulously organized process where 14 professional counselors were paired with clients for a total of 550 metaverse consultations over a specific timeframe. Yatav Enter assumed the pivotal role of overseeing and coordinating the scheduling of counselors and clients while simultaneously addressing any operational challenges during the metaverse sessions. Establishing dedicated technical services was imperative because of the need for technical support and consultation schedules to promptly address potential errors that users might encounter when

utilizing the proposed metaverse solution. Consequently, a comprehensive website was created to facilitate efficient consultation schedule management and seamlessly integrate it with the proposed metaverse system outlined in this paper. Moreover, the participating educational institutions assumed the responsibility of resolving technical errors and conducting experimental evaluations that arose during the operation of the metaverse counseling service, further enriching the collaborative endeavor.

The proposed metaverse counseling system described in this research operates seamlessly across various environments such as Windows 11, macOS, Android, and iOS. Although the Unity engine utilized in the development of this system supports Linux builds, certain functions are incompatible with AI and network libraries, leading to its exclusion from Linux OS implementation. Alternatively, one could contemplate a metaverse based on WebGL as it offers compatibility across all operating systems. However, specific web browsers may only partially support the system despite being a web standard. Moreover, opting for a web browser-based metaverse poses challenges in utilizing the device's multi-threading technology and establishing connections with other libraries. The limitations of this study also include the absence of advanced deep-learning models for complex facial expressions and emotional recognition.

6 Conclusion

A metaverse platform for psychological counseling is proposed. The platform applies voice modulation technology to protect the client's identity and provide effective counseling. Additionally, the platform uses real-time facial expression technology to detect the non-verbal expressions of clients by recognizing facial landmarks using cameras in non-face-to-face counseling. Furthermore, the proposed metaverse psychological counseling system is optimized for mobile devices to facilitate its application in actual metaverse counseling. After evaluating its performance, the proposed system provides metaverse psychological counseling in real-time, utilizing facial expression recognition and voice modulation even on mobile devices with inferior performance. Qualified psychological counseling experts conducted large-scale metaverse counseling using this system, and the results show high satisfaction levels averaging 4.6 on a 5-point scale.

In future research, we plan to address the limitations of the proposed study. Specifically, we aim to develop and apply various deep-learning models to enable more complex expressions and emotional recognition of users' faces. Additionally, we plan to develop voice modulation technologies for enhanced identity protection. Furthermore, beyond its application in counseling, the proposed system holds promising prospects for applications in avatar-based chat systems and educational content delivery through avatars. Its utility can be extended to domains where multiple users engage in interactive discussions and educational activities within the metaverse environment.

Acknowledgement: The authors thank Sungju Kang and HyungJoong Youn (Yatav Enter) for supporting the experiments of the proposed system.

Funding Statement: This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2021RIS-004) and supported by the Technology Development Program (S3230339) funded by the Ministry of SMEs and Startups (MSS, Korea).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: J. Lee and H. Kwon; data collection: J. Lee, H. Lee and S. Lee; analysis and interpretation of

results: J. Lee, H. Lee and H. Kwon; draft manuscript preparation: J. Lee and H. Kwon. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data underlying the results presented in the study are available within the manuscript.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] C. J. Gelso and E. N. Williams, *Counseling Psychology*, fourth edition, American Psychological Association, Washington DC, USA, pp. 1–533, 2021.

[2] D. Best, H. Nicholas and M. Bradley, *Roles and Contexts in Counselling Psychology*, vol. 1. Abingdon-on-Thames, Oxfordshire, England, UK, Routledge, pp. 1–222, 2022.

[3] C. A. Bell, S. A. Crabtree, E. L. Hall and S. J. Sandage, "Research in counselling and psychotherapy post-COVID-19," *Counselling and Psychotherapy Research*, vol. 21, no. 1, pp. 3–7, 2021.

[4] J. Lee, H. Lee, S. Lee and H. Kwon, "Metaverse-based counseling system to protect the identity of clients," in *Proc. of the 6th Int. Symp. on Mobile Internet Security*, Jeju Island, South Korea, pp. 1–2, 2022.

[5] S. M. Lee and D. Lee, "Untact: A new customer service strategy in the digital age," *Service Business*, vol. 14, pp. 1–22, 2020.

[6] M. A. I. Mozumder, M. M. Sheeraz, A. Athar, S. Aich and H. C. Kim, "Overview: Technology roadmap of the future trend of metaverse based on IoT, blockchain, AI technique, and medical domain metaverse activity," in *Proc. of 24th Int. Conf. on Advanced Communication Technology (ICACT)*, PyeongChang, South Korea, pp. 256–261, 2022.

[7] M. U. A. Babu and P. Mohan, "Impact of the metaverse on the digital future: People's perspective," in *Proc. of 7th Int. Conf. on Communication and Electronics Systems (ICCES)*, Coimbatore, India, pp. 1576–1581, 2022.

[8] B. Sisman, J. Yamagishi, S. King and H. Li, "An overview of voice conversion and its challenges: From statistical modeling to deep learning," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 29, pp. 132–157, 2020.

[9] J. B. Li, S. Qu, X. Li, J. Z. Kolter and F. Metze, "Adversarial music: Real world audio adversary against wake-word detection system," in *Proc. of the 33rd Int. Conf. on Neural Information Processing Systems*, Vancouver, BC, Canada, pp. 11931–11941, 2019.

[10] Y. Jin, Z. Li and P. Yi, "Review of methods applying on facial alignment," in *Proc. of IEEE 2nd Int. Conf. on Electronic Technology, Communication and Information (ICETCI)*, Changchun, China, pp. 553–557, 2022.

[11] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. of IEEE Conf. Computer Vision Pattern Recognition*, Columbus, OH, USA, pp. 1701–1708, 2014.

[12] O. M. Parkhi, A. Vedaldi and A. Zisserman, "Deep face recognition," in *Proc. of British Machine Vision Conf.*, Swansea, UK, pp. 6–17, 2015.

[13] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1867–1874, 2014.

[14] D. Thalmann, J. Lee and N. M. Thalmann, "An evaluation of spatial presence, social presence and interactions with various 3D displays," in *Proc. of Computer Animation and Social Agents*, Nanyang, Singapore, pp. 197–204, 2016.

[15] K. Grabowski, A. Rynkiewicz, A. Lassalle, S. Baron-Cohen, B. Schuller *et al.,* "Emotional expression in psychiatric conditions: New technology for clinicians," *Psychiatry and Clinical Neurosciences*, vol. 29, pp. 132–157, 2020.

[16] S. A. Nijman, G. H. M. Pijnenborg, R. R. Vermeer, C. E. R. Zandee, D. C. Zandstra *et al.,* "Dynamic interactive social cognition training in virtual reality (DiSCoVR) for social cognition and social functioning in people with a psychotic disorder: Study protocol for a multicenter randomized controlled trial," *Schizophrenia Bulletin*, vol. 19, pp. 1–11, 2022.
[17] S. A. Nijman, W. Veling, K. G. Lord, M. Vos, C. E. R. Zandee *et al.,* "Dynamic interactive social cognition training in virtual reality (DiSCoVR) for people with a psychotic disorder: Single-group feasibility and acceptability study," *JMIR Mental Health*, vol. 7, no. 8, pp. 1–18, 2020.
[18] J. M. Morales, C. Llinares, J. Guixeres and M. Alcañiz, "Emotion recognition in immersive virtual reality: From statistics to affective computing," *Sensors*, vol. 20, no. 18, pp. 1–26, 2020.
[19] C. N. W. Geraets, S. K. Tuente, B. P. Lestestuiver, M. V. Beilen, S. A. Nijman *et al.,* "Virtual reality facial emotion recognition in social environments: An eye-tracking study," *Internet Interventions*, vol. 25, pp. 1–8, 2021.
[20] A. Howard, M. Sandler, G. Chu, L. C. Chen, B. Chen *et al.,* "Searching for MobileNetV3," in *Proc. of 2019 IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, Seoul, Korea, pp. 1314–1324, 2019.
[21] V. Le, J. Brandt, Z. Lin, L. D. Bourdev and T. S. Huang, "Interactive facial feature localization," in *Proc. of the European Conf. on Computer Vision*, Florence, Italy, pp. 679–692, 2012.
[22] J. Choi and X. Zhang, "Classifications of restricted web streaming contents based on convolutional neural network and long short-term memory (CNN-LSTM)," *Journal of Internet Services and Information Security*, vol. 12, pp. 49–62, 2022.
[23] Y. Lee and S. Woo, "Practical data acquisition and analysis method for automobile event data recorders forensics," *Journal of Internet Services and Information Security*, vol. 12, pp. 76–86, 2022.
[24] J. L. Cabra, C. Parra, D. Mendez and L. Trujillo, "Mechanisms of authentication toward habitude pattern lock and ECG: An overview," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, vol. 13, no. 2, pp. 23–67, 2022.
[25] N. Cassavia, L. Caviglione, M. Guarascio, G. Manco and M. Zuppelli, "Detection of steganographic threats targeting digital images in heterogeneous ecosystems through machine learning," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, vol. 13, no. 3, pp. 50–67, 2022.
[26] D. Pöhn and W. Hommel, "Universal identity and access management framework," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, vol. 12, no. 1, pp. 64–84, 2021.