Department of Computer Engineering
Vishwakarma Institute Of Technology, Pune

8th International Conference on Data Management, Analytics & Innovation

Title 161 - Blending Motion Capture and 3D Human Reconstruction Techniques for Enhanced Character Animation

Authors :-

Anshuman Giramkar	anshuman.giramkar211@vit.edu
Shreyas Godse	shreyas.godse21@vit.edu
Snehal Gupta	snehal.gupta21@vit.edu
Pooja Indulkar	pooja.indulkar21@vit.edu

Guided By - Prof. Pushkar Joglekar

Open Innovation

Blending Motion Capture and 3D Human Reconstruction Techniques for Enhanced Character Animation

VIT, Pune

PROBLEM STATEMENT

We aim to revolutionize motion capture by combining 3D modeling and real-time motion tracking. With state-of-the-art algorithms and hardware, we provide an efficient solution for film and video game production. Our system enables precise tracking and generates high-quality 3D models with minimal lag time, reducing production costs and increasing development speed.

Proposed Solution

Generating 3D human models from 2D images is costly, time-consuming, and requires specialized software and hardware. That's where our solution comes into limelight.

PIFuHD(Pixel aligned Implicit Function for High Resolution 3D Human Digitization) generates 3D models in seconds, greatly improving workflow efficiency compared to traditional methods.
3DDFA is a powerful facial analysis technique that utilizes deep learning algorithms to accurately capture facial landmarks and reconstruct detailed 3D facial structures, enabling realistic and dynamic facial animations.

MeshMarvel Mocap captures precise motion data, useful for animation and virtual reality with high accuracy and fidelity.

Open-source PIFuHD can be modified and used by anyone, allowing for community contributions and improvements.

PIFuHD and MeshMarvel Mocap definitely prove to be affordable and work with standard hardware, unlike expensive traditional motion capture and 3D modeling methods.

Objectives:

- To demonstrate how PIFuHD and 3DDFa can be used to build a high-quality 3D model from a single 2D image and show examples of the resulting models.
- To show how the 3D model built using PIFuHD and 3DDFA can be combined with real-time motion capture data using MeshMarvel to create a realistic and dynamic animation.
- To discuss the potential applications of this technology in various fields, such as virtual reality, gaming, film, and medicine.
- To highlight the challenges and limitations of the project and suggest areas for future research and development .
- To engage the audience by using visual aids, demonstrations, and interactive elements, such as a live motion capture performance or a virtual reality simulation.

What is Machine Learning?

TECH STACK

PIFuHD and 3DDFA through Blender

PifuHD and 3DDFA generates 3D models from 2D images, Blender adds armature and prepares models, while Python automates the workflow.

MeshMarvel MoCap

MeshMarvel MotionCap is a plugin which we have built for Blender to create AR scenes. Import models, add camera, preview on mobile, and export.

Intel® Iris® and Git

Git was used for version control and team collaboration, and computation during the project was performed using Intel Iris Pro 1536MB.

UNDERLYING MATH: PIFUHD

- Consider Image as a variable input I. The 3D space is defined as **X**= (Xx, Yy, Zz) ∀ Xx, Xy, Xz ∈ R
- The Occupancy Values are calculated using the given function:

$$f(\mathbf{X},\mathbf{I}) = \begin{cases} 1 & \text{if } \mathbf{X} \text{ is inside mesh surface} \\ 0 & \text{otherwise,} \end{cases}$$

- The functional image embeddings are extracted from a loci **x =** (Xx, Xy) from the image I, thus defined as Φ(**x**, I).
- Now for **X** space in mesh, the function f is given:

$$f(\mathbf{X},\mathbf{I}) = g\left(\Phi\left(\mathbf{x},\mathbf{I}\right), Z\right),$$

Note: The functions Φ is implemented in the system using CNN architecture and the g function is implemented through Multilayer Perceptron (MLP).

These functions are multi-aligned and used twice: for a lower image resolution (for global structure) as a source and a higher image resolution (for fine detailing).

UNDERLYING MATH: 3DDFA

- The traditional 3D Mutable Structure (3DMM) S can be delineated as

$$\mathbf{S} = \overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp},$$

- T = f [R; t_{3d}], where T ∈ R 3×4 is constructed by a scale factor f, a rotation matrix R and a translation vector t_{3d} = ? [t_{2d}O]

$$V_{2d}(\mathbf{p}) = \mathbf{Pr} * \mathbf{T} * \begin{bmatrix} \overline{\mathbf{S}} + \mathbf{A}\boldsymbol{\alpha} \\ \mathbf{1} \end{bmatrix}$$

- Our regression objective is described as p = [T, α]
- For meta-joint optimization of p:
  - The VDC minimizes the vertex distances between the fitted 3D face and the ground truth

$$\mathcal{L}_{vdc} = \|V_{3d}(\mathbf{p}) - V_{3d}(\mathbf{p}^g)\|^2$$

  - The fWPDC assigns different weights w to each parameter.

$$\mathcal{L}_{fwpdc} = \|\mathbf{w}_T \cdot (\mathbf{T} - \mathbf{T}^g)\|^2 + \|\mathbf{w}_\alpha \cdot (\boldsymbol{\alpha} - \boldsymbol{\alpha}^g)\|^2$$

w = (w1, w2, . . . , wi , . . . , wn),
p de,i = (p g 1 , p g 2 , . . . , pi , . . . , p g n)

UNDERLYING MATH: MESHMARVEL MOCAP

- Quaternions are used for implying rotations.

$$Q = a + bi + cj + dk$$

here a, b, c, d denote real number coordinates in the 3D arena.

- For a quaternion q, the quaternion rotations are calculated using the formula

$$Q_{rotated} = Q \cdot Q_v \cdot Q_{conjugate}$$

- Qv is the vector along which rotations are taking place, expressed in quaternion format for rotatory estimation.

- The quaternions are converted to matrices. For a quaternion q = a + bi + cj + dk, the matrix can be represented as

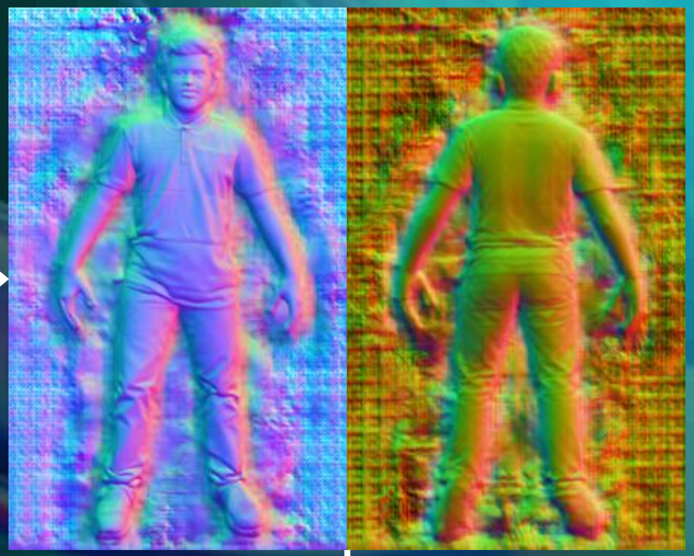
$$Mq = \begin{bmatrix} 1 - 2c^2 - 2d^2 & 2bc - 2ad & 2bd + 2ac \\ 2bc + 2ad & 1 - 2b^2 - 2d^2 & 2cd - 2ab \\ 2bd - 2ac & 2cd + 2ab & 1 - 2b^2 - 2c^2 \end{bmatrix}$$

- The importance of quaternion units in Mocap is their inherent ability to circumvent gimbal lock and facilitate smooth interpolation.

Flowchart for model creation:

Using Selection Tools

Using PiiFuHD in Python



1

SUBJECT PROFILE

Collect high-quality 2D images of the object or subject to be converted into a 3D model.

2

PROFILE PRE-PROCESSING

Use image pre-processing techniques, such as resizing or color correction, to prepare the input images for PifuHD.

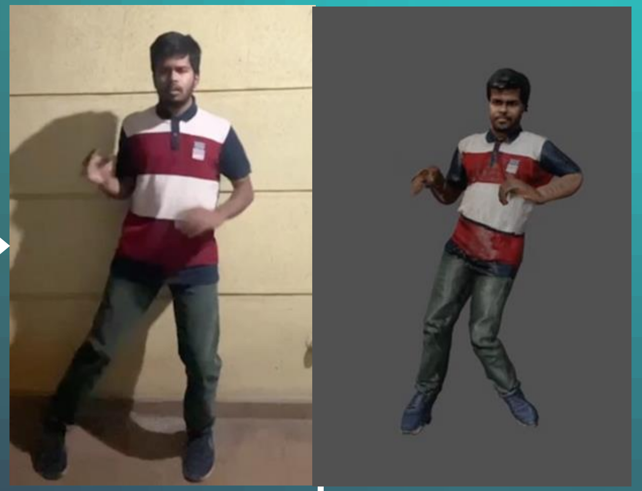
3

GENERATING 3D MESH

PifuHD generates 3D models from images using neural networks trained on 3D models and images.

SUBMISSION 161

Flowchart for model creation(Contd.):



4

5

6

3D RENDERING

Export the generated 3D mesh to a 3D modeling software, such as Blender or Maya, for further editing or refinement.

MESHMARVEL MOCAP

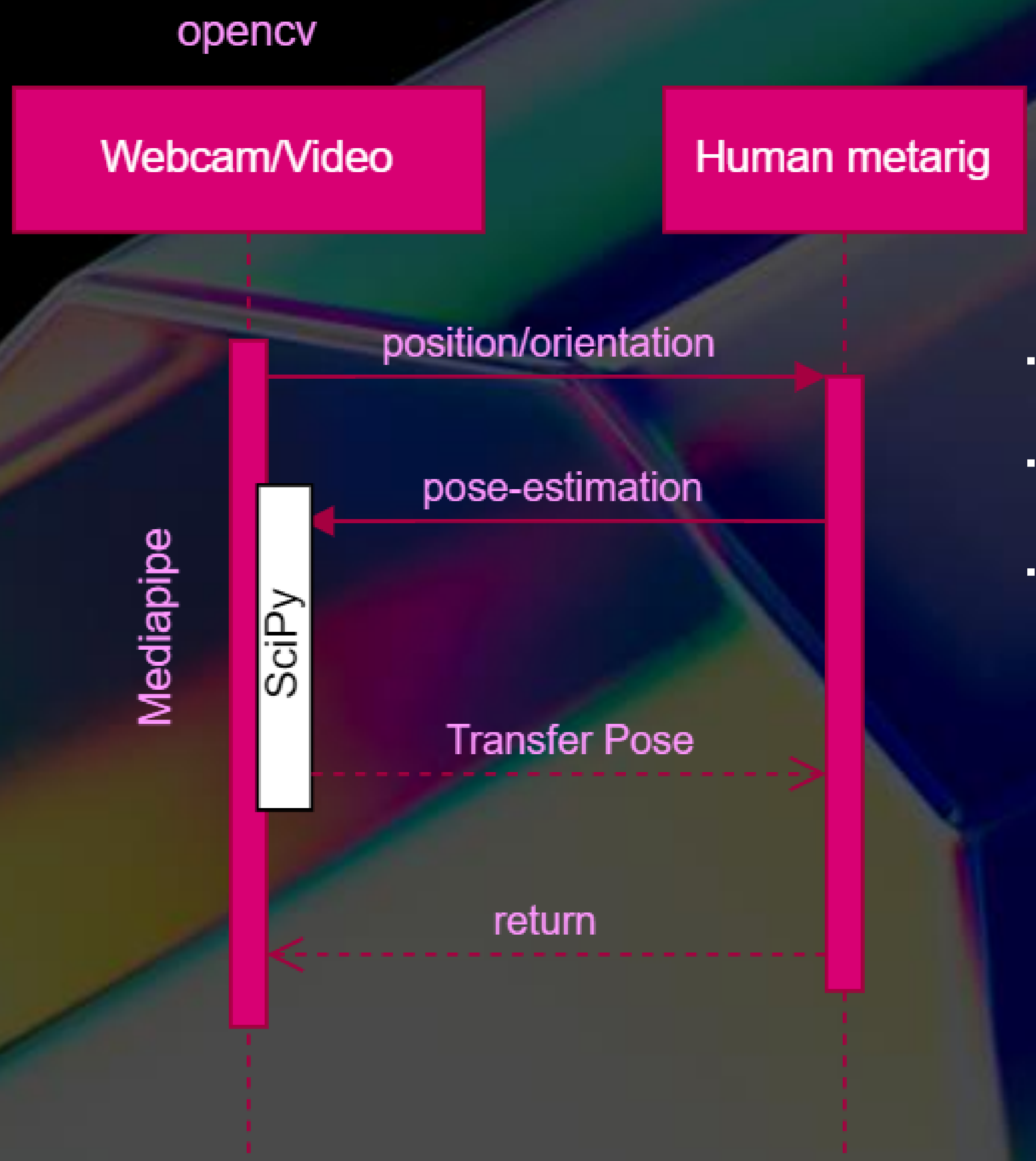
Record motion capture in MeshMarvel Mocap either through video or webcam, integrate it with the model armature rig.

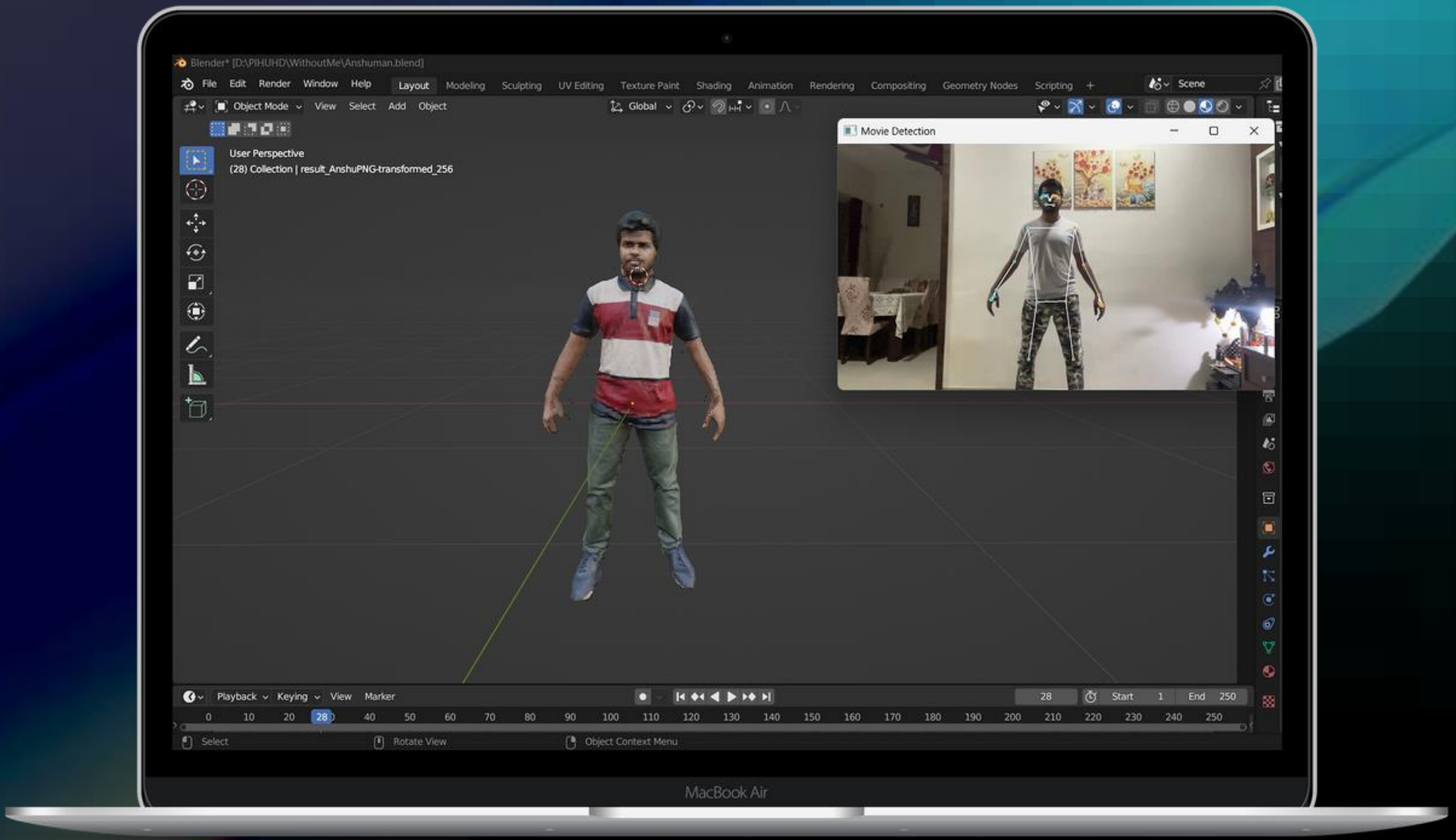
EXPORT

Export the animated 3D model for further use or integration into other projects, such as virtual reality experiences, video games, or animations.

SUBMISSION 161

Flowchart for Motion Capture:

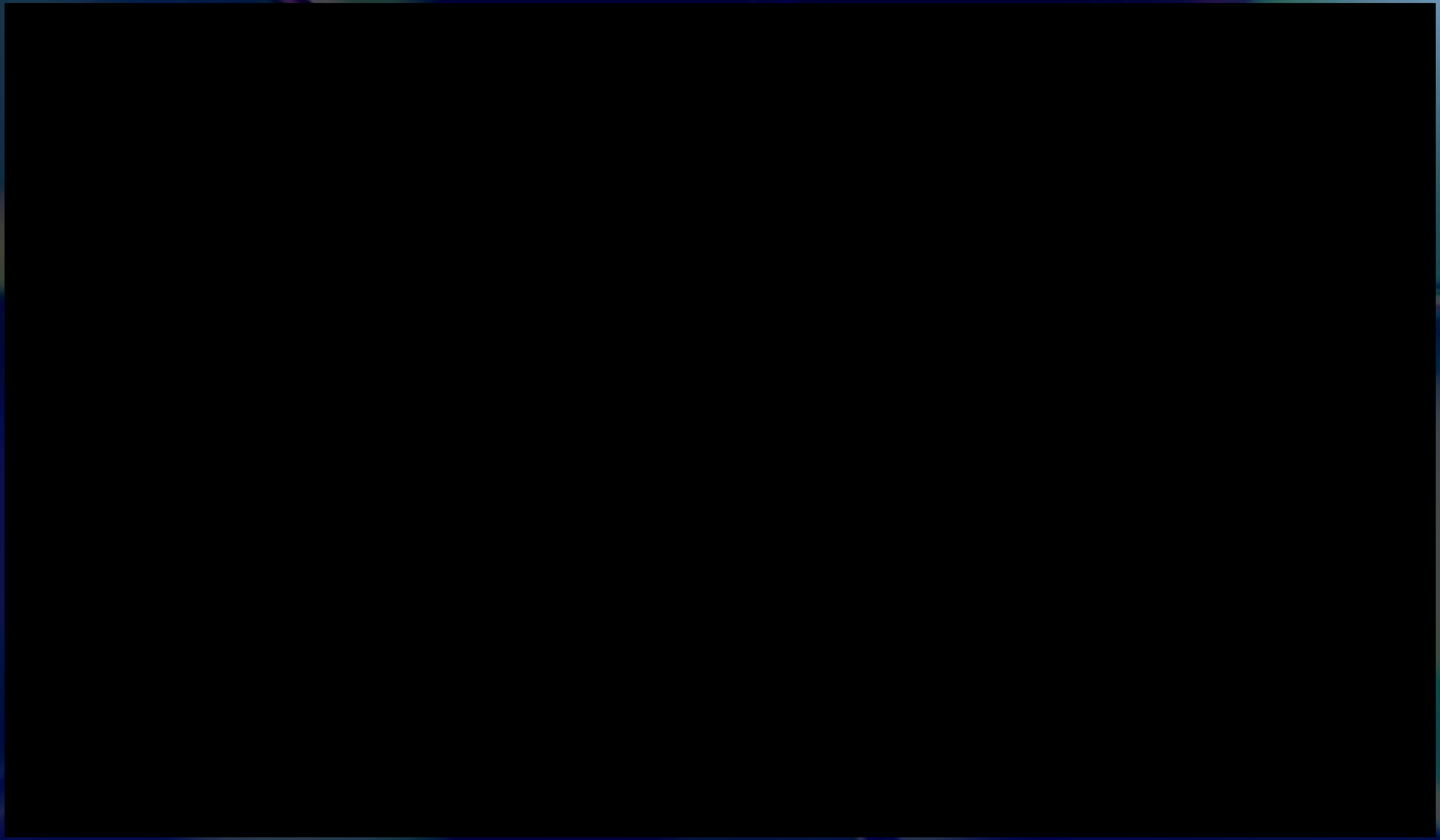




Achieved Result (3DDFA Made Model):

Achieved Result (PIFuHD Made Model):

Achieved Result (3DDFA Made Model):



Challenges Encountered:

Output quality can be affected by input image

Generative Adversarial Networks (GANs) are used to convert low-quality images to high-quality using ML algorithms by training neural networks on a dataset of high-quality images.

Removal of Image background and resolution amplification.

Convolutional neural networks (CNNs) trained on labeled images with foreground and background regions can accurately remove image background

Limited pose estimation accuracy

Blender Mocap will make use of OpenCv and Mediapipe can be used to add motion capture data to the 3D model

Challenges Encountered:

Hardware setup and real-time performance:

Research and carefully select appropriate hardware components based on your requirements and budget. Optimize your code by leveraging parallel processing, multithreading, or GPU acceleration.

Marker tracking alongwith data synchronization:

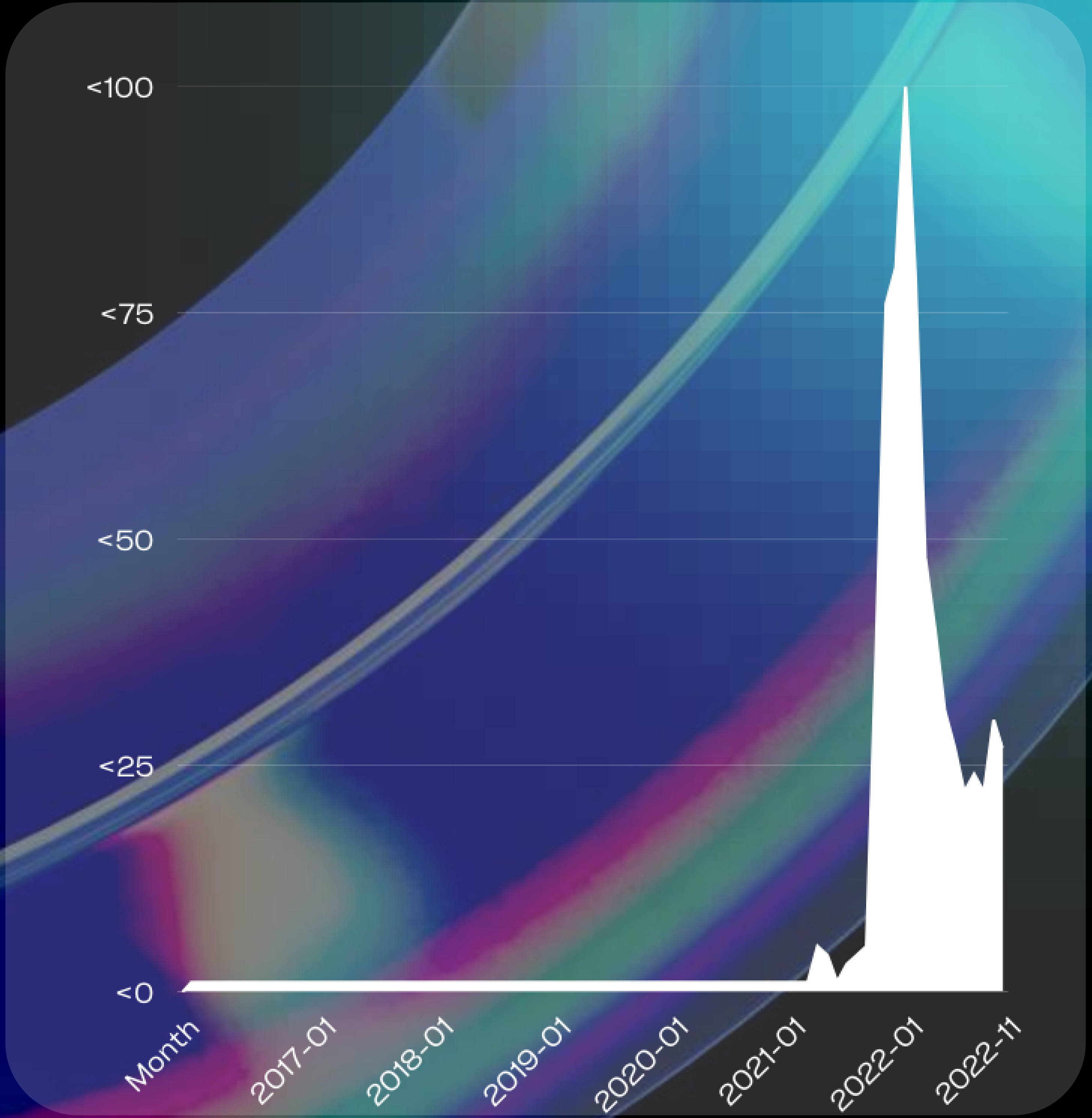
Implement robust computer vision algorithms for marker detection and tracking. Explore techniques like Kalman filtering or model-based tracking to handle occlusions and marker dropout.

Noise and outliers:

Explore signal processing techniques such as low-pass filters or statistical approaches like median filtering to mitigate noise and outliers.

Revenue Model

To offer our tool as a subscription-based service to companies and individuals who require high-quality 3D models of humans with motion capture data. We plan to offer different pricing tiers based on usage and feature access. We also plan to offer customization services and support to enterprise clients.



Source: Google Trends

Student Authors

Anshuman Giramkar

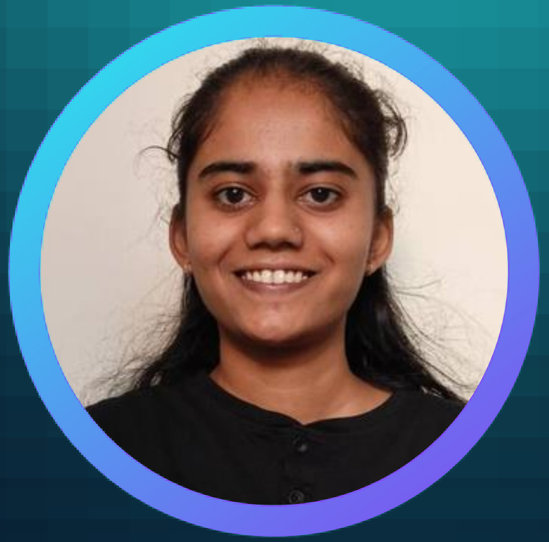
SY Computer Engineering

Shreyas Godse

SY Computer Engineering

Snehal Gupta

SY Computer Engineering



Pooja Indulkar

SY Computer Engineering

SOURCES OF INSPIRATION (References):

- Kato, H., Ushiku, Y., Harada, T. (2019). Moulding Humans: Non-parametric 3D Human Shape Estimation from Single Images. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA.
- Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., Li, H. (2020). PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA.
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA.
- Kolotouros, N., Pavlakos, G., Black, M. J., Daniilidis, K. (2019). Learning to Reconstruct People in Clothing from a Single RGB Camera. IEEE International Conference on Computer Vision (ICCV), Seoul, South Korea.
- Tung, T., Noble, J., Sheikh, Y. (2018). MonoPerfCap: Human Performance Capture from Monocular Video. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA.
- Kanazawa, A., Tulsiani, S., Malik, J. (2018). 3D Human Body Reconstruction from a Single Image via Volumetric Regression. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA.
- Park, J. Y., Florence, P., Straub, J., Newcombe, R., Lovegrove, S. (2019). Learning Implicit Fields for Generative Shape Modeling. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA.
- Chen, L. C., Papandreou, G., Schroff, F., Adam, H. (2017). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834-848.
- Yang, S., Li, Y., Stoyanov, T., Sebe, N., Hauptmann, A. G. (2019). Multi-Garment Net: Learning to Dress 3D People from Images. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA.
- Tan, H., Liu, X., Yu, F., Wang, Y., Darrell, T. (2019). Tex2Shape: Detailed Full Human Body Geometry From a Single Image. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA.
- Guler, R. A., Neverova, N., Kokkinos, I. (2018). DensePose: Dense Human Pose Estimation In The Wild. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA.