Human body 3D avatar generation from an image

Saisamarth Udikeri (01FE21BCS003), Guruprasad Pattar (01FE21BCS226), Deepak Baligar (01FE21BCS041)

Under the guidance of
Uma Mudenagudi and Ramesh Ashok Tabib

KLE Technological University, Vidyanagar, Hubballi-580031, Karnataka, India

June 15, 2024

Overview

- Introduction
- Motivation
- Literature Survey
- Problem Statement and Objectives
- Approach / Solution
- Dataset Analysis / Description
- Experimental Results
- References

Introduction

- The fusion of computer vision and deep learning has propelled the creation of lifelike 3D avatars from 2D human images, marking a significant leap in digital content generation, gaming and virtual reality (VR).
- Through advanced algorithms and techniques, human body 3D avatar generation from images is revolutionizing how we perceive and interact with digital content, offering endless possibilities for immersive experiences and personalized visualizations.

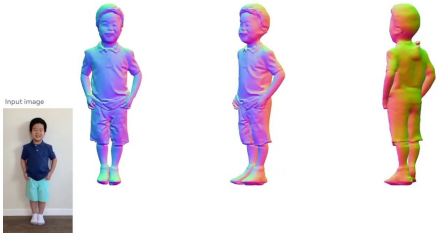

Figure: Image to 3D Avatar generation of a man

Motivation

- Converting 2D medical scans like MRI or CT scans into 3D models helps healthcare professionals analyze the anatomical structures of human body.
- Gaming and VR industries leverage 2D to 3D human avatar generation to create lifelike and customizable characters for players.
- 2D to 3D human avatar generation is used by animators, filmmakers, and digital artists to create animated characters and digital content.

Figure: The left side of the images shown is 2D, while the right one of the images is 3D.

Literature Survey

PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization (Saito Shunsuke et al., CVPR Dec 2019) [2]

The paper presents PIFu (Pixel-Aligned Implicit Function), a novel deep-learning approach for digitizing highly detailed 3D surfaces of clothed humans from one or more input images.

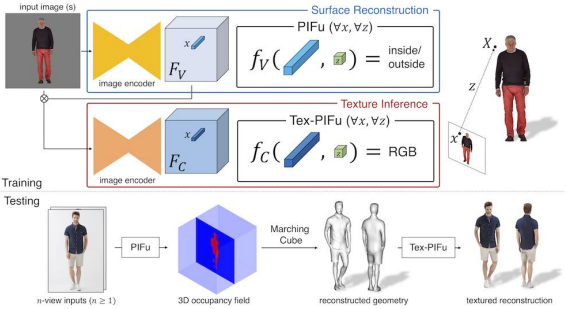

Figure: Image to 3D Avatar generation of a man

Literature Survey

PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization (Saito Shunsuke et al., CVPR April 2020) [3]

The author has proposed a novel multi-level architecture addressing this limitation by combining holistic reasoning from low-resolution images with detailed geometry estimation from high-resolution inputs, leading to superior performance in single image human shape reconstruction.

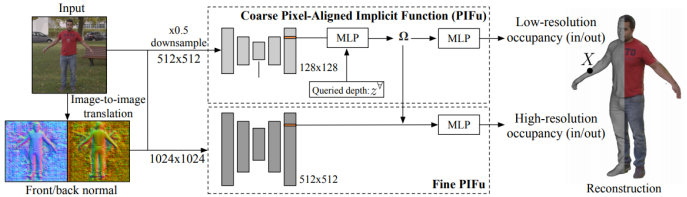

Figure: Two levels of pixel-aligned predictors produce high-resolution 3D reconstructions

Literature Survey

Geo-PIFu: Geometry and Pixel Aligned Implicit Functions for Single-view Human Reconstruction (Tong He et al., June 2020) [1]

- Geo-PIFu is a method that fuses geometry-aligned 3D features and pixel-aligned 2D features within a deep implicit function framework for high-quality reconstruction of clothed 3D human meshes from a single color image input.

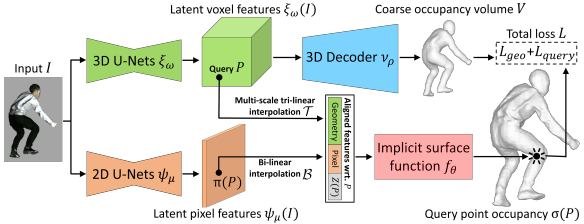

Figure: This method extracts latent 3D voxel and 2D pixel features from a single-view color image.

Problem Statement and Objectives

Problem Statement

To develop a framework for the generation of 3D human avatar using an image.

Objectives

- Using a learning-based architecture for generation of 3D human avatar from a given image.
- Refinement of generated 3D human avatar.

Pipeline for image to 3D human reconstruction

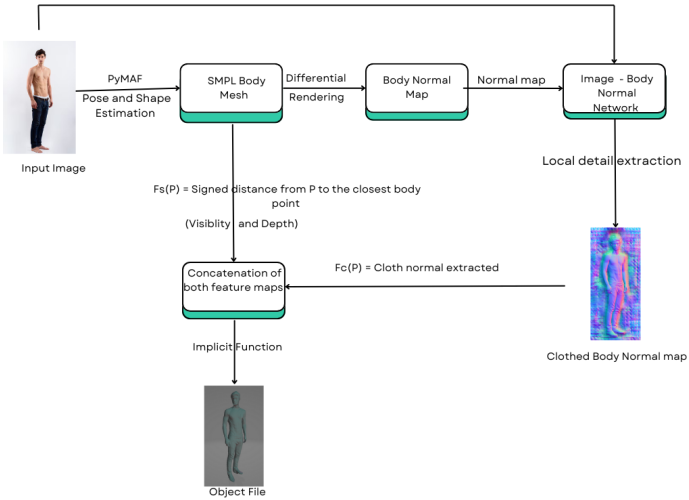

Figure: Proposed architecture using PyMAF and Differential Rendering

Dataset Analysis / Description

- Benchmark dataset used: Renderpeople dataset
- Detailed Explanation:
  - Total Samples: 1020 images
  - Training Samples: 816 images
  - Testing Samples: 204 images
- Sample images from dataset:

Figure: Sample images from Renderpeople dataset

Experimental Results

Figure: Input image along with front and back of generated object file

Experimental Results

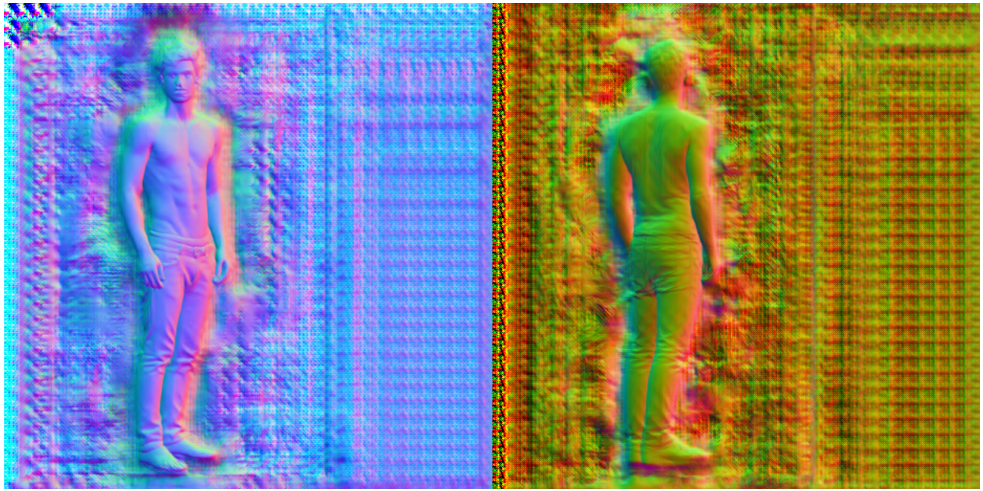

Figure: Generated Normals of front and back from a given input image

Experimental Results

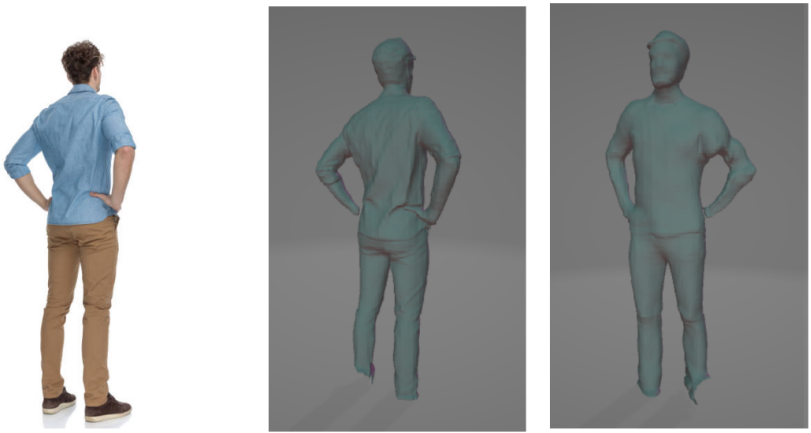

Figure: Input image along with front and back of generated object file

Experimental Results

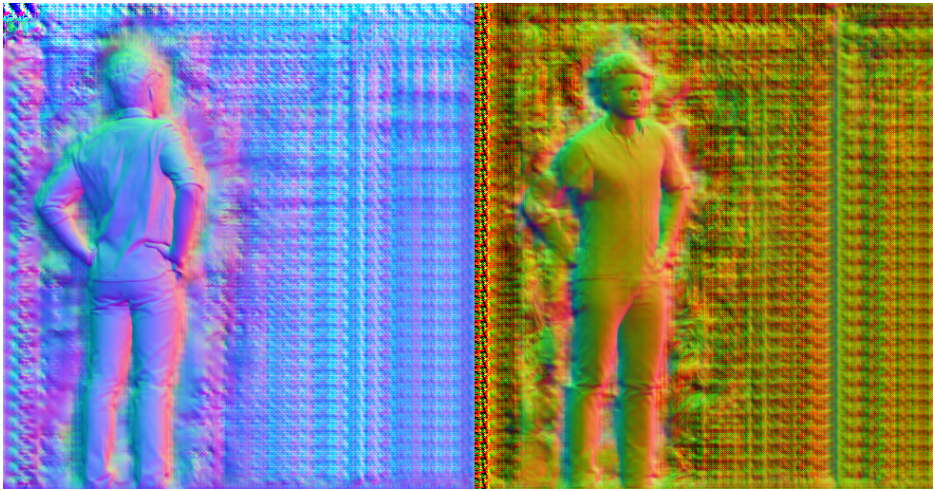

Figure: Generated Normals of front and back from a given input image

Experimental Results

Figure: Input image along with front and back of generated object file

Experimental Results

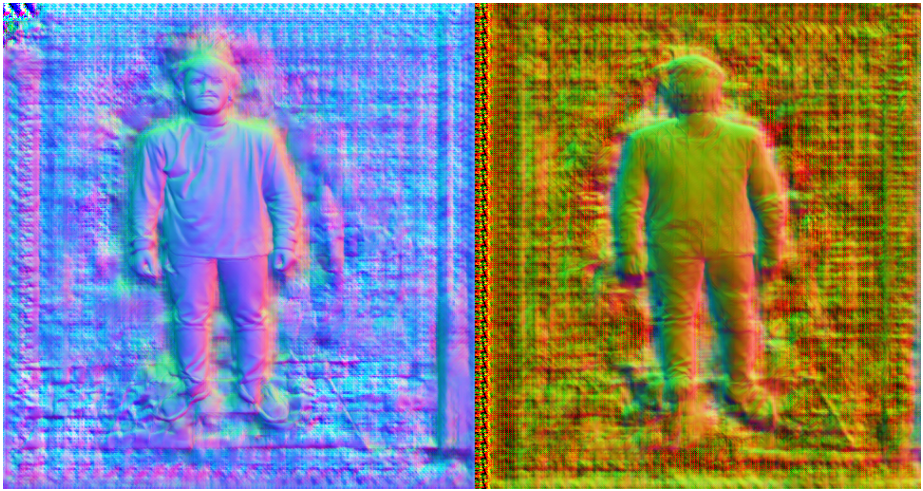

Figure: Generated Normals of front and back from a given input image

Tong He, John Collomosse, Hailin Jin, and Stefano Soatto.
Geo-pifu: Geometry and pixel aligned implicit functions for single-view human reconstruction.
In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9276–9287. Curran Associates, Inc., 2020.

Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li.
Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization.
In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul 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 (CVPR)*, June 2020.

Thank You