IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TEXTURIZED MULTI-LEVEL IMPLICIT MODELLING FOR HIGH RESOLUTION 3D HUMAN DIGITIZATION USING PIFUHD APPROACH

[1]Shreya Junagade, [2]Parth Gorde, [3]Mihir Harne, [4]Roma Thakur

[3]Professor,Dr. S. V. Gumaste , Prof. H. R. Khairnar

Department of Information Technology MET BKC Institute of Engineering Nashik, Maharashtra, India.

Abstract

Our 3D human shape estimation network stands out for integrating volumetric feature transformation, merging diverse image features into 3D space to precisely recover surface geometry. Complemented by a rich dataset of 7000 real-world human models, our method, empowered by unique architecture, excels in single-image 3D human model estimation. Addressing challenges in estimating human pose and body shape from 3D scans over time, we introduce PIFuHD Pixel- aligned Implicit Function. PIFuHD enables end-to-end deep learning for digitizing detailed clothed humans from a single image, surpassing prior work with high-resolution reconstructions on the Render people dataset. Moreover, our innovative approach recovers fine details, even on occluded parts, by transforming shape regression into an aligned image-to-image translation problem. Using a partial texture map as input, our method estimates detailed normal and vector displacement maps, enhancing clothing representation on a low-resolution smooth body model. In the landscape of 3D human shape estimation, our multi-level architecture, balancing broad context and high resolution, significantly outperforms existing techniques, leveraging 1kresolution input images for enhanced single-image reconstructions.

Keywords: PIFuHD, texture map, normal map, multilevel architecture, high-resolution reconstruction, 3D human shape estimation, single-image 3D human digitization.

Introduction

In recent years, the demand for high-resolution 3D human models has grown substantially, driven by applications in various domains, including gaming, animation, virtual reality, medical simulations, anthropometry, and more. How- ever, existing 3D human digitization methods often need to capture the intricate details of human subjects, such as facial expressions and clothing textures. This report introduces an innovative approach, Texturized Multi-Level Implicit Modelling, to address these limitations and achieve high-resolution 3D human digitization PIFuHD is an advanced method in computer vision and graphics that transforms 2D images into highly detailed 3D human models. It stands out for using deep neural networks and a combination of techniques to achieve precise reconstruction of high-resolution human models. At its core, PIFuHD blends fully-connected techniques and deformable convolutional networks to accurately capture pose variations, surface details, and textures from 2D images. This approach surpasses previous methods, providing a significant improvement in the authenticity of 3D human models. PIFuHD excels not

only in its complex architecture but also in its ability to understand spatial relationships effectively. It achieves this through a careful interplay of feature extraction and fusion, resulting in three-dimensional representations that closely resemble the intricacies of the human form. This research delves into PIFuHD's capabilities, examining its architecture, training methods, and applications. The goal is to establish PIFuHD as a key player in the evolution of 3D human modelling. By demystifying its intricacies, this work contributes to the field's knowledge, offering insights that can impact industries such as virtual reality and human-computer interaction, promising a transformative future for 2D-to-3D human model synthesis. It describes a novel approach, Pixel-aligned Implicit Function (PIFu), for 3D deep learning in the context of textured surface inference for clothed 3D humans. In a world dominated by immersive technologies and autonomous systems, the ability to digitize and understand 3D objects is crucial. PIFuHD addresses the challenging task of reconstructing detailed 3D human models from single or multiple input images, focusing on preserving fine-scale details of clothed subjects. The key innovation of PIFuHD lies in aligning individual local features at the pixel level to the global context of the entire object in a fully convolutional manner. Unlike voxel-based representations or global context methods, PIFuHD combines local features with a 3D-aware implicit surface representation, allowing it to reason about 3D shapes accurately, even from a single view. The method employs an encoder to learn per-pixel feature vectors that consider the global context, enabling the preservation of local details while inferring plausible ones in unseen regions. PIFuHD end-to-end and unified digitization approach can predict high- resolution 3D shapes of individuals wearing complex clothing and hairstyles. It handles a variety of clothing types, including skirts and scarfs, capturing high-frequency details like wrinkles at the pixel level. Additionally, PIFuHD can naturally extend to infer per-vertex colours, generating a complete texture of the surface. The method is demonstrated to be effective and accurate on challenging real-world images of clothed subjects, showcasing high-resolution examples of monocular and textured 3D reconstructions from video sequences. Comprehensive evaluations against ground truth 3D scan datasets highlight PIFuHD state-of-the-art performance in digitizing clothed humans.

LITERATURE SURVEY

Several papers focused on reconstructing 3D human models from images or video using methods like skinned multi-person linear models, convolutional neural networks, and volumetric discretization. However, they had limitations in capturing details like facial expressions, hands, and clothing. 2020, Saito etal[1]:They introduced "PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization." Their approach utilized a multi-level pixel-aligned implicit function to achieve high-fidelity 3D human models. However, a notable limitation of this method is its reconstruction of 3D humans without incorporating textures or colours, potentially hindering the realism of the generated models.

2019, Zheng, Yu[2]: Zheng, Yu and their collaborators from Beihang University and Orbbec Company presented "Deep-Human: 3D Human Reconstruction from a Single Image." They implemented a 3D representation using the Skinned Multi-Person Linear (SMPL) model through volumetric discretization. While successful in generating 3D representations, this method faced challenges in re- constructing hands and facial expressions, limiting the comprehensiveness of the reconstructed human models.

2019, Alldieck, Magnor[3]:In the same year, Alldieck, Magnor, and Bhatnagar proposed "Learning to Reconstruct People in Clothing from a Single RGB Camera." They trained a Convolutional Neural Network (CNN) to infer a 3D mesh model while simultaneously reconstructing the subject's 3D model. How- ever, a key limitation was the reliance on synthetic data, raising concerns about the generalization of the model to real-world scenarios. 2018, Joo, Simon and Sheikh[4]:They introduced "Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies." This method utilized the Frank model, integrating body part models to reconstruct a complete 3D human model. Despite its capability to capture overall body movements, the approach had limitations, such as providing limited surface detail and complexity in model integration.

2020, Aymen Mir1 ,Thiemo Alldieck[5]:They provide effective method to automatically transfer textures of clothing images (front and back) to 3D garments worn on top SMPL [42], in real time. Their model opens the door for applications such as virtual try-on, and allows for generation of 3D humans with varied textures which is necessary for learning.

Motivation:

The texturized models provide geometric accuracy; they encapsulate the essence of human subjects by faithfully reproducing the fine details of skin, clothing, and even environmental influences like lighting and shadows. The result is a level of visual realism that was once thought unattainable in the digital do- main. The human 3d model addresses the long-standing challenge of achieving lifelike representation, not just in static forms but also in dynamic, evolving contexts. They empower professionals in these industries to create, simulate, and interact with digital representations that closely mirror reality.

Aim and Objective

Aim

The primary goal of this project is to develop a Texturized Multi-Level Implicit Modelling technique to reconstruct a 3D human with the textures from a single image.

Objectives

1. The realism of 3D human models: The model exhibit realistic rendering of facial expressions, including emotions and subtle movements, contributes to the authenticity of the 3D human model. The model exhibits lifelike skin textures, realistic clothing, and accurate proportions that closely resemble a human.
2. Capture intricate details in both geometry and textures: This involves capturing subtle nuances in the 3D shape, including fine contours, curves, and complex structures It enhances the visual fidelity of the model, making it more convincing and immersive in applications like computer graphics, virtual reality, and simulations where realistic depiction of objects is essential for a high-quality user experience.
3. To enhance the PIFuHD approach by incorporating texture mapping capabilities into the 3D human digitization process: It focuses on reproducing intricate surface textures such as skin tones, clothing patterns, and other visual details. By integrating texture mapping, the PIFuHD approach aims to achieve a more realistic and visually appealing representation of 3D human models, contributing to a higher level of fidelity in the digitization process.
4. Precise capture of dynamic facial expressions: This involves using advanced techniques and technologies to faithfully capture the subtle and rapid changes in facial features, such as smiles, frowns, and other expressions. This precision ensures that the resulting 3D models accurately reflect the dynamic nature of facial movements, contributing to more lifelike and realistic representations in digital environment.

SYSTEMS PROPOSED ARCHITECTURE

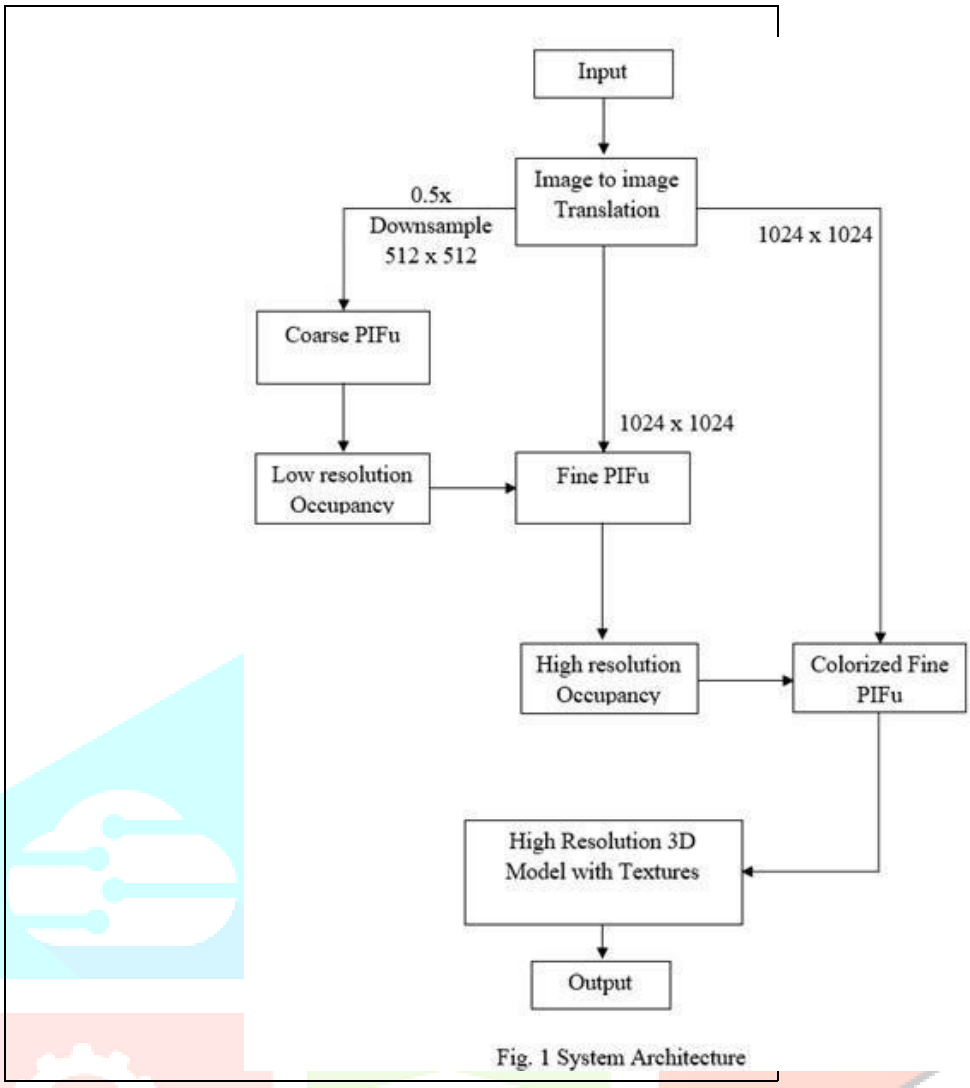

Figure 1: the system model of the proposed design

The method employs the Pixel-aligned Implicit Function (PIFuHD framework to enhance the resolution of 3D human digitization. Initially, PIFuHD processes 512 × 512 resolution images to derive low-resolution feature embeddings 128 × 128. For higher resolution, an additional pixel-aligned prediction module is introduced. The fine module processes 1024 × 1024 resolution images, encoding high-resolution features 512 × 512. The second module utilizes high-resolution feature embeddings and 3D embeddings from the first module to predict an occupancy probability field. To improve reconstruction quality, the method predicts normal maps for the front and back sides in image space, incorporating them as extra input. PIFuHD foundation is briefly described, emphasizing its goal of 3D human digitization by estimating occupancy in a dense 3D volume. PIFuHD models a function f(X), predicting binary occupancy values for any 3D position in continuous camera space X=(Xx,Xy,Xz)ˆaR. outputs 1 if X is inside the mesh surface and 0 otherwise, based on a single RGB image.

Coarse level: The coarse level in the PIFuHD architecture takes low-resolution images as input to cover large spatial context and focuses on holistic reasoning. It provides context to the fine level for estimating highly detailed geometry. The coarse level, integrates global geometric information by processing a down sampled 512×512 image, generating backbone image features of 128×128.resolution. The fine level adds more subtle details by using the original 1024 × 1024 resolution image, producing backbone image features of 512×512 resolution. Notably, the fine level takes 3D embedding features from the coarse level instead of the absolute depth value. The coarse level module is defined similarly to PIFuHD but incorporates predicted frontside and backside normal maps.

Fine level :Fine level in the PIFuHD architecture focuses on adding more subtle details to the 3D human digitization process. It takes the original high-resolution (1024x1024) image as input and produces backbone image features of 512 × 512 resolution, which is four times higher resolution than previous implementations.

The fine level module uses high-resolution input and includes a 3D embedding extracted from the coarse level network. The receptive field of the fine level doesn't cover the entire image, but owing to its fully convolutional architecture, it can be trained with a random sliding window and infer at the original image resolution i.e., 1024 × 1024. These normal maps guide the 3D reconstruction to produce sharper geometry. Normal maps for the back and front are predicted using a pix2pixHD network.

Texturized Fined: It add colourful textures to blank 3D human models. It Uses the original 1024x1024 image for texture mapping. It Creates detailed features at 256 × 256 resolution. It Utilizes 3D information from the Fine level and applies textures to human model.

RESULTS

The qualitative results showcase the effectiveness of the proposed digitization approach using real-world input images from the Render people dataset. the PIFuHD method is demonstrated to handle a diverse range of clothing items such as jackets and dresses. The method excels in producing high-resolution local details and inferring plausible 3D surfaces in previously unseen regions. Notably, the method successfully generates complete textures from a single input image, enabling a comprehensive view of the 3D models from 360 degrees. The results highlight the versatility of the approach in capturing intricate clothing details and faithfully reconstructing 3D surfaces.

CONCLUSION

In conclusion, this project to generate 3D human models from high-resolution images is a testament to human ingenuity and innovation. It promises to revolutionize multiple domains, with entertainment being just one of many beneficiaries. As we navigate the challenges and opportunities that this endeavour presents, we remain steadfast in our commitment to advancing the state of the art in computer graphics and computer vision. Through collaborative teamwork, dedication, and a shared vision of the future, we are on the path to unlocking the immense potential of realistic 3D human models. Together, we are shaping a future where digital realism knows no bounds, offering new creative possibilities and transforming industries across the spectrum. In summary, the PIFuHD approach represents a significant leap in 3D modelling technology, enabling the transformation of 2D images into detailed and realistic 3D models. As research progresses and technology evolves, we can anticipate further improvements in accuracy, speed, and applicability, making it an increasingly valuable tool in the realm of computer graphics and computer vision.

REFERENCES

List all the material used from various sources for making this project proposals:

[1] Shunsuke Saito, University of Southern California, Facebook Reality Labs, Facebook AI Research, PIFuHD: Multi-Level Pixel-Aligned Implicit Function For High-Resolution 3D Human Digitization 2020.

[2] T. Alldieck, M. Magnor, B. L. Bhatnagar, C. Theobalt, and G. PonsMoll. Learning to reconstruct people in clothing from a single RGB camera. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1175ˆa1186, 2019.

[3] T. Alldieck, M. A. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll. Video based reconstruction of 3d people models. In IEEE Conference on Com- puter Vision and Pattern Recognition, pages 8387aˆ 8397, 2018.

[4] T. Alldieck, G. Pons-Moll, C. Theobalt, and M. Magnor.Tex2shape: Detailed full human body geometry from a single image. In The IEEE International Conference on Computer Vision (ICCV), October 2019 .

[5] R.Alp GAˆ z˙ uler, N. Neverova, and I. Kokkinos. Densepose: Dense human pose estimation in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 7297aˆ 7306, 2018 .