Semi-Automated Digital Human Production for Enhanced Media Broadcasting

Semi-Automated Digital Human Production for Enhanced Media Broadcasting

Miriana Martini
Dipartimento di Automatica e Informatica (DAUIN) Politecnico di Torino
Turin, Italy
miriana.martini@studenti.polito.it

Valeria Valentini
Dipartimento di Automatica e Informatica (DAUIN) Politecnico di Torino
Turin, Italy
valeria.valentini@studenti.polito.it

Alberto Ciprian
Centre for Research, Technological Innovation and Experimentation Rai Radiotelevisione Italiana
Turin, Italy
alberto.ciprian@rai.it

Andrea Bottino
Dipartimento di Automatica e Informatica (DAUIN) Politecnico di Torino
Turin, Italy
andrea.bottino@polito.it

Roberto Iacoviello
Centre for Research, Technological Innovation and Experimentation Rai Radiotelevisione Italiana
Turin, Italy
roberto.iacoviello@rai.it

Maurizio Montagnuolo
Centre for Research, Technological Innovation and Experimentation Rai Radiotelevisione Italiana
Turin, Italy
maurizio.montagnulo@rai.it

Alberto Messina
Centre for Research, Technological Innovation and Experimentation Rai Radiotelevisione Italiana
Turin, Italy
alberto.messina@rai.it

Francesco Strada
Dipartimento di Automatica e Informatica (DAUIN) Politecnico di Torino
Turin, Italy
francesco.strada@polito.it

Davide Zappia
Centre for Research, Technological Innovation and Experimentation Rai Radiotelevisione Italiana
Turin, Italy
davide.zappia@rai.it

***Abstract*—In recent years, the application of synthetic humans in various fields has attracted considerable attention, leading to extensive exploration of their integration into the Metaverse and virtual production environments. This work presents a semi-automated approach that aims to find a fair trade-off between high-quality outputs and efficient production times. The project focuses on the Rai photo and video archives to find images of target characters for texturing and 3D reconstruction with the goal of reviving Rai's 2D footage and enhance the media experience. A key aspect of this study is to minimize the human intervention, ensuring an efficient, flexible, and scalable creation process. In this work, the improvements have been distributed among different stages of the digital human creation process, starting with the generation of 3D head meshes from 2D images of the reference character and then moving on to the generation, using a Diffusion model, of suitable images for texture development. These assets are then integrated into the Unreal Engine, where a custom widget facilitates posing, rendering, and texturing of Synthetic Humans models. Finally, an in-depth quantitative comparison and subjective tests were carried out between the original character images and the rendered synthetic humans, confirming the validity of the approach.**

***Index Terms*—AI automation, face reconstruction, Synthetic Humans, media archive, Generative AI**

I. Introduction

Throughout history, the concept of artificial human beings has been a constant in the human imagination, emerging as a recurring theme fascinating different cultures and eras. From ancient mythological stories, such as the legend of Pygmalion, through the ground-breaking pages of Shelley's Frankenstein, the depiction of artificial humans has spanned time, coming to permeate modern science fiction and products of the entertainment industry. These narratives and representations have played a crucial role, continually shaping and reformulating the way we perceive and imagine artificial humanity [1]. Nowadays the advent of Synthetic Humans has intensified the ambition to create artificial human beings, thanks to the opportunities offered by the contemporary technological landscape.[1] Synthetic Humans, also known as Digital Twins of Humans, Embodied Conversational Agents, Virtual Agents or simply Avatars, are essentially digital representations of real individuals. These are not limited to merely reproducing physical appearance, facial expressions and body movements of a person, but, depending on their purpose, they can also extend to the reproduction of human's peculiarities that make each individual unique, such as personality, sensitivity, thoughts

This work been partially funded by the EU Horizon 2020 research and innovation programme under grant agreement No. 101070250 (XReco).

[1] https://www.linkedin.com/pulse/synthetic-humans-great-unknowns-2023-dr-mark-van-rijmenam/ (last accessed March 2024)

and abilities.[2] Synthetic entities arise from the convergence of several disciplines, including computer graphics, computer vision, and artificial intelligence (AI). Together with the use of Generative AI, it is possible to train them using specific information and data, generating a style or a "tone of voice" that reflects the user's needs, ensuring in this way a personalized and relevant communication [1].

Virtual production is an important area in which Synthetic Humans are valuable by simplifying and optimizing the creative process. The TV series "The Mandalorian" is an example of how the pre-visualization of scenes using digital actors can improve the efficiency of production.[3] Furthemore, in the broadcast industry the inclusion of highly realistic digital characters is redefining how content is created, distributed, and perceived, offering new levels of audience engagement and participation. As television broadcasters and news platforms experiment with the use of digital avatars to conduct specific programmes and segments, virtual interviews can be conducted, in which a real person converses with a synthetic character.[4] This proves particularly valuable for interviewing guests from geographically distant places. In the context of livestreaming and real-time conversations, Synthetic Humans open new frontiers of interaction.[5]

This paper presents a semi-automated approach for the creation of photorealistic three-dimensional (3D) head mesh with the main goal of significantly reducing the need for human intervention and ensuring an efficient and flexible creation process that can operate at scale. We have developed a script for the generation of 3D head meshes from two-dimensional (2D) input images of the reference character using a Blender plugin and combined with the generation of suitable images for texture using Stable Diffusion, conditioned on the fine-tuning of the trained models. Finally, we subjectively and objectively evaluated the generated synthetic humans, focusing on fidelity and similarity to the original characters.

II. RELATED WORK

In recent years, significant studies have been conducted in the realm of 3D reconstruction of human models. The 3D reconstruction of digital models of humans poses a complex challenge within the fields of computer vision and graphics. The primary objective is to accurately retrieve the geometry and appearance of humans from visual sources such as images, videos, or depth data, precisely translating 2D information into detailed and realistic 3D representations of human bodies [2]. The traditional reconstruction pipeline is mostly built upon large capture systems in order to guarantee the high standard of the resulting model [3]. The complex capture systems and costly devices employed in the traditional reconstruction process pose significant obstacles to the broader implementation of human reconstruction. These challenges have served as a driving force behind the emergence of cost-effective learning-based methods. The objective of learning-based human reconstruction is to reconstruct 3D human models solely from single or sparsely-viewed 2D images, leveraging human priors acquired from existing data. A notable line of research involves SMPL (Skinned Multi-Person Linear Model)-based methods. SMPL [4] is a statistical deformable human model capable of adapting to the shape and pose of various individuals by adjusting a set of parameters. These approaches typically utilize a geometric prior, leveraging its inherent low dimensionality to generate and animate coarse body meshes from images. Some methods have started exploring alternative representations such as voxels in pursuit of finer geometry [5]. Regrettably, voxels suffer from significant memory consumption. Therefore, implicit functions have garnered increasing interest due to their ability to model intricate details and storage efficiency [6]. However, significant challenges remain, such as the limited generalization capacity of trained neural networks and the consistent need for extensive annotations or 3D ground-truth datasets. Recently, certain research endeavours have aimed to directly reconstruct 3D scene representations from images without relying on 3D ground-truth supervision. A notable milestone in this field is the neural radiance field (NeRF) proposed by [7], which achieves remarkable quality in synthesizing novel views of static scenes using relatively dense input views.

III. METHOD

Our goal is to create a Synthetic Human that closely resembles the target character while also maximizing the degree of automation. The highest level of automation is represented by the ability to use images of a target character as input and obtain the Synthetic Human 3D model as output. The following subsections describe the workflow in detail.

A. Preliminary Settings

First, we query the Rai's archives[6] to search for video frames depicting the target character. Next, we filter out uninformative frames, i.e., those that don't exclusively show the target character, avoiding the existence of other individuals in the same frame. Currently this is a manual process. In this phase, it is essential to select the character's face from different orientations to ensure that the reconstruction of the head mesh accurately reflects the character. These frames are often of low quality, characterised by low resolution and noise, which can affect the accuracy of 3D reconstruction. We used state-of-the-art super-resolution software[7] to ensure that the selected

[2]https://medium.com/@cognidownunder/the-blurring-lines-of-reality-digital-humans-and-digital-twins-be9709b983cc (last accessed March 2024)

[3]THE THIRD FLOOR, Inc. Virtual visualization series – The Mandalorian. https://thethirdfloorinc.com/4206/virtual-visualization-series-the-mandalorian/ (last accessed March 2024)

[4]Amelia Tait. 'Here is the news. you can't stop us': Ai anchor zae-in grants us an interview. https://www.theguardian.com/tv-and-radio/2023/oct/20/here-is-the-news-you-cant-stop-us-ai-anchor-zae-in-grants-us-an-interview (last accessed March 2024)

[5]https://www.prnewswire.com/il/news-releases/d-id-brings-ai-chatbots-to-life-enabling-real-time-conversation-with-digital-humans-301758728.html (last accessed March 2024)

[6]https://www.raicom.rai.it/en/archives-and-production-services/ (Last accessed March 2024)

[7]https://github.com/TencentARC/GFPGAN (Last accessed March 2024)

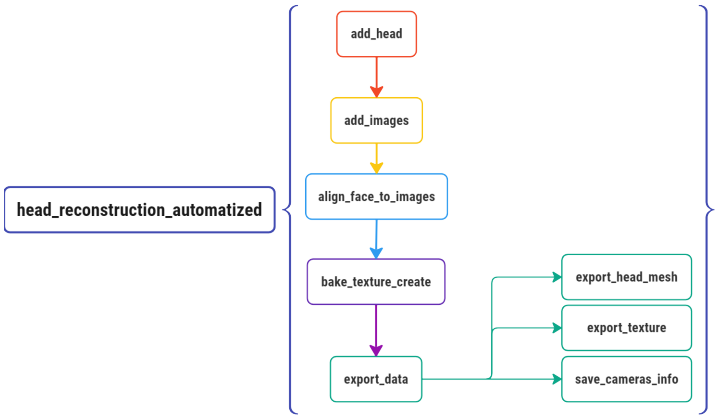

Fig. 1: Workflow of the 3D face reconstruction phase.

frames retain all essential details, thus contributing to a more accurate and realistic 3D face reconstruction.

B. 3D Face Reconstruction

In this phase we transform 2D images into 3D models of the faces of our reference characters. To simplify and automate this complex process, we developed a Python script, which leverages some of the powerful features of FaceBuilder, a Blender plugin created by KeenTools.[8] The script functionalities (see Fig. 1) include inserting images, aligning virtual cameras for each image, creating a texture in accordance with MetaHuman-specific UV mapping and creating the head mesh.[9] The script is fully automated, ensuring that all the features of Blender and FaceBuilder are used without requiring any additional user interaction. The script aligns the virtual cameras associated with each image so that they are oriented, positioned and set with the necessary focal length to achieve a match between the images and the 3D model mesh `(align_face_to_images)`. Alignment is a critical process to ensure the quality of the final result. Next, the script handles the creation of the texture of the 3D head model `(bake_texture_create)`. Before starting the texture creation, some optimizations are made to ensure a more photorealistic and accurate appearance. These optimizations include excluding the areas related to the eyeballs, the inside of the mouth, and the lower part of the neck from the mesh and consequently from the texture. Finally, the script outputs a JSON file `(export_data)` that contains the cameras info that are used to make a comparison between the MetaHuman obtained at the end of the workflow, and the source images with which it was generated.

C. From 3D Model to MetaHuman

The next step in our process is a manual step to transform the 3D models obtained from the 3D face reconstruction phase into MetaHuman. It allows to transform a customised mesh into a complete MetaHuman with skeleton. For this purpose, we import the character mesh into the Unreal Engine project. This results in a textured mesh ready for the subsequent steps in the "Mesh to MetaHuman" pipeline. In creating realistic 3D models of human faces the texture of a 3D human face is critical in order to make a Synthetic Human look realistic and closely resemble the real individual they are intended to represent.

D. Generating Textures for Synthetic Humans

As previously described, the 3D Face Reconstruction phase relies on the texture extraction algorithm provided by FaceBuilder. This algorithm integrates camera views, aligning the 3D facial mesh with previously uploaded photos, and generates the texture by projecting each pixel of the 3D mesh's UV map onto the photo according to the model's position. This process is repeated for each uploaded photo, and therefore, each camera aligned with the 3D model. Our script significantly streamlines the texture creation process, however, there are issues related to the resulting textures that make those obtained via the FaceBuilder algorithm unsuitable for direct use on Synthetic Humans. These issues can be attributed to two main factors:

- Quality of reference images: creating a high-quality texture with FaceBuilder necessitates exceptionally high-quality reference images. These images should be sharp, high-resolution, and captured under consistent lighting conditions. Blurry, low-resolution, or differently lit images can lead to artifacts and colour inconsistencies.
- Availability of various facial orientations: the images used for the creation of the 3D mesh need to cover various angles of the individual's face to align virtual cameras with the 3D model. Missing some facial angles can lead to distorted and irregular textures.

E. Hybrid Approach: Combining Texture from FaceBuilder and Metahuman Creator

To overcome the above issues and achieve a balance between specific facial details and overall skin realism algorithm we adopted an hybrid approach by integrating its results with the features of MetaHuman Creator. We align the specific facial details from FaceBuilder with the overall skin realism from MetaHuman Creator. In our study, we compare two solutions: we employ sophisticated graphic tools, including Adobe Photoshop, that enable the superimposition of textures acquired from FaceBuilder and those generated by MetaHuman Creator, following a sequence of manual procedures. Alternatively, we have developed an in-house Blender project that streamlines the semi-automatic fusion of these two textures. It's important to note that, compared to using Photoshop, which offers more detailed control, the Blender approach is more automated, but precision might be slightly lower.

F. Adding Images With Generative AI

In this phase, we address challenges and objectives such as achieving high-quality textures, maintaining uniform lighting

[8]Facebuilder for blender guide. https://medium.com/keentools/facebuilder-for-blender-guide-cbb10c717f7c (last accessed March 2024)

[9]Unreal Engine. Delivering high-quality facial animation in minutes: Metahuman animator. https://www.unrealengine.com/en-US/blog/delivering-high-quality-facial-animation-in-minutes-metahuman-animator-is-now-available (last accessed March 2024)

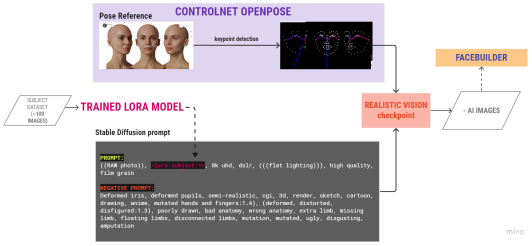

Fig. 2: Workflow for creating AI images of Maria Callas with Stable Diffusion.

Fig. 3: Texture obtained from original images (left) and AI images (right).

conditions, accommodating various orientations of the character, and handling character age. To tackle these challenges, we propose an innovative solution: leveraging AI generated images. These images cover diverse angles and lighting conditions, enabling the direct creation of high-quality textures using FaceBuilder. We have addressed three key aspects: usage of Stable Diffusion,[10] the integration of ControlNet [8] with OpenPose[11] for pose definition, and the finetuning of generative models with LoRA [9]. Stable Diffusion is a text-to-image model, meaning that when provided with a text prompt, it generates an image that corresponds to the given text. ControlNet can condition the noise predictor with information like human poses, achieving excellent control over image generation. For our purpose, we have focused on human pose detection using OpenPose, a keypoint detection model for the human body, capable of extracting positions of hands, legs, and the head. Keypoints are extracted from the input image using OpenPose and saved as a control map containing the keypoint positions. These are then provided to Stable Diffusion as an additional condition, alongside the text prompt. Overall, image generation through AI, using the mentioned techniques allows the generation of subject images in the same poses given by the reference image. Moreover, we have fine-tuned Stable Diffusion using LoRA, to introduce new concepts in the models (i.e., faces of known individuals), enabling them to generate images incorporating these concepts. To illustrate how we have applied the aforementioned techniques, let's consider an example. We first finetuned Stable Diffusion through LoRA using a dataset consisting of 100 photos of Maria Callas and leveraging the "Realistic Vision V5.1" checkpoint. We have then used the resulting model (see Fig. 2), defining a set of desired features for the resulting image in the textual prompt. To ensure the generation of the output image with the subject in the specific desired poses, we leverage reference images depicting any subject in the same poses we want for Maria Callas. At the end of this process, we obtain an image of Maria Callas with all the characteristics defined in the prompt, in high quality and in the desired pose (see Fig. 3).

Once the AI-generated images are obtained, they serve as a starting point to FaceBuilder's Texture Reconstruction algorithm. Textures derived from these images exhibit consistent lighting uniformity and reconstruction coherence, providing an initial high-quality texture that requires minimal adjustments for readiness on a Synthetic Human.

IV. Evaluation of the Proposed Approach

This Section describes the evaluation and analysis of the proposed approach. The analysis focuses on the fidelity and similarity of the Synthetic Humans to the reference characters. This evaluation has been conducted using both objective and subjective methods by comparing the original images depicting the characters with the resulting Synthetic Humans.

To conduct a meaningful evaluation of the proposed approach, we defined a representative dataset of nine TV celebrities. The dataset was diversified to include individuals with different demographic attributes such as ethnicity, age and gender. The diversification of these features aims to demonstrate that the workflow can generate Synthetic Humans reflecting the traits of a broad spectrum of individuals in the global population. For each celebrity, we collected 80 to 100 pictures, depicting them in a range of poses and expressions, from close-ups to long shots, without strict limitations on the variety of images. Those images were utilized in finetuning a Stable Diffusion model.

We have used our script to automatically generate the 3D model of the individual's face, utilizing the character's original images and images generated by the Generative AI. This head mesh is a base for creating a MetaHuman (MH) into Unreal. It is important to note that 3D head models are subject to change as a result of the "Mesh to Metahuman" transformation. In this process, a 3D face mesh is transformed into a MetaHuman in the Unreal Engine, where the original geometry is adapted to the structure of the MetaHuman based on the preset customization options. Such changes to the facial geometry may result in a loss of characteristic features. After completing the "Mesh to Metahuman" workflow, the MetaHuman texture, hereinafter referred to as Original MH, can be exported. In addition to Original MH, two other textures – Original FB and Original AI – are derived from earlier stages. These three textures form the basis for the subsequent texture reconstruction phase. The base textures are as follows:

- Original MH: The base texture provided by the MetaHuman framework, designed for high compatibility with its models.

[10]CompVis. Stable diffusion. GitHub repository, n.d. https://github.com/CompVis/stable-diffusion. (last accessed March 2024)

[11][CMU-Perceptual-Computing-Lab. Openpose. GitHub repository, https://github.com/CMU-Perceptual-Computing-Lab/openpose (last accessed March 2024)

TABLE I: TEXTURE BLENDING METHODS OVERVIEW

Textures Generated from Original (orig.) Images			
Name	***Blending Method***	***Texture A***	***Texture B***
Orig. Blender Mix	Blender	Orig. FB	Orig. MH
Orig. Photoshop Mix	Photoshop	Orig. FB	Orig. MH
Textures Generated from AI Images			
Procreate AI	Procreate	Orig. MH	Orig. AI
Blender AI	Blender	Orig. MH	Orig. AI
Photoshop AI	Photoshop	Orig. MH	Orig. AI

- Original FB: Textures obtained from FaceBuilder mapping using the character's images taken from the Rai archive.
- Original AI: Textures generated by FaceBuilder mapping using at least fifteen images of each character generated by Stable Diffusion.

After obtaining the MH, several facial textures were generated through a series of blending operations. The blending process involved combining the base textures to address any gaps and ensure a cohesive appearance. Table I details the blending methods and the specific textures combinations.

During the testing phase, all obtained textures were used to objectively and subjectively evaluate which texture generation procedure is more effective in ensuring similarity with the target character. We conducted a survey with 46 participants. Moreover, the effectiveness of the system in generating Synthetic Humans was measured using the Cosine similarity, widely used to evaluate face recognition methods. ArcFace [10] and Retinaface [11] are used to extract a 512-dimensional feature array (i.e., face embedding) and 106 pairs of (x,y) coordinates array (i.e., face landmark), respectively.

V. Test Results

First, we selected the best picture generated through Stable Diffusion and compared with the corresponding original images of a character. Then we compared the two pictures using an objective measure using the Cosine similarity and subjective user evaluation (see Fig. 4). The cosine similarity was consistently above 0.5, the threshold value that normally indicates that two people are the same person [10]. We asked the participants to "Assign a score from 1 to 5 to express how much you think the AI-generated image resembles the original character". The score of 5, which represents the highest level of resemblance, was given by 68% of the respondents. This demonstrates the remarkable capability of the generative AI in producing images that closely match the original characters in appearance.

The second test examined the validity of the 3D meshes obtained during the automated face reconstruction phase, using our script. To conduct this evaluation, we have focused on the comparison between an original image of the subject and the rendering of the 3D model from the same viewpoint. The Cosine similarity was used to measure the affinity between the rendered 2D image of the 3D head mesh (with Procreate AI texture) and the original images of the character (see Fig. 5). As before, we asked the participants to "Assign a score from 1 to 5 to express how faithful you think the 3D model of the face is to the original character". The Cosine similarity scores between the original 2D images and their corresponding 3D mesh reconstructions generally approached 0.5, indicating a reasonable resemblance. The slight decrease in cosine similarity when comparing the original 2D images with rendered 3D meshes can be attributed to several factors such as pose, skin reflectance, lighting and finer scale surface details. Moreover, the original images are not always captured in a frontal pose, thus impacting the direct comparison. Despite these challenges, the 3D meshes maintain a high level of fidelity to the original characters, as confirmed by the subjective evaluation, getting a score greater than 4 for 64% of the participants. This demonstrates that a significant majority of respondents recognize a strong likeness between the 3D models and the original images.

Fig. 4: Cosine similarity between the ArcFace embeddings.

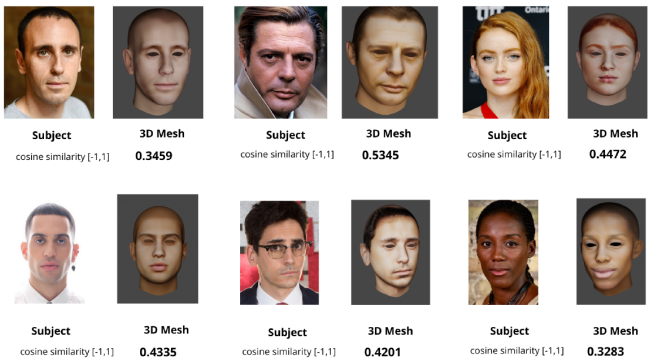

Fig. 5: Cosine similarity between 3D meshes and photos.

The aim of the third test was to determine which texture creation method best preserves the similarity of the reference subject in the MetaHuman. As the textures are not fully refined in their original state, additional processing is required to improve their quality and achieve the desired level of realism. Thus, three main techniques to blend these textures together are employed:

- Photoshop: this method requires user input to adjust the layer blending modes and opacity to achieve the seamless integration of textures.

TABLE II: Original picture Vs. MH based on original picture

Person	*Or. MH*	*Or. FB*	*Or. Blender Mix*	*Or. Photoshop Mix*
(1)	0.1038	0.2688	0.1847	0.0852
(2)	0.0832	0.3444	0.2929	0.1939
(3)	0.0702	0.1760	0.1145	0.0950
(4)	0.0311	0.3166	0.1874	0.0651
(5)	0.0687	0.3442	0.2447	0.1005
(6)	0.0517	0.4258	0.3413	0.1271
(7)	0.1194	0.2930	0.1879	0.0983
(8)	0.1556	0.3447	0.3306	0.1377
(9)	0.0359	0.3108	0.2166	0.0201
Mean	0.0800	**0.3138**	0.2334	0.1025

TABLE III: Original picture Vs. MH based on synthetic picture

Person	*Original AI*	*Procreate AI*	*Blender AI*	*Photoshop AI*
(1)	0.1539	0.2084	0.1292	0.0845
(2)	0.3155	0.3333	0.2857	0.1989
(3)	0.1125	0.0957	0.0649	0.1285
(4)	0.2565	0.2110	0.1103	0.0912
(5)	0.5506	0.4180	0.3820	0.2940
(6)	0.4032	0.3781	0.3150	0.1524
(7)	0.2384	0.3095	0.1737	0.1249
(8)	0.3127	0.3408	0.2377	0.1391
(9)	0.2469	0.1945	0.1453	0.0297
Mean	**0.2878**	**0.2774**	0.2049	0.1381

- Procreate:[12] uses the app's touch interface and advanced brush system, which, when used with a stylus on a tablet, gives a tactile approach to texture refinement.
- Blender: this employs Blender's node-based compositing system to combine textures in a semi-automatic approach.

The goal is to mix the FB Original and AI Original textures with the MH Original one to create refined textures for the Metahuman model.

As highlighted in Tables II and III, the Cosine similarity between the subject's original image and the facial texture blending techniques applied to the MetaHuman—using both original images (Table II) and AI-generated images (Table III)—is not always good. One of the reasons is certainly the loss of information about face geometry once the MetaHuman is generated, as anticipated earlier. The Original FB (Table II) textures, stands out with an average Cosine similarity value of 0.31, which keeps most of information from the original photos taken from Rai archives, but results not particularly realistic as it is a collage of multiple photos. The Procreate AI and Original AI (Table III) also show noteworthy performance, with average Cosine similarity of 0.27 and 0.28, respectively. The latter contains all the information obtained from AI-generated images mapping, while the other one shows some inaccuracies that are fixed manually on Procreate. Participants were asked to "Select one (best) to seven (all) images that you think represent a realistic facial texture from the available options". Photoshop AI scored 26.52% and Procreate AI 29.86%. This indicates that despite the initial higher Cosine similarity of Original FB textures (0.3138, see Table II), participants showed a stronger preference for textures that realistically blend AI-generated details with manual corrections.

[12]https://procreate.com/ (last accessed March 2024

VI. CONCLUSIONS

We presented an advanced workflow for the generation of Synthetic Humans, aimed at minimizing human intervention and ensuring an efficient and flexible process exploiting Rai's archive. Initial study focused on the complete automation of the Face Reconstruction stage. Subsequently, emphasis was placed on improving textures through the adoption of finetuned Generative AI to obtain additional realistic images of the reference subject. Images generated by AI demonstrated significant similarity to their original counterparts, both objectively and through subjective user evaluations. The fidelity of 3D meshes obtained during the automated Face Reconstruction phase remained at a high level. However, the comparison between the obtained MetaHuman and the original reference character presented some critical issues, e.g. loss of facial geometry information during conversion of 3D mesh to MetaHuman.

REFERENCES

[1] E. Spadoni, M. Carulli, M. Mengoni, M. Luciani, and Mo. Bordegoni. 2023. Empowering Virtual Humans' Emotional Expression in the Metaverse. In Universal Access in Human-Computer Interaction: 17th International Conference, UAHCI 2023. Proceedings, Part II. Springer-Verlag, Berlin, Heidelberg, 133–143. https://doi.org/10.1007/978-3-031-35897-5_10

[2] L. Chen, S. Peng, and X. Zhou. Towards efficient and photorealistic 3d human reconstruction: A brief survey. Visual Informatics, 5(4):11–19, 2021. https://doi.org/10.1016/j.visinf.2021.10.003.

[3] Guo, K., Lincoln, P., Davidson, P., Busch, J., Yu, X., Whalen, M., Harvey, G., Orts-Escolano, S., Pandey, R., Dourgarian, J., et al., 2019. The relightables: Volumetric performance capture of humans with realistic relighting. ACM TOG

[4] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J., 2015. Smpl: A skinned multi-person linear model. ACM Trans. Graph. 34 (6), 1–16.

[5] Zheng, Z., Yu, T., Wei, Y., Dai, Q., Liu, Y., 2019. Deephuman: 3d human reconstruction from a single image. IEEE/CVF International Conference on Computer Vision, pp. 7739–7749.

[6] Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., Li, H., 2019. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. IEEE/CVF International Conference on Computer Vision, pp. 2304–2314.

[7] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., Nerf: Representing scenes as neural radiance fields for view synthesis. ECCV 2020.

[8] L. Zhang and M. Agrawala. Adding conditional control to text-to-image diffusion models. https://arxiv.org/pdf/2302.05543.pdf (last accessed March 2024)

[9] E. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. Lora: Low-rank adaptation of large language models. https://arxiv.org/pdf/2106.09685.pdf (last accessed March 2024).

[10] Jing Yang Niannan Xue Irene Kotsia Jiankang Deng, Jia Guo and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. https://arxiv.org/pdf/1801.07698v4.pdf (last accessed March 2024).

[11] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, 'Retinaface: Single-shot multi-level face localisation in the wild', in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5203–5212.