Real-World Applications for PIFuHD: In 3D Digitization of Humans

Arya Agrawal [1], Gunavathi C [1], Saket Balabhadruni [1], Harsh Singh [2]
[1] *School of Computer Science and Engineering,* [2] *School of Electronics Engineering*
Vellore Institute of Technology, Vellore - 632014
Vellore, Tamil Nadu, India
arya.agrawal2021@vitstudent.ac.in, gunavathi.cm@vit.ac.in, saketb2003@gmail.com, harshu21122003@gmail.com

***Abstract*—This paper presents applications of Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization (PIFuHD), a novel approach for high-resolution 3D human digitization from single images. The objective is to reconstruct detailed object images. The approach employed in PIFuHD utilizes a multi-level pixel-aligned implicit function, enabling comprehensive analysis of both global context and intricate local details. This methodology aims to achieve detailed and high-resolution 3D reconstructions of clothed individuals from single images, eliminating the need for supplementary post-processing or extra contextual information. The key findings are that the multi-level approach is needed for accurate and precise reconstructions. This paper aims to explore future work based on the recommendation of the authors for PIFuHD, to explore PIFu's versatility extends to computer animation, virtual try-on, medical imaging, and numerous applications in forensic science, Augmented Reality (AR), Virtual Reality (VR), Human-Computer Interaction (HCI), offering precision and customization potential across various industries.**

***Index Terms*—High-fidelity human digitization, Multi-level PIFu method, Applications of Human Digitization, Neural Networks**

I. Introduction

The widespread adoption of HD human digitization has the potential to revolutionize a wide range of applications, spanning from imaging in healthcare to virtual reality.

3D human shape estimation is the process of reconstructing the 3D shape of a human body from one or more 2D images or videos. This is a challenging task due to the complexity of the human body and the inherent ambiguity in 2D images. [1] The goal of 3D human shape determination is to produce a 3D model of the human body that accurately captures its shape and pose, which has presented itself to be a challenge in the past. To address this challenge, the research community has increasingly turned to the integration of powerful deep-learning models. These models show remarkable promise in generating reconstructions even from a single image. [2]

3D object reconstruction has widespread application, depth and shape estimation during autonomous tasks like driving, or even industrial robots use cases, to 3D reconstruction for applications like Augmented Reality(AR), Virtual Reality, Medical Imaging, and many more. On one hand, autonomous driving and robots require point cloud 3D reconstruction of humans and nonhumans alike, whereas AR requires 2D to 3D regeneration. [3] This paper focuses more on 2D to 3D regeneration use cases, and to test whether existing algorithms present to be a viable option for such use cases. The authors of this paper are researching to check whether the practical application of theoretical models is possible or not. For this, the benchmark set for testing 3D reconstruction is focussed on 3D Human Reconstruction from 2D datasets. [4] [5]

The research problem addressed in this paper revolves around the real-world applications of PIFuHD in digitizing humans. Specifically, it aims to investigate the extent to which PIFuHD can be utilized in various fields based on the recommendations provided by its authors. [6] [7] The objective of this paper is to extensively examine and assess the practical applications of PIFuHD, an efficient method for high-resolution 3D human digitization from single images. To achieve this overarching purpose, the paper outlines the following specific objectives:

- To assess the versatility of PIFuHD in real-world scenarios by delving into its applications in computer animation, virtual try-on experiences, medical imaging, and more.
- To examine the potential impact of PIFuHD in fields such as forensic science, where precise 3D reconstructions can enhance investigative processes.
- To explore how PIFuHD contributes to the realms of augmented reality (AR), virtual reality (VR), and human-computer interaction (HCI), investigating its role in enhancing user experiences and interaction within these domains.
- To highlight the significance of PIFuHD in terms of offering unparalleled precision and customization opportunities across various industries, potentially revolutionizing established practices. By addressing these objectives, this paper aims to explore on the diverse real-world implementations of PIFuHD and analyze whether PIFuHD can be used as a benchmark for the above-mentioned applications or not. [7]

II. RELATED WORK

Before deep diving into PIFuHD applications,, it is important to review all past implementations and understand how the PIFuHD algorithm is the best suited to test. The papers used as a reference represent significant breakthroughs in 3D human shape estimation and avatar generation through deep learning. Some of the existing approaches include PIFu, RGB Camera reconstruction, and reconstruction using monocular video. A

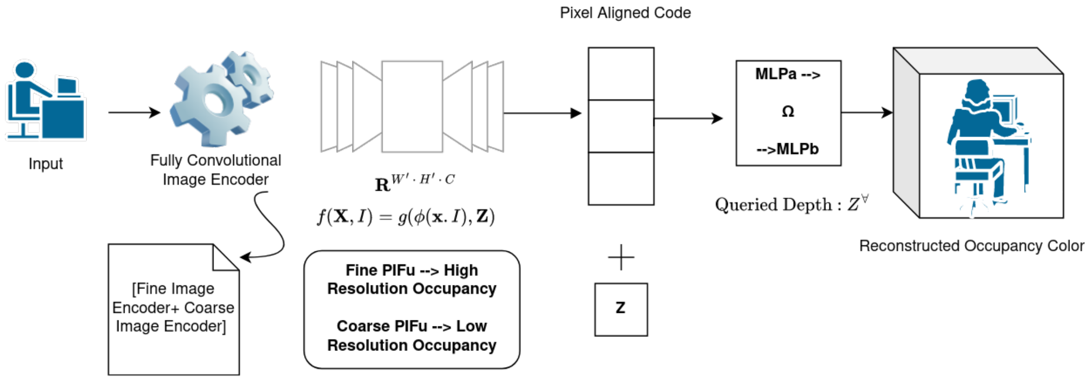
Pixel Aligned Code
Input
Fully Convolutional Image Encoder
$\mathbf{R}^{W' \cdot H' \cdot C}$
$f(\mathbf{X}, I) = g(\phi(\mathbf{x}. I), \mathbf{Z})$
MLPa -->
Ω
-->MLPb
Queried Depth : $Z^{\forall}$
Reconstructed Occupancy Color
[Fine Image Encoder+ Coarse Image Encoder]
Fine PIFu --> High Resolution Occupancy
Coarse PIFu --> Low Resolution Occupancy
+
Z

Fig. 1. PIFuHD diagrammatic representation

few techniques for 3D reconstruction include the usage of sensor fusion, Open3D library, etcetera. [8]

In recent research, several innovative methods have emerged for digitizing and reconstructing high-resolution clothed human models. PIFu (Pixel-Aligned Implicit Function) presents an implicit representation approach to align pixels with 3D objects, enabling detailed reconstructions of clothing and hairstyles from single images [9]. PIFuHD introduces multi-level architecture and high-resolution imagery to enhance the reconstruction of clothed humans, eliminating the need for post-processing [7]. Geo-PIFu combines 3D and 2D features for improved mesh reconstruction and topology robustness [8]. Octopus revolutionizes personalized body shape estimation from monocular video, reducing image requirements and computation time [10]. IntegratedPIFu advances single-view human reconstruction by leveraging depth and human parsing information [11]. The 'Detailed Human Avatars' method refines parameterized body models for high-detail avatars, featuring natural faces, hairstyles, and clothes [12]. These innovations offer valuable solutions for various applications in fields like virtual reality and medical imaging.

Among the implementations discussed in Table 1, 'PIFuHD' takes the spotlight. It distinguishes itself by utilizing high-resolution imagery and a multi-level architecture to achieve detailed reconstructions of clothed people from individual images, eliminating the need for post-processing or extra data. This paper will serve as the primary reference for exploring and testing the viability of the envisioned applications in our research. [7]

Method	Advantages	Limitations
PIFu [9]	- Innovative implicit representation for high-resolution clothed human digitization. Recovers intricate variations from a single input image.	- Limited to digitizing clothed humans. Only works with Front Facing images.
PIFuHD [7]	- High-resolution imagery and multi-level architecture for detailed reconstruction. Eliminates the need for post-processing.	- Computationally intensive, limiting real-time applications.
Geo-PIFu [8]	- Combines geometry-aligned 3D and pixel-aligned 2D features for enhanced occupancy estimation and mesh reconstruction. Improves global topology robustness and local surface details.	- Increases computational complexity due to the combination of 3D and 2D features.
IntegratedPIFu [11]	- Leverages depth and human parsing info for improved single-view human reconstruction. Predicts critical human features with reduced noise.	- Increased computational load and resource requirements.
Detailed Human Avatars [12]	- Creates high-detail human avatars from monocular video.	- Limited to specific body poses and movements. Ineffective for complex clothing topologies. Works with people in minimal clothing.

TABLE I
Comparison of Methods

III. Methodology

The PIFuHD algorithm requires a single high-resolution image of a clothed human as input. The image should be taken from a fixed viewpoint and should contain a full-body view of the person. The algorithm also requires a pre-trained deep-learning model for feature extraction. [7]

The PIFuHD algorithm has two levels: a coarse level and a fine level. At the coarse level, the algorithm estimates the silhouette of a human body by the help of a low-resolution image. At the fine level, the algorithm refines the shape estimate using a high-resolution image. The two levels are connected by a feature pyramid network that allows the algorithm to leverage information from both levels. [7]

The algorithm requires a single high-resolution image of a clothed human as input. The image should be taken from a fixed viewpoint and should contain a full-body view of the person. The algorithm also requires a pre-trained deep-learning model for feature extraction. Before feeding the input image to the algorithm, some pre-processing steps are required. The input image is first resized to a fixed resolution of 1024x1024 pixels. The image is then normalized to have zero mean and unit variance. The normalized image is then passed through the pre-trained feature extraction network to obtain a feature map. The PIFuHD algorithm uses a multi-level architecture that consists of a coarse level and a fine level.

At the coarse level, the algorithm approximates the overall shape of the human body through the analysis of a low-resolution image. This is achieved through the following equation:

$$f^L(X) = g^L(\phi^L(x_L, I_L, F_L, B_L), Z) \tag{1}$$

At the fine level, the algorithm refines the shape estimate using a high-resolution image. The equation that follows is:

$$f^H(X) = g^H(\phi^H(x_H, I_H, F_H, B_H), \Omega(X)) \tag{2}$$

The two levels are connected by a feature pyramid network that allows the algorithm to leverage information from both levels. The PIFuHD algorithm uses an implicit function to represent the 3D shape of the human body. The implicit function is defined as a continuous function that takes a 3D point as input and provide a signed distance to the outer boundary of the human body.

The neural network refines the implicit function by diminishing the difference between the estimated surface and the genuine ground truth surface. The PIFuHD algorithm also uses a pixel-aligned sampling strategy to ensure that the implicit function is aligned with the input image. This approach includes the projection of the 3D surface onto the 2D image plane, followed by sampling the image at the locations corresponding to the projections.

The experimentation in this paper revolves around the practical application of the proposed method across diverse datasets with varying applications, as detailed in the Experimentation section.

This approach aims to showcase the method's effectiveness and versatility by demonstrating its performance and adaptability across a range of data types and use cases. The paper substantiates the method's ability to provide meaningful insights and solutions applicable to a broad spectrum of real-world scenarios, and real-world applications.

IV. EXPERIMENTATION

The broad applications for testing are identified as Medical Imaging, Virtual Try-On, Computer Graphics and Animation, AR/VR, Forensic Science, and Surveillance. The applications are beyond those mentioned previously, but the results and inference of this paper rely on the mentioned ones.

The dataset collected for the above applications was taken from the Max Planck Institute for Informatics (MPII) Human Pose Dataset and Crowd Pose, and real-time testing was also conducted to note the difference in outputs and accuracy.

The next steps involved included defining and refining performance metrics specific to each application. For example, in medical imaging, metrics might focus on accuracy and sensitivity, while in fashion prototyping, style transfer quality and realism could be critical. Along with performance metrics, conducting user studies to gather feedback and insights from end-users in each application domain, for example, in virtual try-on, getting feedback on the level of satisfaction achieved when the 3D model for the test case was generated, along with how helpful the 3D model was, for virtual try-on use case.

V. RESULT

A. *Medical Imaging*

The following test image was used to check whether 3D reconstruction could be done using PIFuHD, but the experimentation failed and the results were broken. This meant that PIFuHD's future scope could not include medical imaging as an application.

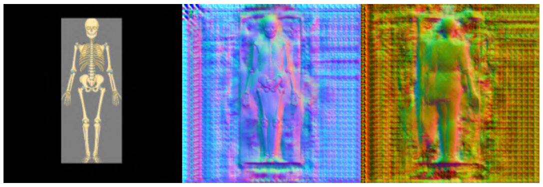

Fig. 2. PIFuHD, when applied on medical imaging dataset.

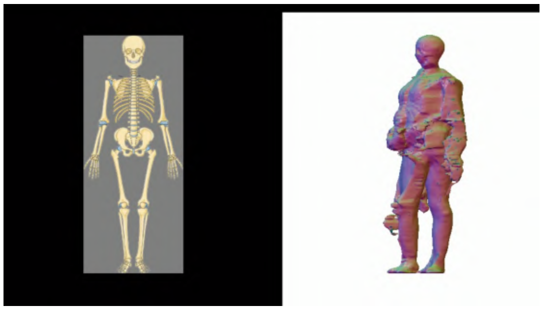

Fig. 3. Result of an input medical imaging image generated on PIFuHD which depicts frontal view with accuracy, but the side view isn't generated according to predicted 3D reconstruction

B. *Virtual Try-On*

The above result was obtained for testing the "business attire" try-on. Still, the results were incomplete for the use case since the results generated did not have the same image colors as the input. Hence, the users faced a limitation in seeing the exact outfit details, defeating the purpose of a virtual try on. This can be treated as a part of future developments for PIFuHD.

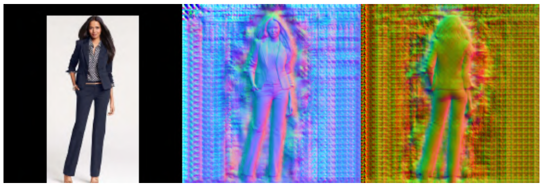

Fig. 4. PIFuHD application on 2D avatars on people dataset which gave incomplete results but can be considered as a future development idea to generate coloured 3D avatars to be used as a virtual try on application.

C. Computer Graphics and Animation

The application of PIFuHD can be done in the animation of images, the only limitation that arises is the same as that faced in virtual try-on.

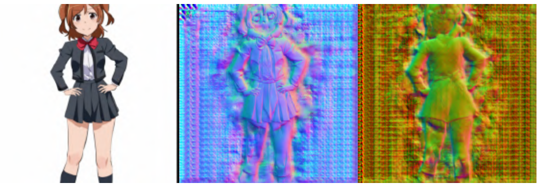

Fig. 5. 3D reconstruction can be done on animated images dataset as well, results gave an inference of trying out coloured 3D reconstruction for accurate RGB reconstructed 3D images as part of future development.

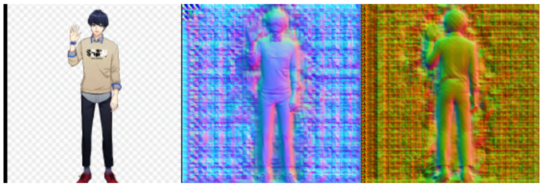

Fig. 6. Test 2 for Animated dataset testing.

D. Augmented Reality and Virtual Reality

The limitation that arose in AR/VR applications was that In AR applications, especially those involving interaction with the physical environment, low latency is critical to maintaining the illusion of objects being present in the real world. The time taken for 3D reconstruction can introduce latency, which can harm the user experience.

Optical limitations were also identified, along with the calibration of PIFuHD with the AR/VR hardware and tracking systems, with the side note that PIFuHD doesn't generate RGB images in 3D, and only gives the 3D reconstruction of the image in a different format.

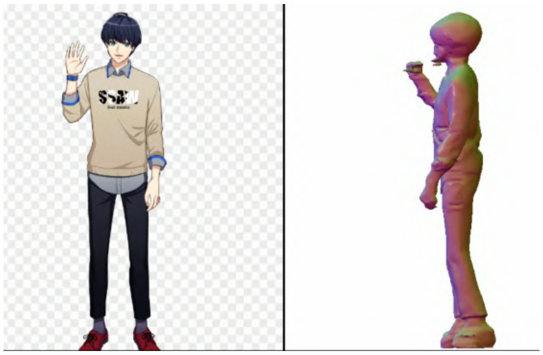

Fig. 7. Result of PIFuHD application on an image to be tested in VR systems, however, no conclusion can be drawn until these results are tested on hardware systems in order to check the compatibility and accuracy of the 3D images with respect to the VR and the AR hardware and software.

E. Forensic Science/Surveillance

Forensic Science and surveillance through PIFuHD can only be successful when better pose estimation along with RGB mask creation can be integrated into this algorithm since this line of work carries a great need for accuracy and clarity in the generated images, which couldn't be identified in the images tested.

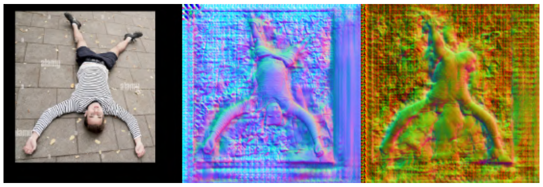

Fig. 8. Testing on forensic and surveillance datasets also gave an inconclusive result since the precise estimations for this use case isn't viable through just these results.

VI. Conclusion

The utilization of a multi-level approach in PIFuHD plays a crucial role in ensuring accurate and precise reconstructions. By addressing ambiguity in images, this approach significantly enhances the consistency of 3D reconstruction details, particularly in occluded regions. These findings have important implications for a wide range of applications, spanning from medical imaging to virtual reality.

In the field of medical imaging, the ability to create high-resolution 3D models of patients using PIFuHD could lead to more accurate diagnoses and better treatment outcomes. In the realm of virtual reality, the ability to create realistic avatars using PIFuHD could enhance the user's sense of presence and improve the overall quality of the virtual environment. In the fashion industry, the ability to create virtual try-on systems using PIFuHD could lead to a reduction in returns and an increase in customer satisfaction.

The results presented in this paper also have implications for the ongoing development of PIFuHD. The identified limitations collectively underline the ongoing research required

Application	Metrics	Description
Medical Imaging	Accuracy, Sensitivity	Accuracy measures the match between 3D reconstructions and actual patient anatomy. Sensitivity assesses PIFuHD's ability to detect subtle anatomical changes.
Virtual Try-On	Satisfaction Level, Model Helpful-ness	Satisfaction Level gauges user satisfaction with the virtual try-on experience. Model Helpful-ness evaluates how well the 3D model represents the clothing item during virtual try-on.
Computer Graphics	Accuracy of RGB-Reconstructed 3D Images	Measures how closely the 3D images generated by PIFuHD match the actual RGB images.
AR/VR Applications	Latency Issues, Optical Limitations	Latency Issues assess the delay between user actions and system response. Optical Limitations evaluate the constraints of the optical system used for scene capture.
Forensic Science	Precision, Clarity	Precision assesses the accuracy of 3D reconstructions in representing the subject being analyzed. Clarity evaluates how clear and interpretable the 3D reconstructions are.

TABLE II
PERFORMANCE METRICS FOR PIFUHD EVALUATION

to refine PIFuHD and seamlessly integrate 3D reconstructions into various applications. For instance, in medical imaging, achieving accurate 3D reconstructions proved challenging, indicating the need for further enhancements in this domain. Similarly, in virtual try-on, the inability to replicate exact image colors from the input restricted a seamless user experience. The application in computer graphics and animation faced a similar constraint, warranting future development for more accurate RGB-reconstructed 3D images. In AR/VR applications, PIFuHD's viability was hampered by latency issues and optical limitations, necessitating integration improvements for real-time use. Moreover, the potential in forensic science and surveillance relies on better pose estimation and RGB mask creation to meet the precision and clarity demands of these applications.

The limitations identified in the results section can be improved through further research and development. For instance, in medical imaging, achieving accurate 3D reconstructions proved challenging, indicating the need for further enhancements in this domain. One potential solution could be to incorporate additional imaging modalities, such as MRI or CT scans, to improve the accuracy of the 3D reconstructions.

Similarly, in virtual try-on, the inability to replicate exact image colors from the input restricted a seamless user experience. This limitation could be addressed by developing more advanced algorithms that can accurately capture and reproduce color information from the input images.

The application in computer graphics and animation faced a similar constraint, warranting future development for more accurate RGB-reconstructed 3D images. This could be achieved through the development of more advanced neural network architectures that can better capture the complex relationships between RGB images and 3D geometry.

In AR/VR applications, PIFuHD's viability was hampered by latency issues and optical limitations, necessitating integration improvements for real-time use. This could be addressed through the development of more efficient algorithms that can process images in real-time and the use of more advanced optical systems that can capture more detailed information about the scene.

Moreover, the potential in forensic science and surveillance relies on better pose estimation and RGB mask creation to meet the precision and clarity demands of these applications. This could be achieved through the development of more advanced pose estimation algorithms that can accurately capture the pose of the subject and the use of more advanced RGB mask creation techniques that can better separate the subject from the background.

The future scope of PIFuHD is vast and promising. The technology has the potential to revolutionize a wide range of applications, spanning from medical imaging to the realm of virtual reality. One potential future direction for research is to explore PIFu's versatility in computer animation, virtual try-on, and numerous applications in forensic science, Augmented Reality (AR), Virtual Reality (VR), and Human-Computer Interaction (HCI).

To sum up, the paper underscores the significance of the multi-level approach in achieving accurate and precise reconstructions. It emphasizes that mitigating ambiguity in the image domain significantly improves the consistency of 3D reconstruction details, particularly in occluded regions. These findings have important implications for a wide range of applications, spanning from medical imaging to the realm of virtual reality. Further advancements and focused research will undoubtedly play a pivotal role in unlocking the true potential of PIFuHD across a spectrum of domains, potentially offering a clearer direction for future research. Overall, this paper offers a succinct wrap-up of the research presented, highlighting the importance of high-fidelity human digitization and its transformative impact on various industries.

REFERENCES

[1] Y. Cao, G. Chen, K. Han, W. Yang, and K.-Y. K. Wong, "Jiff: Jointly-aligned implicit face function for high quality single view clothed human reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2729–2739.

[2] H. Zhao, J. Zhang, Y.-K. Lai, Z. Zheng, Y. Xie, Y. Liu, and K. Li, "High-fidelity human avatars from a single rgb camera," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 15 904–15 913.

[3] R. Liu, J. Shen, H. Wang, C. Chen, S.-c. Cheung, and V. Asari, "Attention mechanism exploits temporal contexts: Real-time 3d human pose reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5064–5073.

[4] Z. Zheng, T. Yu, Y. Wei, Q. Dai, and Y. Liu, “Deephuman: 3d human reconstruction from a single image,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7739–7749.

[5] M. Zanfir, A. Zanfir, E. G. Bazavan, W. T. Freeman, R. Sukthankar, and C. Sminchisescu, “Thundr: Transformer-based 3d human reconstruction with markers,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 971–12 980.

[6] C.-H. Hsieh, Y. Song, Z. Wang, and C. Li, “An adaptive bsco algorithm of solid color optimization for 3d reconstruction system with pifuhd,” in *International Conference on Wireless Algorithms, Systems, and Applications*. Springer, 2022, pp. 16–27.

[7] S. Saito, T. Simon, J. Saragih, and H. Joo, “Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 84–93.

[8] T. He, J. Collomosse, H. Jin, and S. Soatto, “Geo-pifu: Geometry and pixel aligned implicit functions for single-view human reconstruction,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 9276–9287, 2020.

[9] S. Saito, Z. Huang, R. Natsume, S. Morishima, A. Kanazawa, and H. Li, “Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 2304–2314.

[10] T. Alldieck, M. Magnor, B. L. Bhatnagar, C. Theobalt, and G. Pons-Moll, “Learning to reconstruct people in clothing from a single rgb camera,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1175–1186.

[11] K. Y. Chan, G. Lin, H. Zhao, and W. Lin, “Integratedpifu: Integrated pixel aligned implicit function for single-view human reconstruction,” in *European conference on computer vision*. Springer, 2022, pp. 328–344.

[12] T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll, “Detailed human avatars from monocular video,” in *2018 International Conference on 3D Vision (3DV)*. IEEE, 2018, pp. 98–109.