Replacing Real Faces with Virtual Humans: A New Paradigm for Facial De-Identification

Paulo Knob
PUCRS / INCT-SANI
Porto Alegre, RS, Brazil
Email: paulo.knob@edu.pucrs.br

Erick Menezes
UNIT / INCT-SANI
Aracaju, SE, Brazil
Email: erick.marck@souunit.com.br

Rafael Mecenas
UNIT
Aracaju, SE, Brazil

Gabriel Ferri Schneider
PUCRS
Porto Alegre, RS, Brazil
Email: gabriel.ferri@edu.pucrs.br

Ana Carolina Xavier
PUCRS
Porto Alegre, RS, Brazil
Email: a.xavier004@edu.pucrs.br

Victor Araujo
UNIT / INCT-SANI
Aracaju, SE, Brazil
Email: victor.flavio93@souunit.com.br

Soraia Musse
PUCRS
Porto Alegre, RS, Brazil
Email: soraia.musse@pucrs.br

***Abstract*—Face de-identification seeks to protect individual privacy in visual data while preserving useful facial features. In this work, we propose a novel approach that replaces original facial images with realistic 3D virtual human representations. These persistent virtual faces, generated via automatic reconstruction (Deep3D) and manual tailoring (MetaHumans), serve as identity representatives, enabling face recognition without storing real biometric data. We created multiple datasets of virtual faces based on two public datasets (CFD and MEAD), while also applying transformations on a few selected virtual faces, such as age progression, skin tone variation, and emotional expression changes. Our experiments, using Adam's face recognition algorithm, show that even modified virtual faces consistently match their real counterparts with high accuracy.**

I. INTRODUCTION

In recent years, the advancement and proliferation of technology have been changing many areas. The widespread availability of digital imagery, particularly with the growing use of cameras, combined with Artificial Intelligence (AI) methods, enables the identification of various events and supports us in our daily routines, whether at home, in study, or at work. However, all this flow of personal data has raised significant concerns about individual privacy. On this topic, face de-identification has emerged as a critical research area aimed at addressing these concerns by altering facial images in a way that conceals personal identity, while preserving other useful attributes such as expression, pose, or demographic features [1]–[3].

Face de-identification techniques have evolved from traditional methods, such as pixelation and blurring, to more sophisticated approaches based on machine learning and deep generative models. For instance, the work of Zhang et al. [4] introduces a framework that disentangles identity and attribute features to balance privacy protection and data utility, named Realistic-Generation and Balanced-Utility GAN (RBGAN), which ensures that the generated faces maintain essential facial attributes while effectively concealing identity-specific information. Despite these advancements, achieving a balance between de-identification effectiveness, visual quality, and preservation of non-identifying facial attributes, remains a significant technical challenge. It is essential to note that the importance of face de-identification extends beyond privacy protection to encompass issues of ethics, legality, and fairness in the context of artificial intelligence and technological applications. Regulations, such as the General Data Protection Regulation (GDPR) [1] in Europe, the California Consumer Privacy Act (CCPA) [2] in the United States, and the General Personal Data Protection Law (LGPD) [3] in Brazil, emphasize the need for robust data anonymization techniques to safeguard personal information.

In this work, we propose a new paradigm for face de-identification by introducing 3D virtual humans as digital representatives for real individuals. Instead of storing original facial images, our method generates a realistic and persistent virtual face that represents each user. This study also introduces modifications, such as age progression and skin tone variation, to assess recognition consistency across varying conditions. The primary objective is to de-identify individuals without compromising facial recognition performance. By decoupling stored representations from real biometric data, our method enhances privacy while maintaining the functionality and accuracy of face-based recognition systems.

II. RELATED WORK

Newton et al. [5] is widely recognized as one of the first formal works on face de-identification. In their work, the authors present an algorithm, named k-Same, which aims to limit the ability of face recognition solutions to recognize faces in images while also keeping facial details. Nowadays, methods are being applied to prevent someone's personal identity from being revealed, both at the software and hardware levels. The software level is the most common, which means that the software performs some processing on images. As commented in the recent literature [6], [7], most common

[1] https://gdpr-info.eu/
[2] https://oag.ca.gov/privacy/ccpa
[3] https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

techniques include blurring, mosaicing, masking, pixelation, and replacement of the face or a given object. Although these methods can be effective in concealing a person's identity, they have several drawbacks. For instance, Jiang et al. [8] note that these methods can introduce artifacts that compromise the performance of downstream visual tasks. Generative adversarial network (GAN) based methods tend to provide a better privacy/utility trade-off compared to traditional approaches [9]. The main idea of these methods is to generate new faces to replace the original ones. Several works can be found in the literature using this approach [10], [11] More et al. [12] propose a model for video de-identification, using object detection and tracking to locate the faces. The authors use a generative adversarial network (GAN) to anonymize the located faces. It can be used with real-time video sequences and was tested on real-world datasets, such as the Flickr-Faces-HQ (FFHQ), demonstrating its effectiveness in preserving privacy while maintaining the visual integrity of the video content. While previous work on face de-identification has primarily focused on altering or replacing facial regions in 2D images through traditional techniques, such as blurring and pixelation, or more advanced approaches like GAN-based face synthesis, our method introduces a fundamentally different strategy. Instead of storing the original face images, our system generates and stores high-fidelity virtual human faces that serve as identity representatives in facial recognition systems. To the best of our knowledge, this is the first method to leverage 3D virtual humans as a de-identification mechanism, combining privacy preservation with adaptability and realism across multiple identity scenarios.

III. Proposed Method

Figure 1 presents an overview of our method. To generate a virtual face of a real person, we first need to have images of the individuals we wish to transform. Moreover, facial images should be presented in a controlled environment to avoid variations in lighting, facial orientation, and other factors. wo facial datasets were used in this work: the Chicago Face Dataset (CFD) and the Multi-view Emotional Audio-visual Dataset (MEAD), both illustrated in Figure 1.

The CFD used in this work is comprised of three datasets: the original Chicago Face Dataset (CFD) [13], and two extensions: the Multiracial expansion (CFD-MR) [14], and the India Face Set (CFD-INDIA) [15]. The main CFD set comprises 597 images of people portraying a neutral expression, along with a subset featuring varying facial expressions (i.e., happy, angry, etc.), totaling 1205 images. It includes self-identified Asian, Black, Latino, and White individuals, both male and female, from the United States. The CFD-MR extension comprises 88 images of people portraying a neutral expression, who self-reported having multiracial ancestry. Finally, the CFD-INDIA extension comprises 142 images of people with neutral expressions, recruited in Delhi, India. In turn, the Multi-view Emotional Audio-visual Dataset (MEAD) [16] comprises videos of 60 actors and actresses speaking with eight different emotions at 3 different intensity levels. The videos are captured at seven different view angles, inside a controlled environment. Being in possession of the images of these two datasets, our method can be divided into four general steps: **1)** Select images from the CFD and MEAD datasets to create R-CFD and R-MEAD, respectively. For CFD, one image per person was chosen; for MEAD, a subset was selected due to the manual creation of MetaHumans, which limited full reconstruction of all individuals; **2)** Use the images from R-CFD and R-MEAD to generate the virtual faces, which are stored inside two new datasets: V-CFD and V-MEAD, respectively; **3)** Generate modifications in the face of the models from V-MEAD, altering their facial expressions, skin tone, and age. We use these modifications in our experiments to determine if a face recognition algorithm can identify the real person, even when such modifications are applied. The new faces are stored in V-MEAD-EMOTIONS, V-MEAD-MONK, and V-MEAD-AGE, respectively; **4)** Use a Face Recognition algorithm to identify real faces (R-CFD and R-MEAD), based on virtual faces (V-CFD, V-MEAD, V-MEAD-EMOTIONS, V-MEAD-MONK, and V-MEAD-AGE).

In step 1, we select all images from the CFD dataset (including the two extensions) where the individuals were unique and presented a neutral expression (as specified by the authors), totalizing 771 images. These images comprise R-CFD. From the MEAD dataset, we select eight video sequences of different people, capturing a neutral face for each, for a total of eight images. It is worth noting that, since this dataset will be used for the manual generation method (i.e., manually tailoring a MetaHuman from a real face), fewer samples were selected compared to the CFD dataset. These images comprise R-MEAD.

In step 2, we use these images to generate virtual faces of each person. For the images selected from the R-CFD dataset, we used Deep 3D, proposed by Deng et al. [17]. It functions as an automatic 3D face reconstruction system, utilizing a Convolutional Neural Network (CNN) with weakly supervised learning. Their method is capable of reconstructing a face with only a single image. This method was used to create 771 virtual faces from R-CFD, originating the V-CFD. In its turn, for R-MEAD dataset, we used MetaHuman, from Epic Games, which is the state-of-the-art for realistic virtual humans. In this method, we selected images of real faces and manually tailored their characteristics to match those of their respective MetaHumans. The goal was to test realistic virtual humans to obtain more diverse data, allowing for comparison with the simplest visual output delivered by the automatic process in facial recognition. The shapes of the faces were extracted from the real face images and processed by the same method used for the automatic process (Deep 3D). This result is pre-processed and used inside Unreal Engine and transferred into a Metahuman, where the most similar characteristics, based on the original images, were selected within the limited options of the tool. This method was used to create eight virtual faces from R-MEAD, originating the V-MEAD.

In step 3, we modify the virtual faces from V-MEAD to generate three new datasets: V-MEAD-EMOTIONS, V-

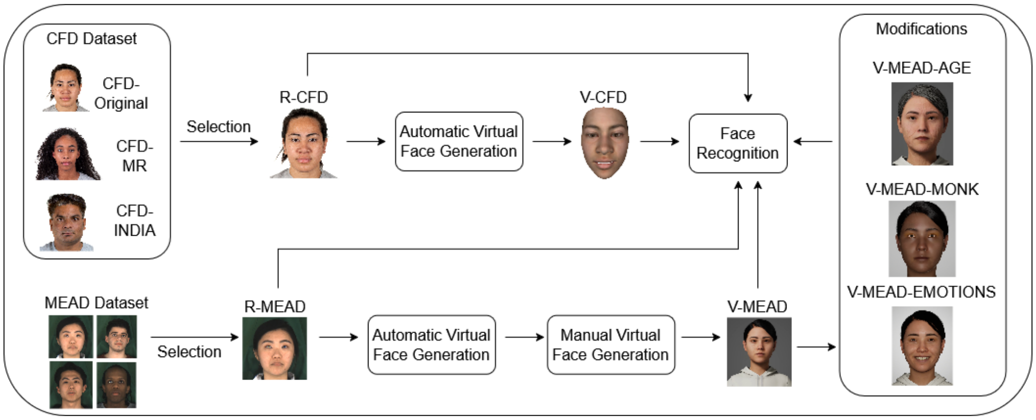

Fig. 1. Overview of our method. CFD and MEAD Datasets are used as input to generate R-CFD and R-MEAD, respectively. R-CFD is our selection from CFD and is used to generate virtual faces for V-CFD, while R-MEAD is our selection from MEAD and does the same for V-MEAD. In its turn, V-MEAD is used to generate three more datasets: V-MEAD-AGE, V-MEAD-MONK, and V-MEAD-EMOTIONS.

MEAD-MONK, and V-MEAD-AGE. V-MEAD-EMOTIONS is generated by altering the facial expression of the virtual faces (i.e., happy, sad, angry, etc.). V-MEAD-MONK is generated by altering the skin tone of the virtual faces, following the Monk Skin Tone (MST) scale [18], which is comprised of 10 levels of skin tones. V-MEAD-AGE is generated by altering the visual appearance of the virtual faces to make them look older. More details are provided in the following sections.

Fig. 2. Datasets examples. The leftmost side shows an example from R-CFD (Top) and V-CFD (Bottom). Left-center shows R-MEAD (Top) and V-MEAD (Bottom). Right-center shows V-MEAD-MONK (Top) and V-MEAD-EMOTIONS (Bottom). The rightmost side shows V-MEAD-AGE.

Finally, in step 4, we apply a Face Recognition method using real face images (R-CFD and/or R-MEAD) and virtual face images (V-CFD, V-MEAD, V-MEAD-EMOTIONS, V-MEAD-MONK, and V-MEAD-AGE). Figure 2 presents one example of each of our datasets. The face recognition method used in this work is based on the approach proposed by Adam Geitgey[4], which is capable of identifying a face using a single reference image. The model, trained with deep learning techniques, extracts facial features and computes the distance between them to assess similarity between two images. This distance ranges from 0% (identical) to 100% (completely different), with lower values indicating higher similarity. The core assumption is that, within a dataset containing multiple individuals (e.g., X, Y, Z), the image of person X will be most similar to another image of X. The method includes two primary functions: (1) a function that attempts to recognize a face by comparing it to a stored dataset, using a default distance threshold of 0.7—faces with a higher distance are considered unrecognized; and (2) a function that returns the top N most similar faces, regardless of the threshold. In this work, we adopt the second function with N = 3 for our experiments.

[4]GitHub: Face Recognition

A. Skin Tone

One of our hypotheses is that virtual human avatars, even when generated with different skin tones, should still be recognized as the same individual by face recognition systems, demonstrating their potential for identity-preserving de-identification. To achieve this, we used the 8 Metahumans from V-MEAD and modified their skin tone. Ten levels of tones were used, as suggested by Monk Skin Tone (MST) scale [18], totalizing 80 (8 MHs x 10 skin tones) virtual faces. MST provides an RGB Reference value for each of the ten skin tones. To apply this color in Unreal Engine, we manually set the base skin color of each of the 8 MHs to $(0.5, 0.5)$ inside MH Creator, generating a new texture map for each. Then, this texture map is normalized by dividing its RGB values by their average. Finally, the target RGB (provided by MST) is multiplied, creating a new texture with the desired skin tone to be applied to the modified MetaHuman within Unreal Engine, as proposed by Zhou et al. [19]. Those 80 new MHs compound a new dataset, named V-MEAD-MONK.

B. Facial Expressions

Our hypothesis here is that changes in facial expression do not prevent face recognition systems from identifying the original person through their virtual avatar. We also used the same 8 Metahumans stored in V-MEAD, modeling the six basic emotions defined by Ekman [20]. To achieve this, we used the default facial expression animations provided by MetaHuman Creator, specifically the Facial Pose Library. This tool can be imported together with a given MetaHuman inside Unreal Engine. Additionally, it provides different levels of intensity for each emotion. For standardization, we chose the third level of each emotion, as it best represents its respective emotion in our subjective evaluation. Each of the 8 MetaHumans is imported into Unreal Engine and, for each one, 6 emotional facial expressions (i.e., happy, sad, disgusted, afraid, surprised, and angry) are applied using the Facial Pose Library. In the end, 48 new MetaHumans are generated (8 MHs X 6 emotional facial features), which compound a new dataset, named V-MEAD-EMOTIONS.

C. Aging

Another hypothesis is that age-related modifications in virtual avatars do not hinder their ability to be recognized as the same individual in real images. Once again, we used the 8 Metahumans stored in V-MEAD, applying facial aging at two levels. Based on visual criteria commonly associated with aging, such as the appearance of wrinkles, changes in skin texture and roughness, and hair graying [21], the *Metahuman Creator* software was used to adjust the corresponding parameters manually. For each original character, two progressively aged versions were created: the first simulated an increase of approximately 10–15 years, and the second added 10–15 years of aging over the first version.

More information about the parameters used (such as the exact values) can be provided on demand. These adjustments were applied uniformly across all characters to maintain consistency and create a suitable dataset for subsequent recognition and perceptual experiments. The justification for the choice of such parameters goes as follows: **Skin Texture and Roughness:** Increased to simulate aging indicators such as wrinkles, loss of elasticity, and dermal irregularities [21]; **Contrast:** Enhanced to emphasize lines and facial depth, aiding the visual perception of age [21]; **Graying and Brightness:** Hair and facial hair were gradually grayed, with brightness increased to mimic typical aging-related color loss [22]; **Facial Hair Roughness:** Modified to represent coarser and thinner hair textures observed in older individuals [23].

In the end, 16 new MHs (8 MHs X 2 aging levels) were created, compounding a new dataset named V-MEAD-AGE.

IV. Experiments and Results

This work proposes generating and storing a virtual representation of a person for use in their de-identification, while maintaining data utility. Therefore, we propose some experiments aimed at verifying whether a face recognition method can accurately identify the correct person, even when the stored image is not a true representation of their actual appearance. We used the face recognition method proposed by Adam Geitgey[5], but any other available method could be used. The first set of experiments was divided into three parts, depending on the datasets that were used, as follows: **1)** Trying to identify the real faces from R-CFD, using the virtual faces from V-CFD; **2)** Trying to identify the real faces from R-MEAD, using the virtual faces from V-MEAD; **3)** Trying to identify the real faces from both R-CFD and R-MEAD, using the virtual faces from both V-CFD and V-MEAD;

Following the way the face recognition method was built, the face Rx of a person X is considered identified if the lowest calculated distance between Rx and all virtual faces V available is a virtual face Vx of the same person X. As commented in Section III, we select the first three lowest distances calculated by the face recognition algorithm, considering the image with the lowest calculated distance to be identified as the person in the real face image. For experiment 1, we used the 771 real face images selected from the CFD dataset, which comprises R-CFD, alongside 771 virtual face images from V-CFD. As commented in Section III, only the virtual face images (V-CFD) are stored in the face recognition database, while the real face images are used as input for identification. The algorithm was unable to detect the face of one real person; therefore, 770 real face images were used in the experiment. The real person was correctly identified with the counterpart virtual (i.e., as the first occurrence) in about 95% of the images (736 from 770), while only four people could not be identified inside the Top-3 best distances. Moreover, the correct person was identified as the second-best occurrence 24 times (3.12%) and as the third-best occurrence 6 times (0.78%). Table I shows how the face recognition performs in the experiment, also showing the average distances (alongside standard deviations) calculated by the face recognition algorithm, for first, second, and third occurrences.

For experiment 2, we used the eight real face images selected from the MEAD dataset, which comprises R-MEAD, alongside eight virtual face images from V-MEAD. Table I shows the results generated by the face recognition method. The real person was correctly identified for all eight images. The average distance between the real face image and its virtual counterpart was 0.55, with a standard deviation of 0.05. We also selected a second real face image from the MEAD Dataset for each person in R-MEAD, allowing us to compare the distances between real-real face recognition and real-virtual face recognition. All 8 people were found as the first occurrence, with an average distance of 0.25 and a standard deviation of 0.08. As expected, the distances between a real face image and its virtual counterpart were higher when compared with its respective second real face image.

For Experiment 3, we used all 778 real face images, from both Experiment 1 (770) and Experiment 2 (8), as well as all 778 virtual face images from both previous experiments. Table I shows the results. The real person was correctly identified (i.e., as the first occurrence) in approximately 94% of the images (738 from 778), while only 9 people could not be identified within the Top 3 best distances. Moreover, the correct person was identified as the second-best occurrence 25 times (3.21%) and as the third-best occurrence 6 times (0.77%). Finally, although we did not use the threshold value from the face recognition method (Section III), it is interesting

[5]GitHub: Face Recognition

TABLE I

RESULTS FROM THE FACE RECOGNITION FOR EXPERIMENTS 1, 2, AND 3, RESPECTIVELY, FROM TOP TO BOTTOM. FOR EXPERIMENT 1, THE REAL PERSON WAS CORRECTLY IDENTIFIED (AS THE FIRST OCCURRENCE) IN APPROXIMATELY 95% OF THE IMAGES, WHILE ONLY 4 PEOPLE COULD NOT BE IDENTIFIED WITHIN THE TOP 3 BEST DISTANCES. FOR EXPERIMENT 2, THE REAL PERSON WAS CORRECTLY IDENTIFIED FOR ALL IMAGES. FOR EXPERIMENT 3, THE REAL PERSON WAS CORRECTLY IDENTIFIED (AS THE FIRST OCCURRENCE) IN APPROXIMATELY 94% OF THE IMAGES, WHILE ONLY 9 PEOPLE COULD NOT BE IDENTIFIED WITHIN THE TOP 3 BEST DISTANCES. DISTANCES (BOTH AVERAGE (AVG) AND STANDARD DEVIATION (STD)) BETWEEN THE REAL FACE IMAGE AND ITS VIRTUAL COUNTERPART ARE ALSO PROVIDED, WHEN THEY WERE FOUND INSIDE THE TOP-3. DISTANCES RANGE BETWEEN 0 AND 1.

	Images	**First**	**Second**	**Third**	**Not Found**
Total	770	736	24	6	4
Percentage	100%	95.58%	3.12%	0.78%	0.52%
Distances (AVG)	-	0.48	0.53	0.56	-
Distances (STD)	-	0.05	0.04	0.01	-
Total	8	8	-	-	-
Percentage	100%	100%	-	-	-
Distances (AVG)	-	0.55	-	-	-
Distances (STD)	-	0.05	-	-	-
Total	778	738	25	6	9
Percentage	100%	94.85%	3.21%	0.77%	1.15%
Distances (AVG)	-	0.49	0.53	0.58	-
Distances (STD)	-	0.05	0.04	0.03	-

TABLE II

RESULTS FROM THE FACE RECOGNITION WHEN MODIFICATIONS WERE APPLIED, FROM TOP TO BOTTOM: SKIN TONE, EXPRESSIONS, AGE. EXTREME SCALES (MONK 1 AND MONK 10) WERE MORE CHALLENGING, WHILE MONK 6, 7, AND 8 PRESENTED THE BEST PERFORMANCES. CONCERNING EXPRESSIONS, MOST OF THE EXPRESSIONS YIELDED A PERFECT SCORE, WHILE THE SAD EXPRESSION PRESENTED THE WORST SCORE. PERFECT SCORES WERE ACHIEVED IN THE AGING CONDITION. ALL SCALES FOR SKIN TONE, EXPRESSIONS, AND AGE, WHEN TOGETHER, ALSO PRESENTED A PERFECT SCORE (ITEMS IN BOLD).

	First	**Second**	**Third**	**Not Found**
Monk 1	5	2	1	0
Monk 2	6	1	1	0
Monk 3	6	2	0	0
Monk 4	6	2	0	0
Monk 5	6	2	0	0
Monk 6	8	0	0	0
Monk 7	8	0	0	0
Monk 8	8	0	0	0
Monk 9	6	2	0	0
Monk 10	5	3	0	0
All Monk levels	8	0	0	0
Happy	8	0	0	0
Sad	5	2	1	0
Disgusted	8	0	0	0
Surprised	8	0	0	0
Angry	7	1	0	0
Afraid	8	0	0	0
All Expressions	8	0	0	0
First age level	8	0	0	0
Second age level	8	0	0	0
Both Age levels	8	0	0	0

to cite that all first occurrences were found below the default threshold value (i.e., 0.7), with the maximum distance found being equal to 0.63.

A. Skin Tone Analysis

As previously commented, we hypothesize that a person's face should be identifiable, even when modifications are made to its skin tone. This experiment was conducted to verify whether the face recognition method can identify the correct person, even when their face has a different skin tone, and to investigate possible skin tone bias. The face recognition was run with each monk scale: first, we used only the virtual faces generated for Monk 1; then, only the virtual faces for Monk 2; and so on, a total of 10 times (10 skin tones). In the end, we also run the recognition one more time, using all 80 virtual faces from V-MEAD-MONK.

Table II presents the results obtained with the 11 runs. It is noticeable that the extremes of the Monk scale (Monk 1 and Monk 10) were the most challenging for facial recognition, with each model correctly guessing only 5 out of 8 people in the first occurrence. Following this trend, it appears that the upper middle scales (Monk 6, Monk 7, Monk 8) were better suited for recognition, as they identified all people as the first occurrence. The reduced recognition performance at the extremes of the Monk scale (Monk 1 and Monk 10) suggests a potential bias in the facial recognition system, raising concerns about fairness and equity in systems trained predominantly on mid-range skin tones. Using all Monk scale levels improved recognition accuracy, as having more virtual face variations for each person increased the chances of correctly identifying them, even with more competing faces from other individuals.

B. Facial Expressions Impact

As commented before, we also hypothesize that a person's face should be identifiable, even when its facial expressions are modified. This experiment was conducted to verify whether a face recognition method can identify the correct person, even when the person presents different facial expressions. The face recognition was run with each facial expression: first, we used only the virtual faces generated for the happy expression; then, only the virtual faces for the sad expression; and so on, a total of six times (six facial expressions). In the end, we also run the recognition one more time, using all 48 virtual faces from V-MEAD-EMOTIONS.

Table II presents the results obtained with the 7 runs. It is possible to notice that all people were correctly identified, as in the first occurrence, for most of the expressions (Happy, Disgusted, Surprised, Afraid). Sad expression was the most challenging, with only 5 correct first occurrences. As observed in the skin tone experiment, using all expressions also yielded promising results, enabling the identification of all people in the first occurrence.

C. Aging Impact

Finally, we hypothesize that a person's face should be identifiable, even when the person appears to be older. This experiment was conducted to verify whether the face recognition method can identify the correct person under aging modification. The face recognition was run with each of the two age levels: first, we used only the first age level, and then only the second age level. In the end, we also run the recognition one more time, using all 16 virtual faces from V-MEAD-AGE.

Table II presents the results obtained with the 3 runs. It is noticeable that both age levels achieved a perfect score, correctly identifying all individuals on the first attempt. Using both levels combined also yielded a perfect score. Finally, we also ran the face recognition method with all MHs combined (i.e., V-MEAD, V-MEAD-MONK, V-MEAD-EMOTIONS, V-MEAD-AGE), which also yielded a perfect score, with all 8 people being found as the first occurrence.

V. Final Considerations

This work introduces a new paradigm for face de-identification: instead of using a person's real face, we generate and store a virtual representation, which is then used for de-identification while preserving data utility. Although the real face is not retained, its virtual counterpart maintains facial characteristics that allow for identity preservation in recognition systems. Our findings demonstrate that identity can be consistently preserved even when the virtual face undergoes changes in appearance, such as variations in skin tone, facial expression, and age. These results highlight the potential of realistic avatars as traceable and privacy-aware representations in applications involving facial data. The proposed approach yielded promising outcomes, with the face recognition system correctly identifying most individuals despite the visual modifications applied to their avatars. One current limitation is the small number of images available in V-MEAD, due to the lack of an automated method for generating MetaHuman avatars from real images. Additionally, further investigation is needed to assess how our virtual identities interact with different face recognition algorithms and to ensure that the avatar generation process does not encode or amplify existing biases in the underlying data.

Acknowledgment

This study was partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior – Brazil (CAPES) – Finance Code 001; by the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (CNPq) - Process Numbers 309228/2021-2; 406463/2022-0; 153641/2024-0.

References

[1] T. Orekondy, M. Fritz, and B. Schiele, "Connecting pixels to privacy and utility: Automatic redaction of private information in images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8466–8475.

[2] K. Brkic, I. Sikiric, T. Hrkac, and Z. Kalafatic, "I know that person: Generative full body and face de-identification of people in images," in *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*. IEEE, 2017, pp. 1319–1328.

[3] R. Hasan, P. Shaffer, D. Crandall, E. T. Apu Kapadia *et al.*, "Cartooning for enhanced privacy in lifelogging and streaming videos," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 29–38.

[4] Y. Zhang, Y. Fang, Y. Cao, and J. Wu, "Rbgan: Realistic-generation and balanced-utility gan for face de-identification," *Image and Vision Computing*, vol. 141, p. 104868, 2024.

[5] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.

[6] S. Ravi, P. Climent-Pérez, and F. Florez-Revuelta, "A review on visual privacy preservation techniques for active and assisted living," *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 14 715–14 755, 2024.

[7] R. Zhao, Y. Zhang, T. Wang, W. Wen, Y. Xiang, and X. Cao, "Visual content privacy protection: A survey," *ACM Computing Surveys*, vol. 57, no. 5, pp. 1–36, 2025.

[8] B. Jiang, B. Bai, H. Lin, Y. Wang, Y. Guo, and L. Fang, "Dartblur: Privacy preservation with detection artifact suppression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 16 479–16 488.

[9] J. Lopez, C. Hinojosa, H. Arguello, and B. Ghanem, "Privacy-preserving optics for enhancing protection in face de-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 12 120–12 129.

[10] S. Barattin, C. Tzelepis, I. Patras, and N. Sebe, "Attribute-preserving face dataset anonymization via latent code optimization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 8001–8010.

[11] J. Cao, B. Liu, Y. Wen, R. Xie, and L. Song, "Personalized and invertible face de-identification by disentangled identity information manipulation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 3334–3342.

[12] R. More, A. Maity, G. Kambli, and S. Ambadekar, "Privacy-preserving video analytics through gan-based face de-identification," in *2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)*. IEEE, 2024, pp. 1–6.

[13] D. S. Ma, J. Correll, and B. Wittenbrink, "The chicago face database: A free stimulus set of faces and norming data," *Behavior research methods*, vol. 47, pp. 1122–1135, 2015.

[14] D. S. Ma, J. Kantner, and B. Wittenbrink, "Chicago face database: Multiracial expansion," *Behavior Research Methods*, vol. 53, pp. 1289–1300, 2021.

[15] A. Lakshmi, B. Wittenbrink, J. Correll, and D. S. Ma, "The india face set: International and cultural boundaries impact face impressions and perceptions of category membership," *Frontiers in psychology*, vol. 12, p. 627678, 2021.

[16] K. Wang, Q. Wu, L. Song, Z. Yang, W. Wu, C. Qian, R. He, Y. Qiao, and C. C. Loy, "Mead: A large-scale audio-visual dataset for emotional talking-face generation," in *ECCV*, 2020.

[17] Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, and X. Tong, "Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set," in *IEEE Computer Vision and Pattern Recognition Workshops*, 2019.

[18] C. Schumann, G. O. Olanubi, A. Wright, E. Monk Jr, C. Heldreth, and S. Ricco, "Consensus and subjectivity of skin tone annotation for ml fairness," *arXiv preprint arXiv:2305.09073*, 2023.

[19] C. Li, K. Zhou, and S. Lin, "Intrinsic face image decomposition with human face priors," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 218–233.

[20] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[21] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 4352–4360.

[22] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 365–372.

[23] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3730–3738.