Foundational Models 3/3

3D and Robotics

Presenter:

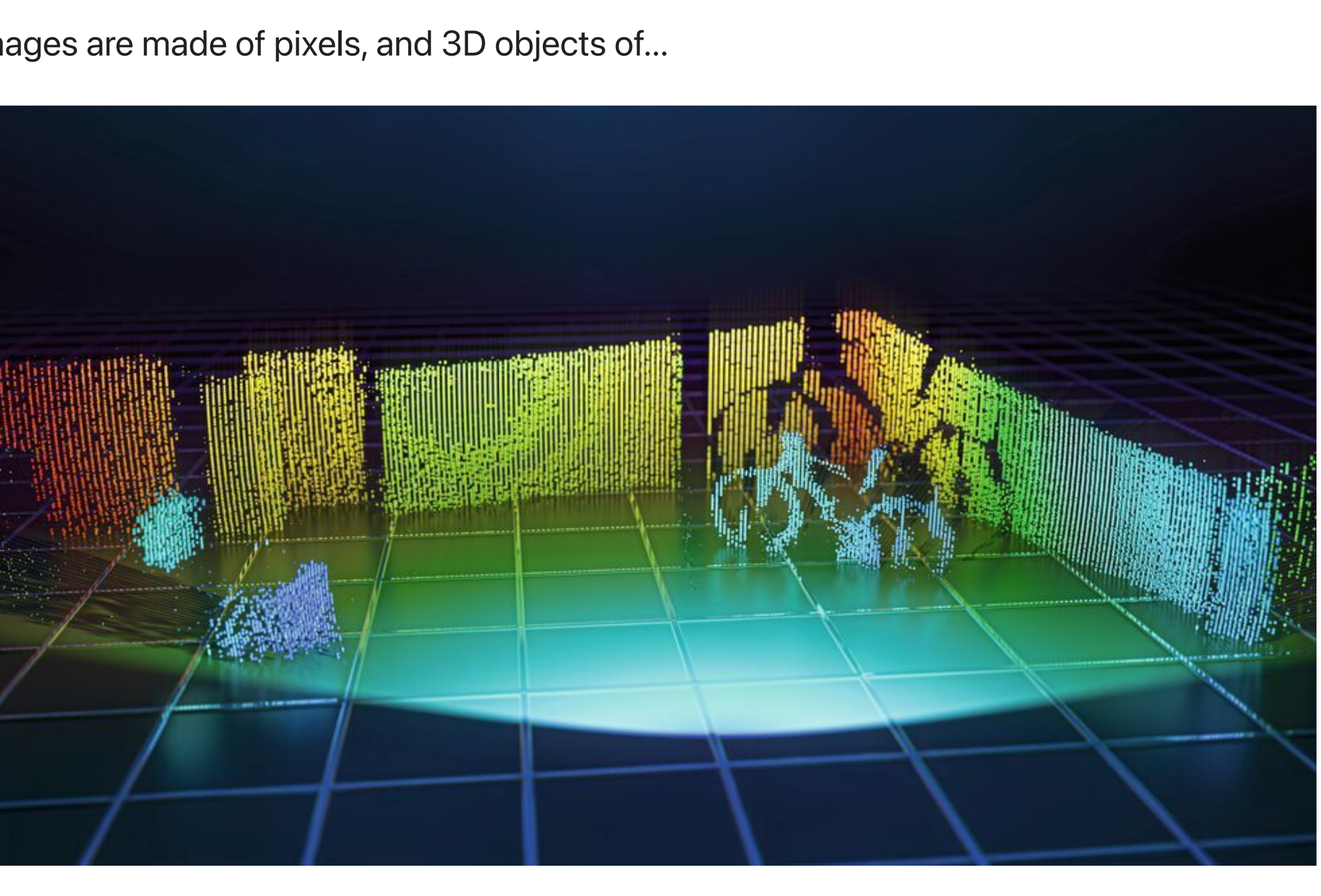
> Anton Akusok (PhD), **Senior Lecturer in IT / Senior Data engineer at Zalando**
> (*email*: anton.akusok@arcada.fi)

Goals

- Learn about foundational models for 3D and actions
- Understand how these models are created

Basics of 3D Sensing

Images are made of pixels, and 3D objects of...



1. Voxels (volumetric pixels)

Dense uniform grid in 3D space.

- **very** heavy on memory
- low-dimensional models e.g. 100 vx in each direction (1,000,000 samples altogether)
- cover full space
- native format for medical 3D scans, some 3D printing

Used for robot action planning that benefits from covering full space

2. Point Cloud

Collection of points in 3D space. Usually represents object surface.

- sparse, very memory efficient
- any resolution
- represent any geometry
- native format for LiDAR scanning or 3D mapping

Used for autonomous vehicles, robot navigation, 3D scanning

2.b. Depth Map

2D image where pixels are distances from object to camera. Basically a special point cloud where all points are visible from one location.

- image representation - easy to store and process by existing methods
- can be computed from stereo vision

Useful as first data input, and as extra input in AI image processing.

3. Multi-view images

Several images of the same scene, usually from a known and fixed camera location.

- very easy to capture
- easy to store and process
- native arrangement for robotics

Useful as data input for 3D reconstruction and robotics.

4. Mesh

Collection of vertices (points), edges (connecting lines) and faces (usually triangles). Typical for 3D computer graphics.

- detailed surface details
- support textures
- extra features for enhanced realism

Used in 3D computer graphics, gaming, 3D modeling.

5. Implicit Surfaces

Mathematical function that gives the distance to nearest surface for every point in 3D space.

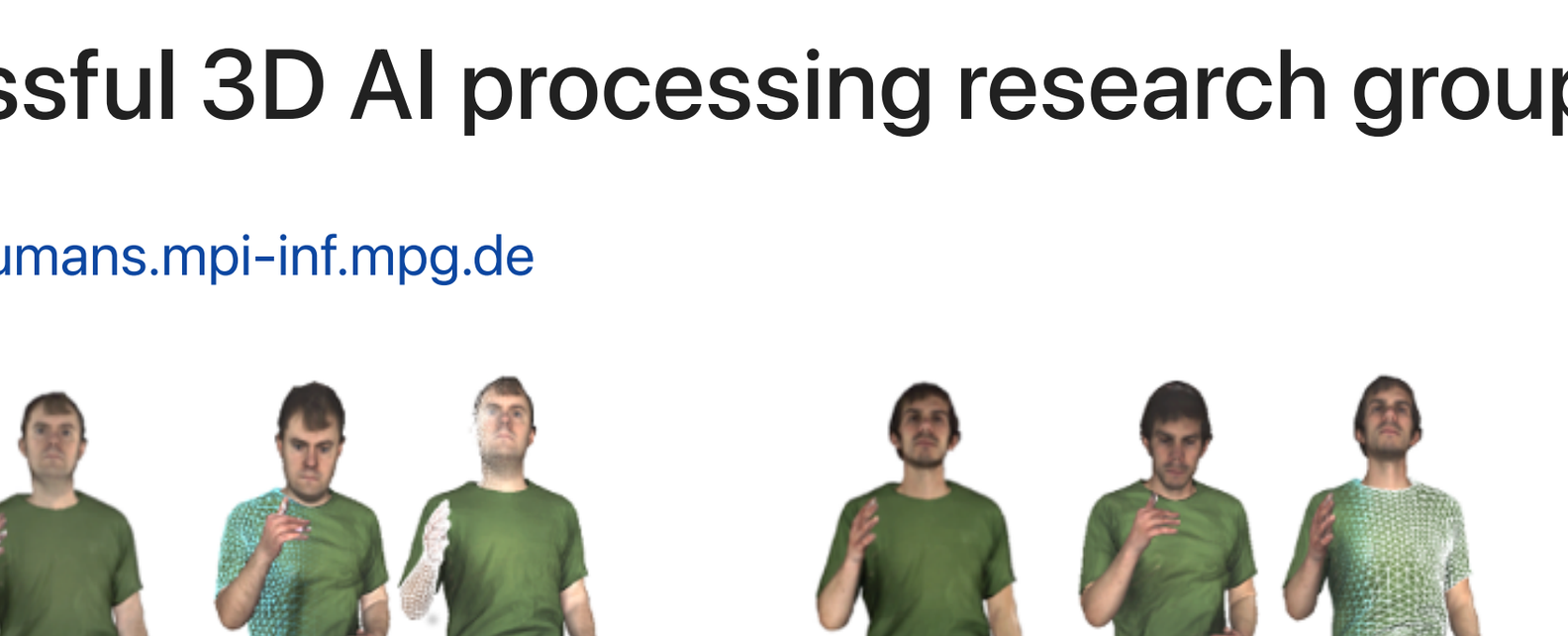
- smooth and continuous surface representation
- very compact
- can represent complex shapes

Useful for physical simulations, computer graphics.

Real Virtual Humans

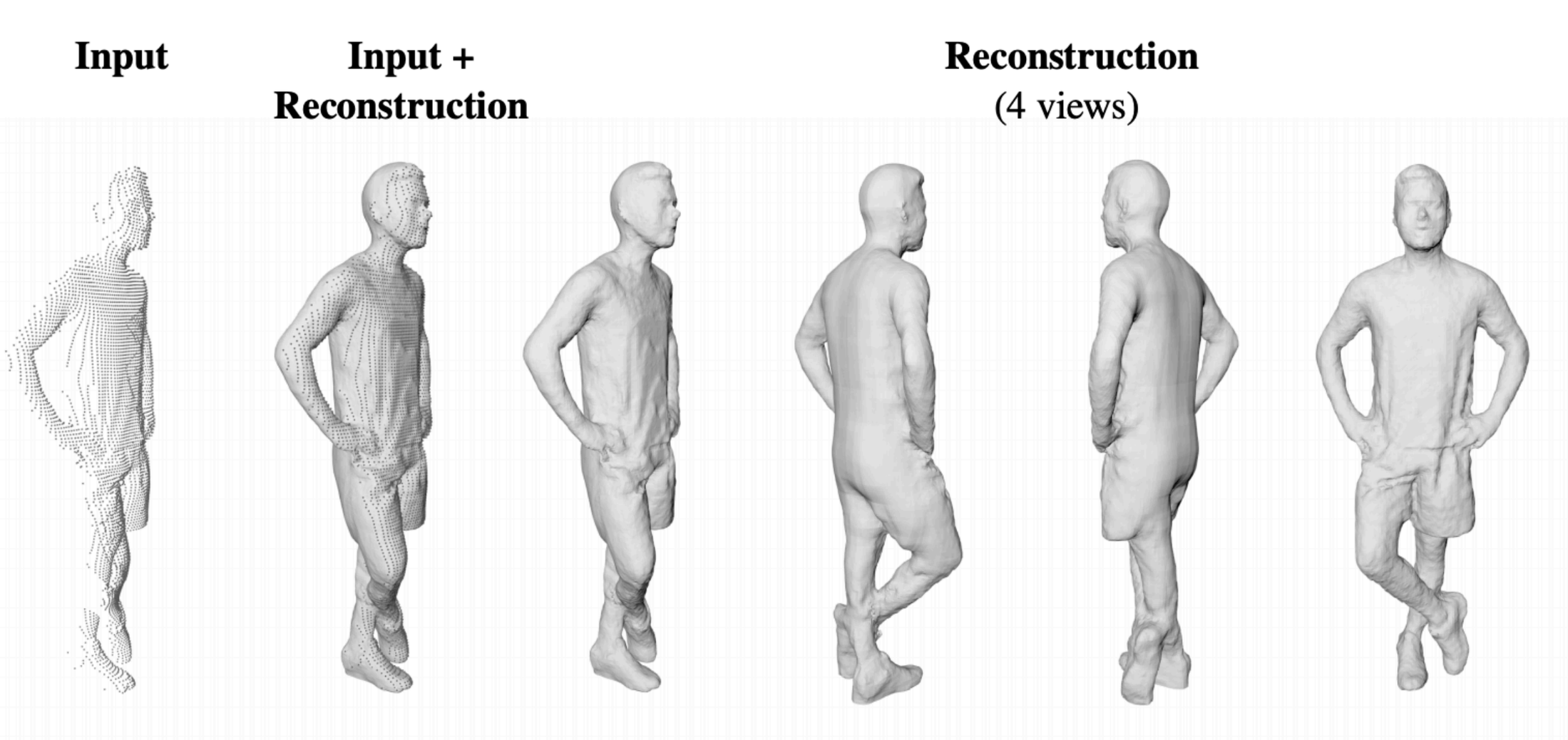
A successful 3D AI processing research group

https://virtualhumans.mpi-inf.mpg.de



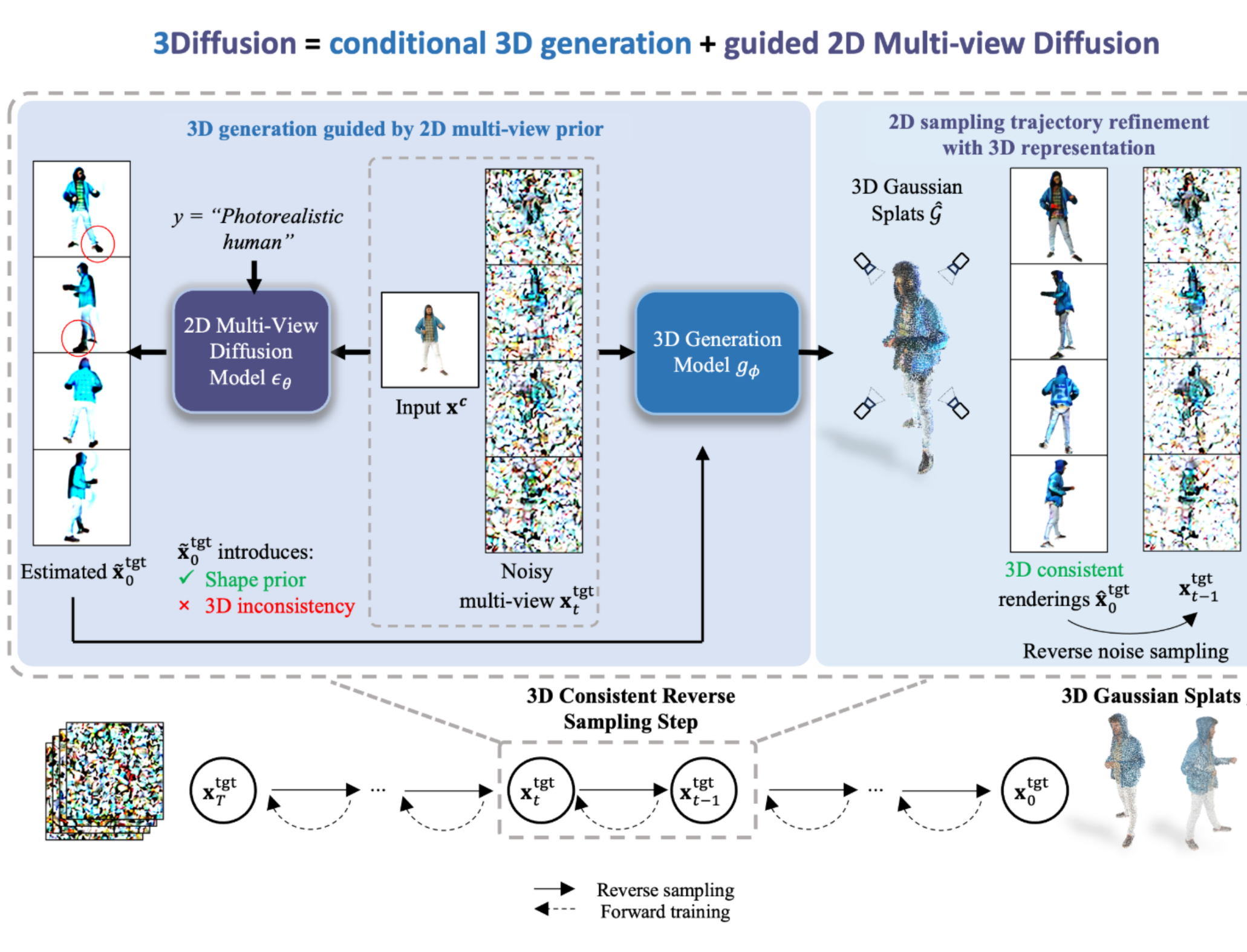
AI human body shape learning

https://virtualhumans.mpi-inf.mpg.de/topics/implicit-shape-and-shape-representation-learning.html



Human 3DDiffusion

https://yuxuan-xue.com/human-3diffusion/
https://www.youtube.com/watch?v=xK9Ebr19jF0

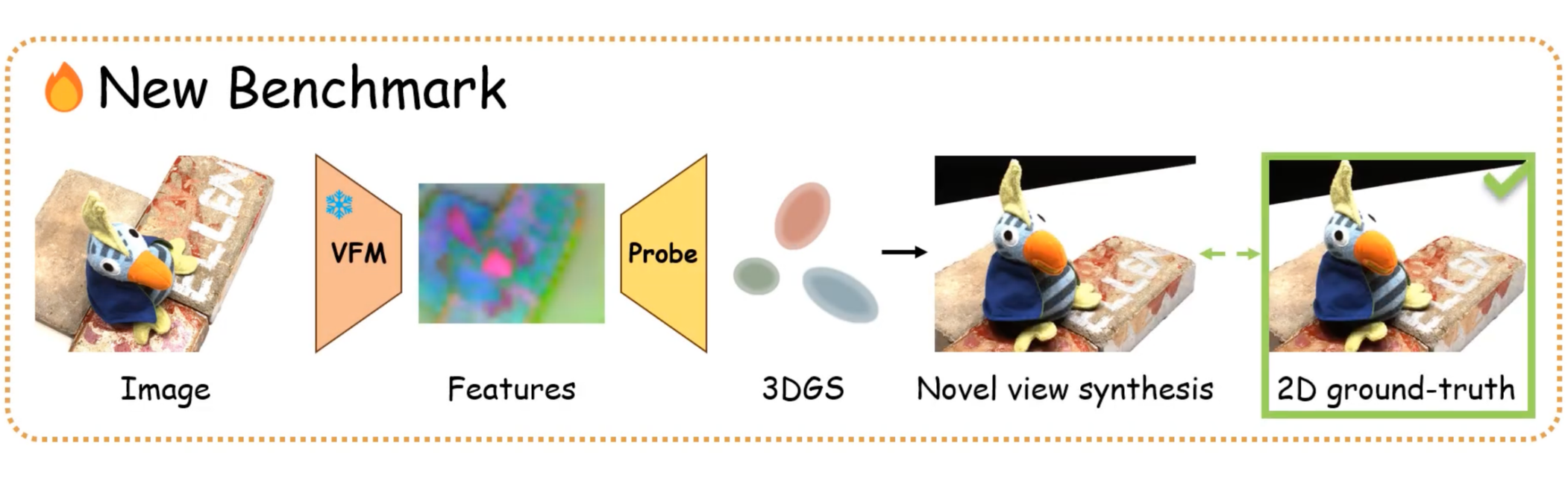


Understanding Visual Foundation Models

Evaluate Visual Foundation Models by generating 2D "views" from 3D model representations. Basically gives a "camera" to walk around and explore the 3D latent space of Visual Foundation Models.

(also gives a good list of visual foundation models)

https://fanegg.github.io/Feat2GS/
https://www.youtube.com/watch?v=4fT5lzcAJqo



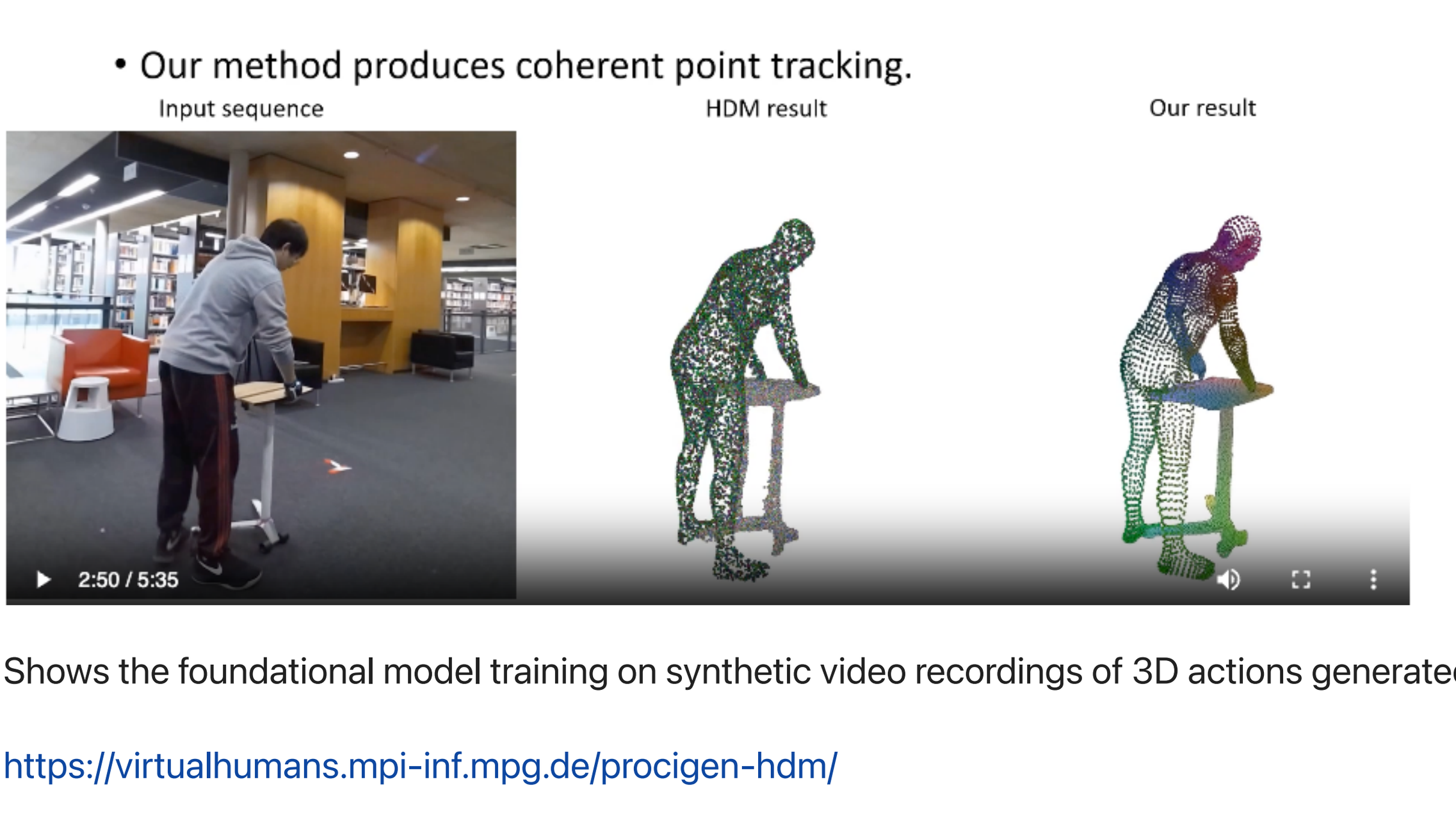
Interactive demo!

https://huggingface.co/spaces/endless-ai/Feat2GS

Demo can be run locally (needs CUDA)
https://github.com/fanegg/Feat2GS

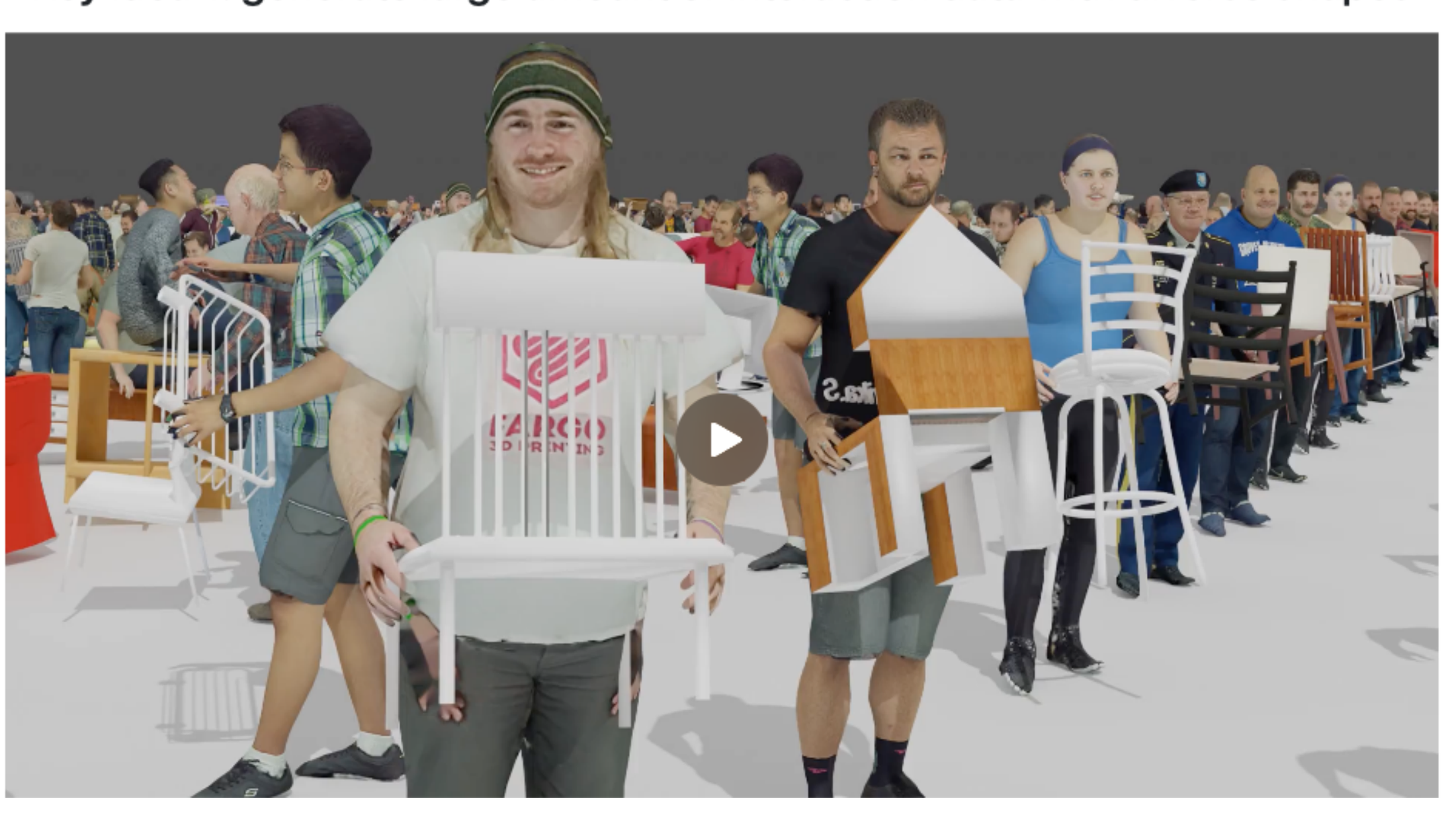
Human-Object Interaction without templates

Coolest presentation!
https://virtualhumans.mpi-inf.mpg.de/InterTrack/
https://www.youtube.com/watch?v=x50HSQJt8DA



Shows the foundational model training on synthetic video recordings of 3D actions generated in a 3D software

https://virtualhumans.mpi-inf.mpg.de/procigen-hdm/



More software and talks from these people

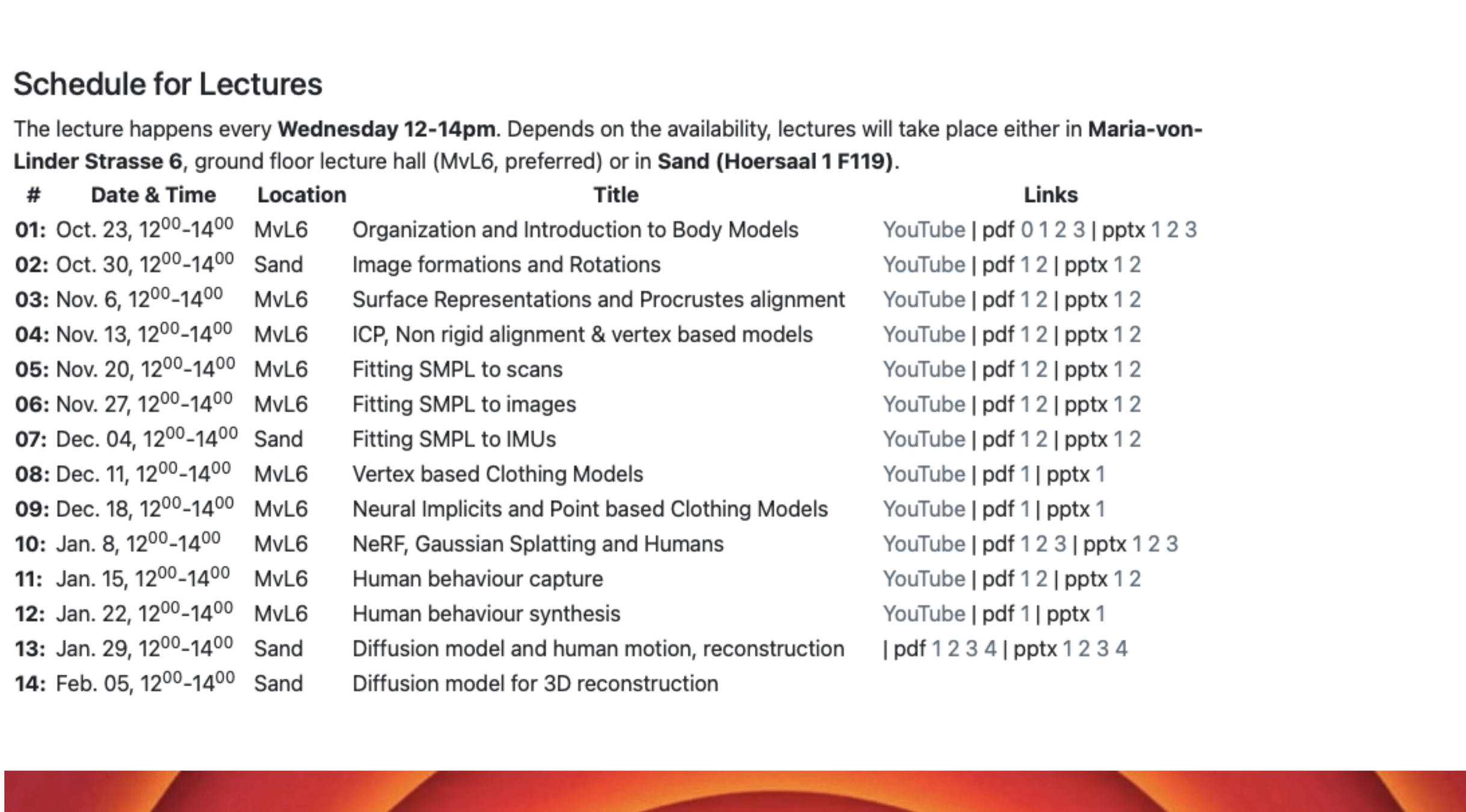
https://virtualhumans.mpi-inf.mpg.de/software.html

https://virtualhumans.mpi-inf.mpg.de/talks.html

"Virtual Humans" the *whole freaking course* with videos and PDFs!

https://virtualhumans.mpi-inf.mpg.de/VH24/

https://www.youtube.com/c/TübingenML/playlists



Schedule for Lectures

The lecture happens every **Wednesday 12-14pm**. Depends on the availability, lectures will take place either in **Maria-von-Linder Strasse 6**, ground floor lecture hall (MvL6, preferred) or in **Sand (Hoersaal 1 F119)**.

#	Date & Time	Location	Title	Links
01:	Oct. 23, 12⁰⁰-14⁰⁰	MvL6	Organization and Introduction to Body Models	YouTube | pdf 0 1 2 3 | pptx 1 2 3
02:	Oct. 30, 12⁰⁰-14⁰⁰	Sand	Image formations and Rotations	YouTube | pdf 1 2 | pptx 1 2
03:	Nov. 6, 12⁰⁰-14⁰⁰	MvL6	Surface Representations and Procrustes alignment	YouTube | pdf 1 2 | pptx 1 2
04:	Nov. 13, 12⁰⁰-14⁰⁰	MvL6	ICP, Non rigid alignment & vertex based models	YouTube | pdf 1 2 | pptx 1 2
05:	Nov. 20, 12⁰⁰-14⁰⁰	MvL6	Fitting SMPL to scans	YouTube | pdf 1 2 | pptx 1 2
06:	Nov. 27, 12⁰⁰-14⁰⁰	MvL6	Fitting SMPL to images	YouTube | pdf 1 2 | pptx 1 2
07:	Dec. 04, 12⁰⁰-14⁰⁰	Sand	Fitting SMPL to IMUs	YouTube | pdf 1 2 | pptx 1 2
08:	Dec. 11, 12⁰⁰-14⁰⁰	MvL6	Vertex based Clothing Models	YouTube | pdf 1 | pptx 1
09:	Dec. 18, 12⁰⁰-14⁰⁰	MvL6	Neural Implicits and Point based Clothing Models	YouTube | pdf 1 | pptx 1
10:	Jan. 8, 12⁰⁰-14⁰⁰	MvL6	NeRF, Gaussian Splatting and Humans	YouTube | pdf 1 2 3 | pptx 1 2 3
11:	Jan. 15, 12⁰⁰-14⁰⁰	MvL6	Human behaviour capture	YouTube | pdf 1 2 | pptx 1 2
12:	Jan. 22, 12⁰⁰-14⁰⁰	MvL6	Human behaviour synthesis	YouTube | pdf 1 | pptx 1
13:	Jan. 29, 12⁰⁰-14⁰⁰	Sand	Diffusion model and human motion, reconstruction	| pdf 1 2 3 4 | pptx 1 2 3 4
14:	Feb. 05, 12⁰⁰-14⁰⁰	Sand	Diffusion model for 3D reconstruction