MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

University of Cyprus

University of Cyprus

MAI645 - Machine Learning for Graphics and Computer Vision

Andreas Aristidou, PhD

Spring Semester 2023

Co-financed by the European Union
Connecting Europe Facility

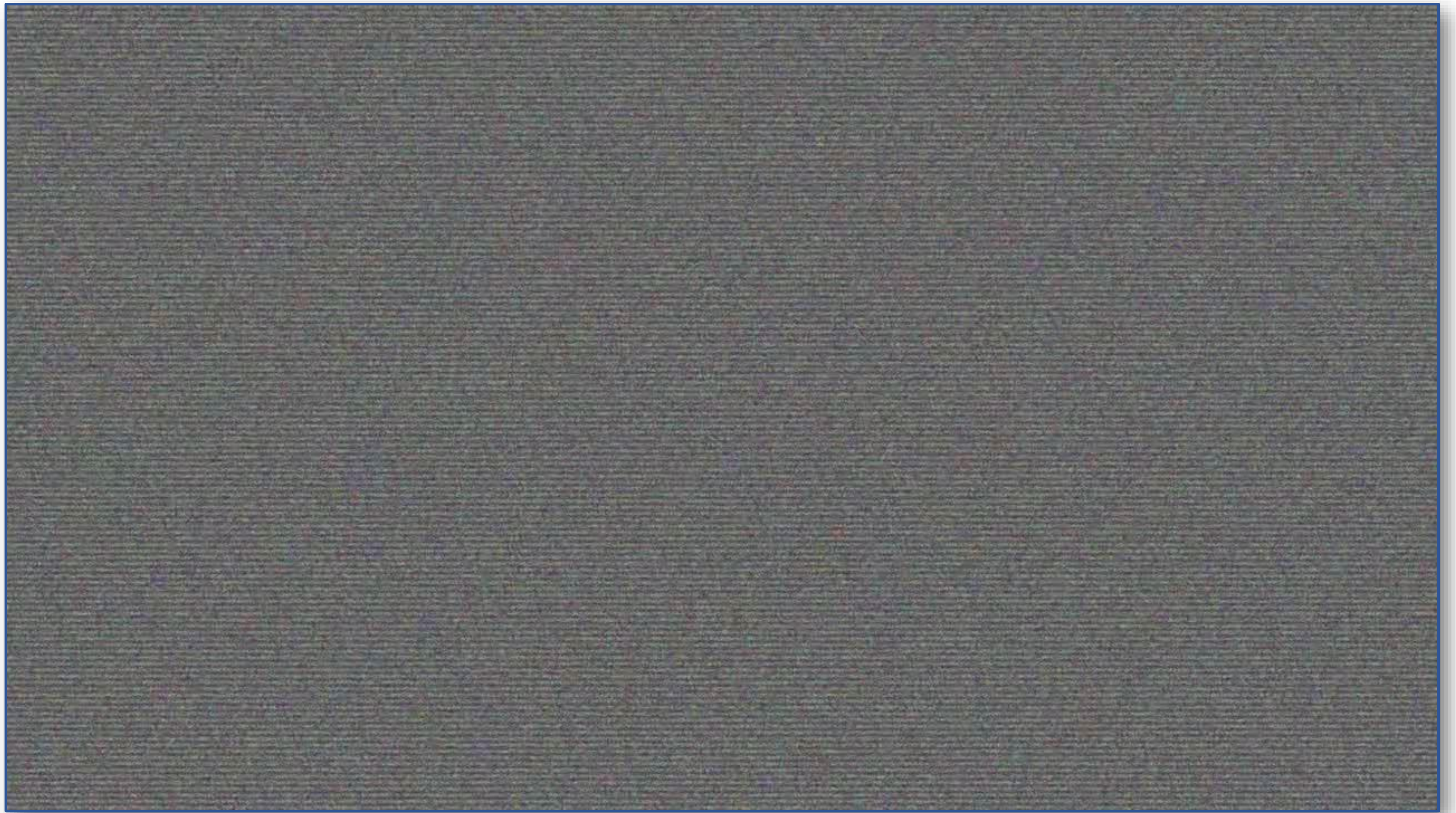
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Character Animation

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How does the magic happen?



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How do we make this movie?

Uncharted 4: https://youtu.be/zL46dpNEPPA

Modeling

- Geometry
- Materials
- Lighting

Animation

- How do they move?

Rendering

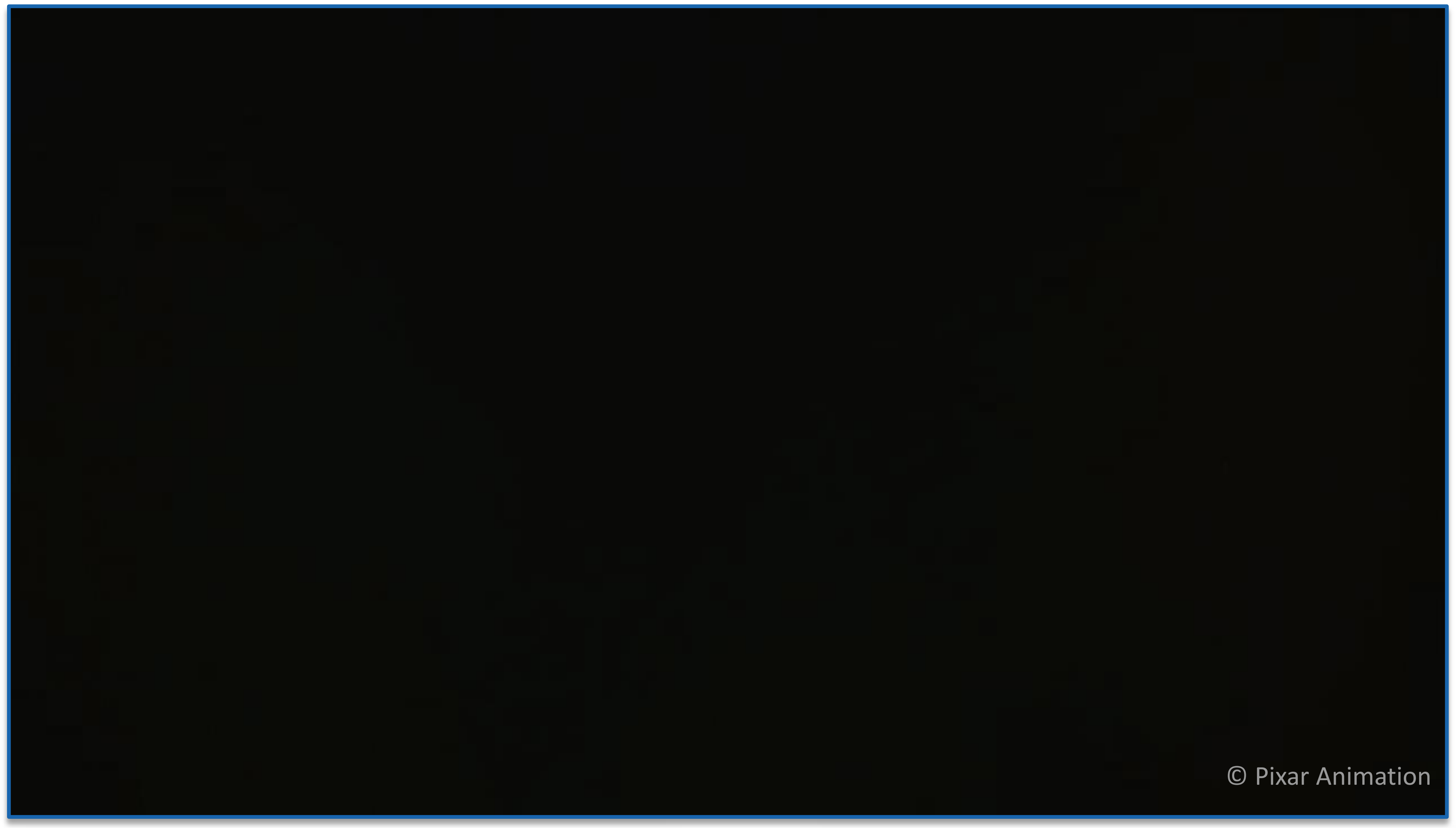
- Shadows
- Camera
- Special effects
- Post-processing

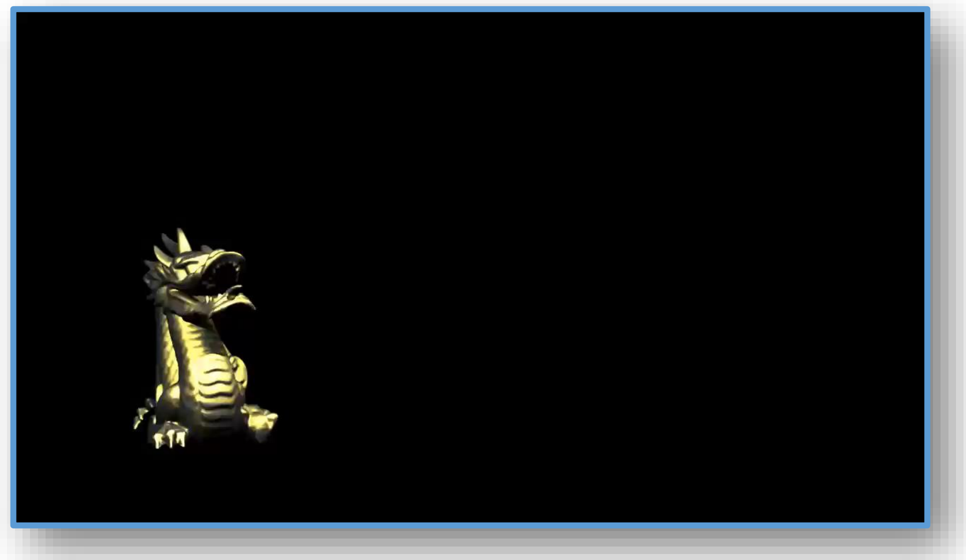
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Motivation

- Bring Animated characters to life
  - Animator analogous to film actors
- Many applications use **character** or **object** animation
  - Entertainment technology (e.g., films, games)
  - Virtual, or augmented reality
  - Simulations, demonstrations, or training systems
- Other forms of animation?
  - Trees, liquids, animals, clouds, etc.
- Other Important factors in character animation
  - Lighting, Rendering, etc.





Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Introduction to Animation

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Moving Picture & Animation

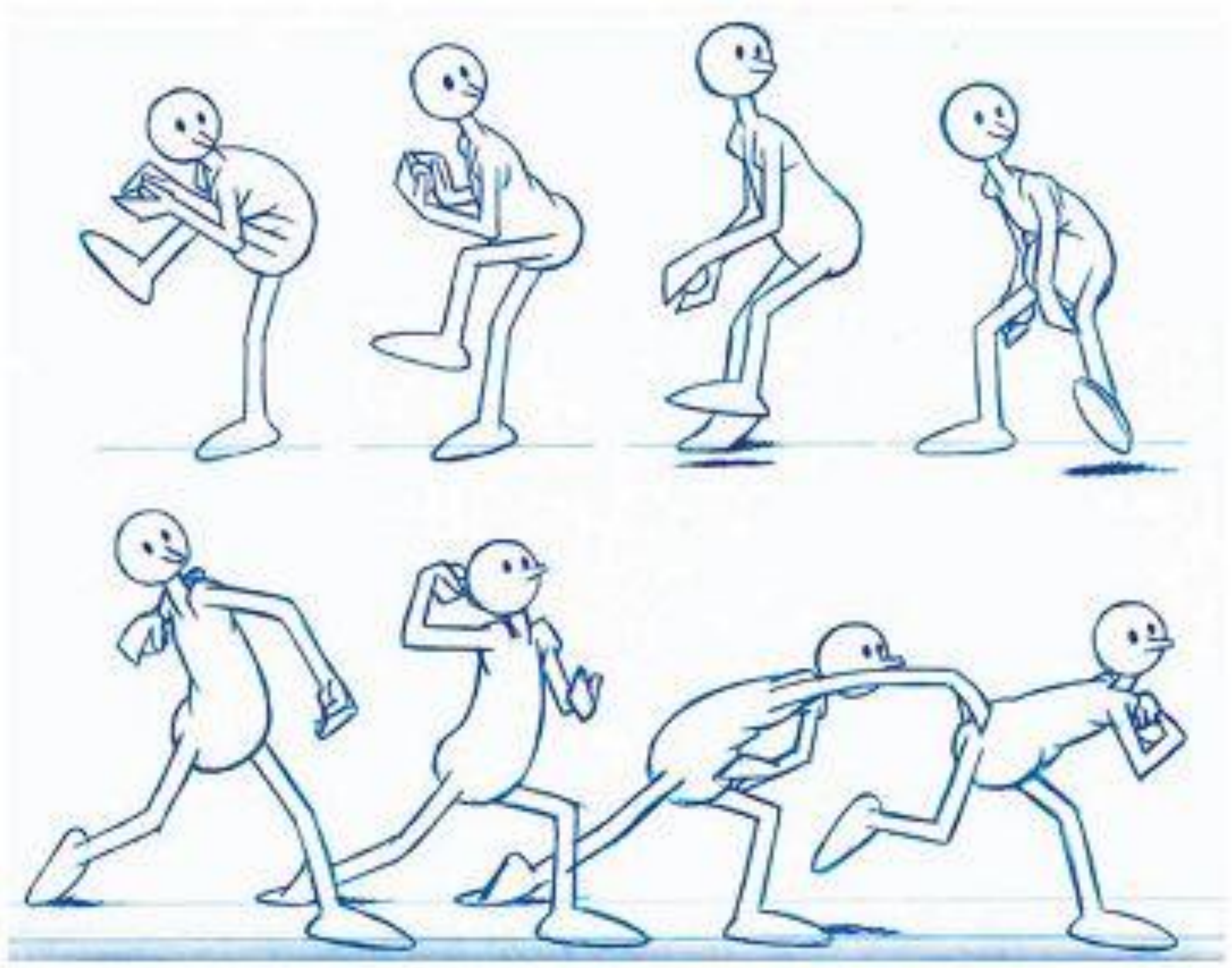
- The perception of motion is based on two optical illusions, the phi phenomenon and beta movement.
  - **phi** is an optical illusion whereas we perceive motion from fast luminous impulses in sequence. Our visual system “fills in” the missing information.
  - **beta** movement is the illusion of motion created when stimuli changes position in a sequence of images. Instead of being perceived as a series of images we perceive movement.
  - Quick succession of images (frames) causes this sensation of movement (1/25sec)

Phi Phenomenon

Beta Movement

Moving Picture & Animation

- Below 16 images/sec flicker is observed.
- Movies play at 24 images/sec.
- ~10 images/sec still provide sensation of movement.
- Traditional animation was created “on twos”
  - A new image every second frame.
- Faster motions are executed “on ones”
  - A new image every frame.

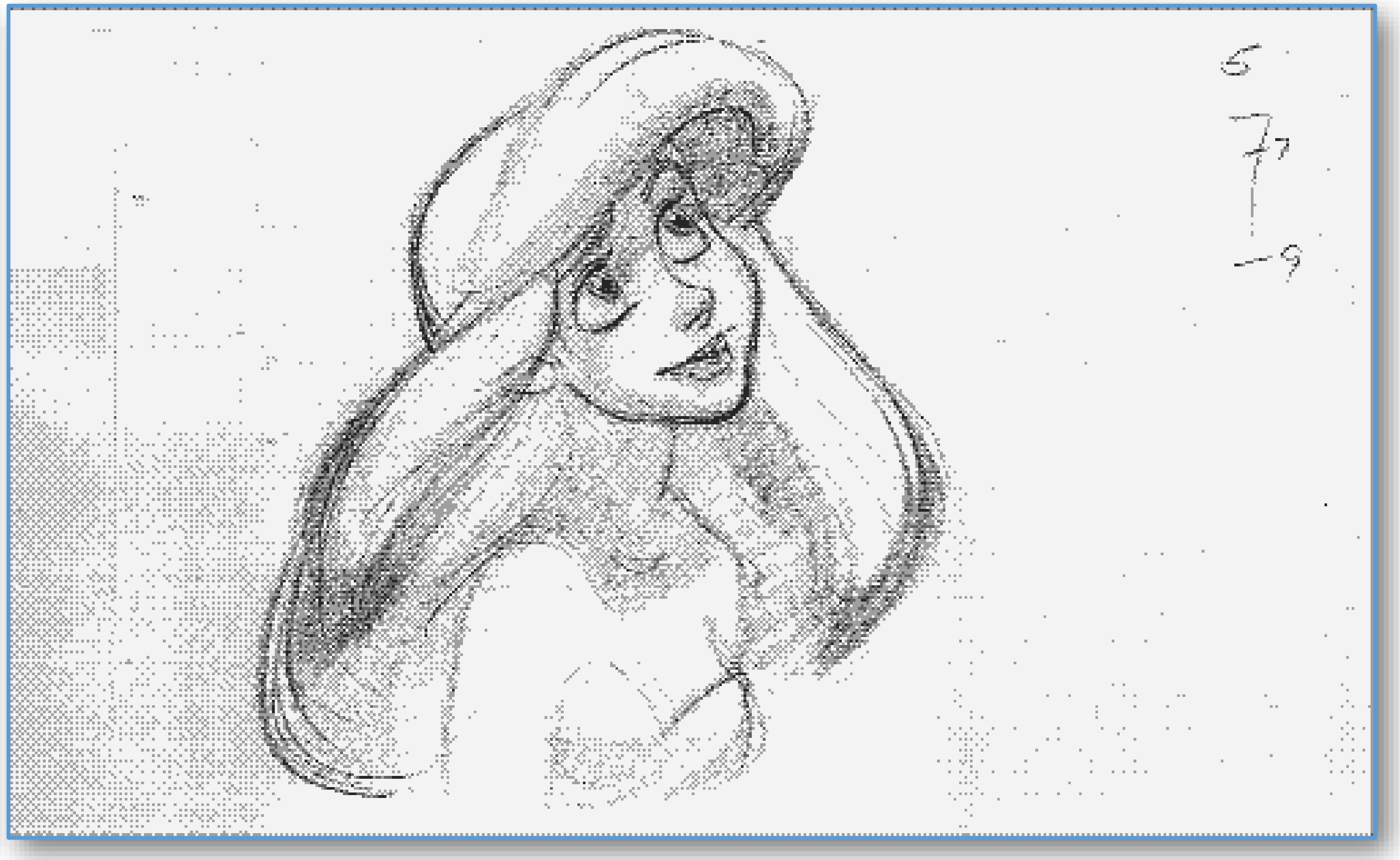


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Moving Picture & Animation

Moving Picture & Animation



Moving Picture & Animation

- Senior animators draw keyframes (important/extreme shots)
- Junior animators (inbetweeners) fill in the in-between frames

A cartoon animation may require <u>thousands</u> of hand-drawn images

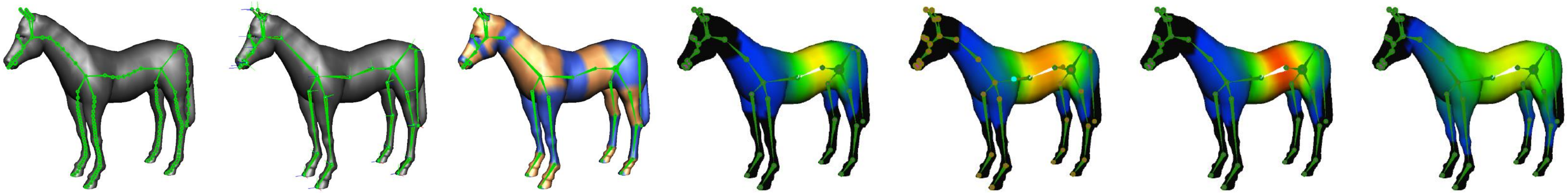
What is Computer Animation?

- **Computer Animation** is the branch of computer graphics interested in developing techniques for creating moving images.
- Computer Animation is the modernized brother of traditional animation.

Co-financed by the European Union
Connecting Europe Facility

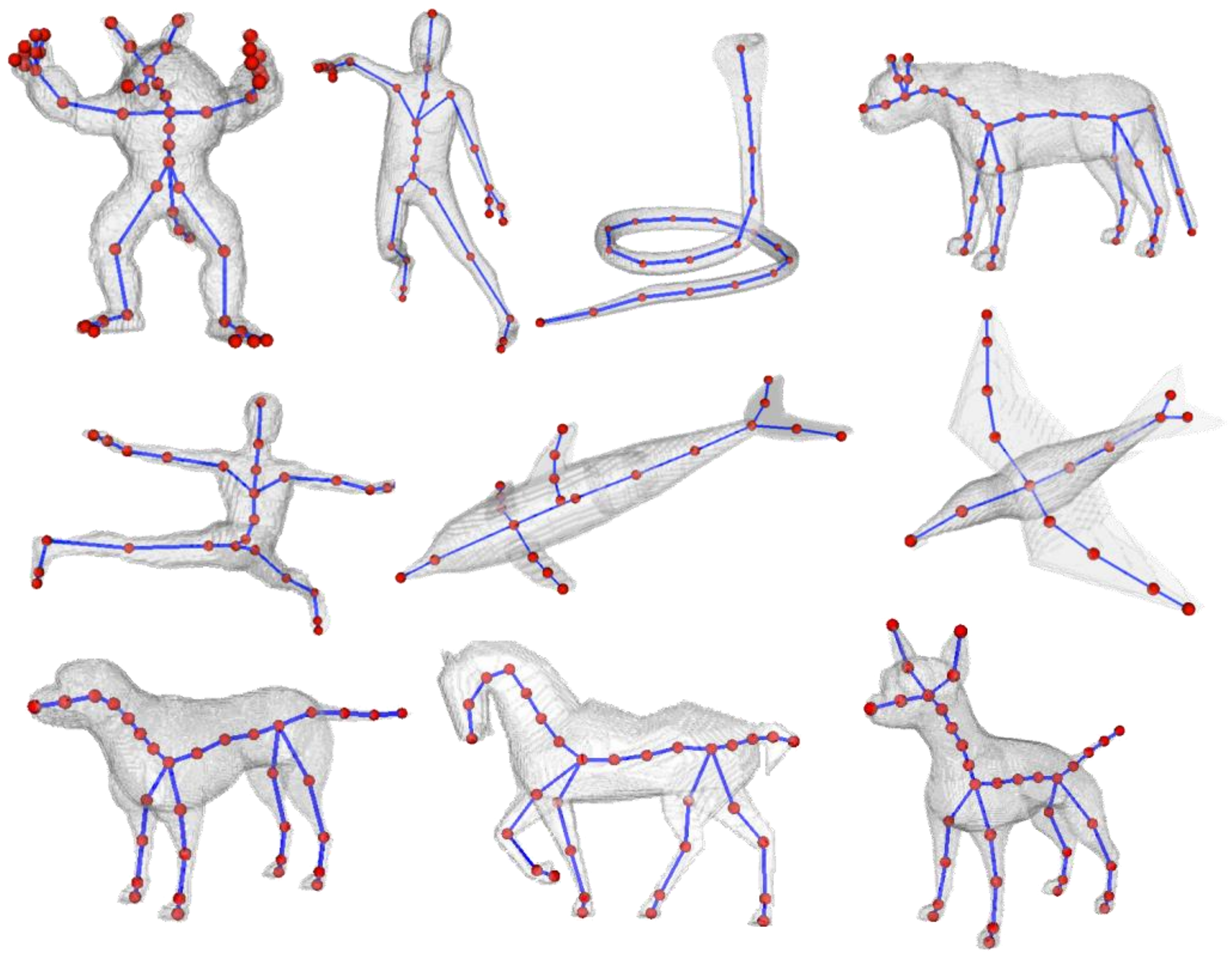
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Character rigging & skinning



Rigging

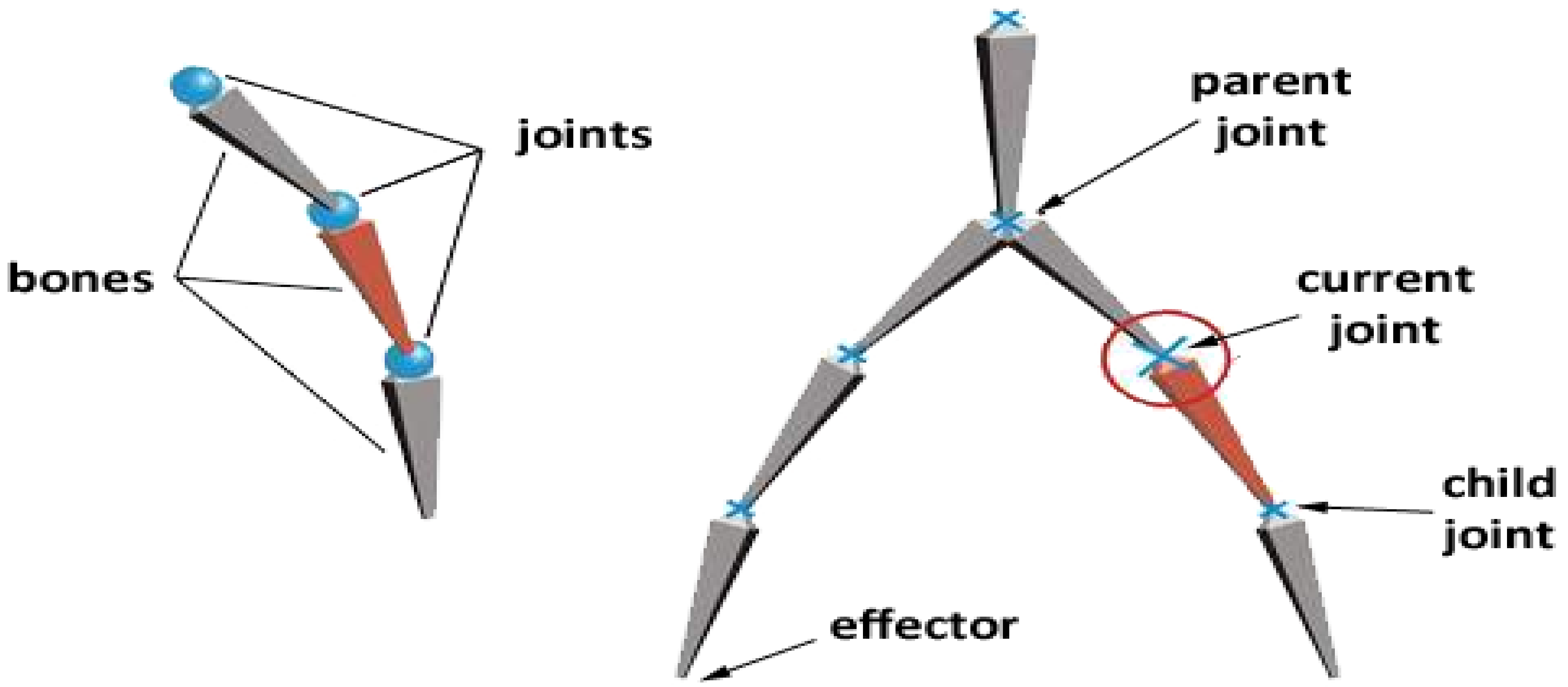
- 3D **rigging** is the process of creating a skeleton for a 3D model so it can move.
- A 'rig' has numerous degrees of freedom (DOFs) that can be used to control various properties.
- One character could have several rigs. One rig could control several characters…



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Rigging: *The Rig*

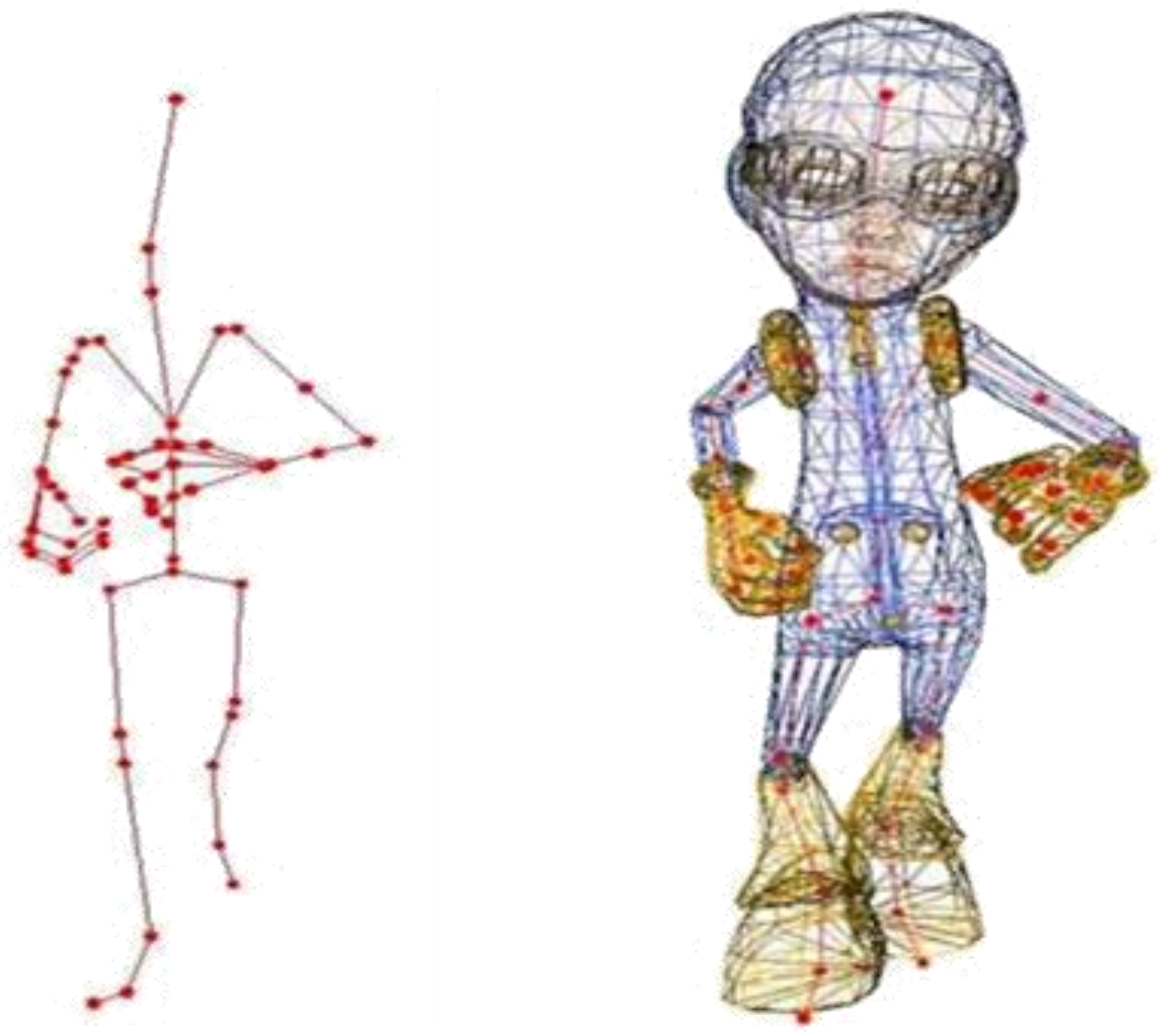
- A skeletal system (**rig**) is comprised of kinematic chains:
  - A hierarchical set of interconnected bones
  - A chain:
    - starts from a **root**,
    - it has multiple **bones**,
    - connected by **joints**, and
    - ends at the **end-effector**.



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Rigging: *The Rig*

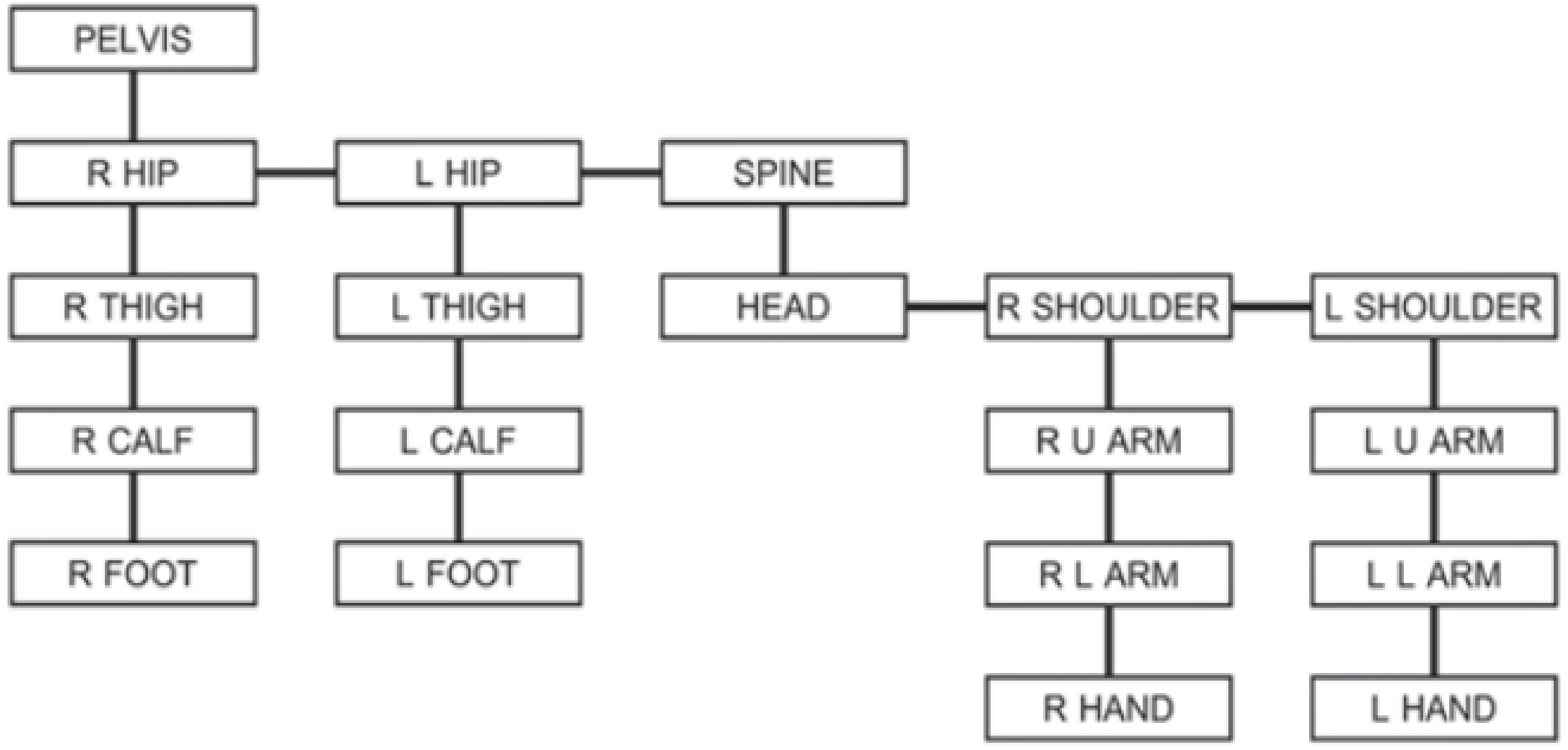
- A skeleton allows higher-level control of the character's animation.
- The skeleton is only a control mechanism – it is not rendered into the final image.
- Typically, there are many constraints.

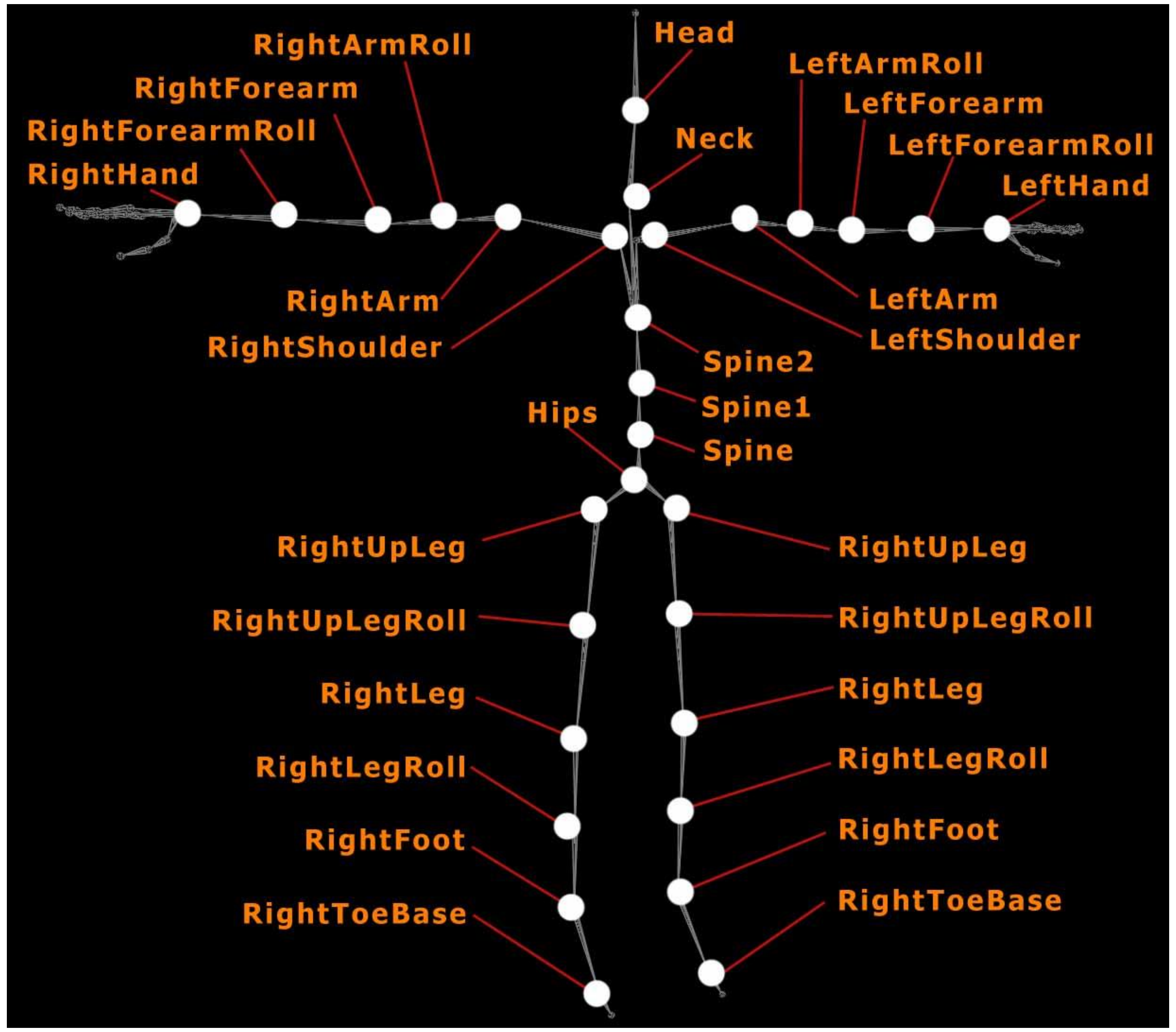


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

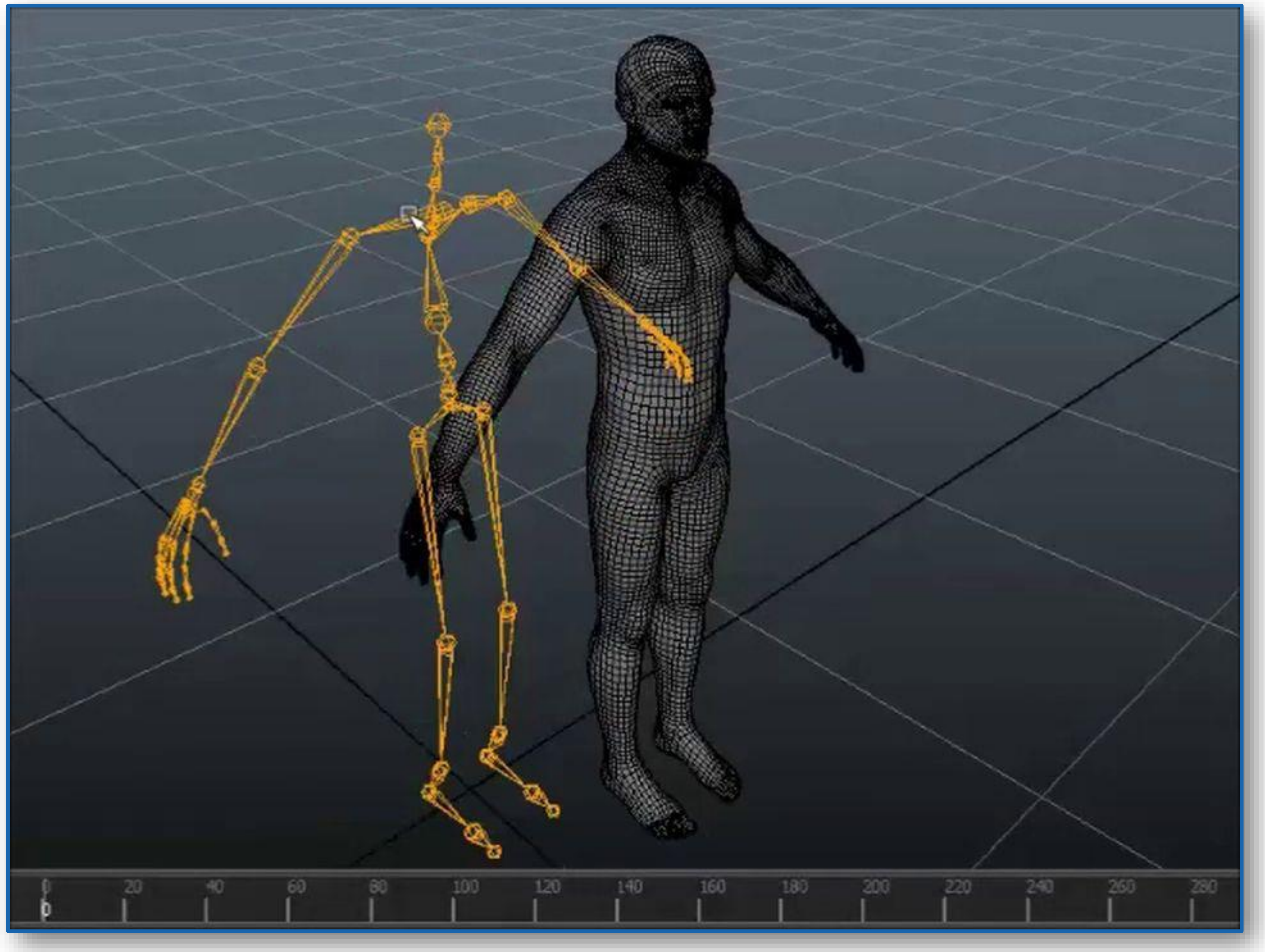
Rigging: *The Rig*





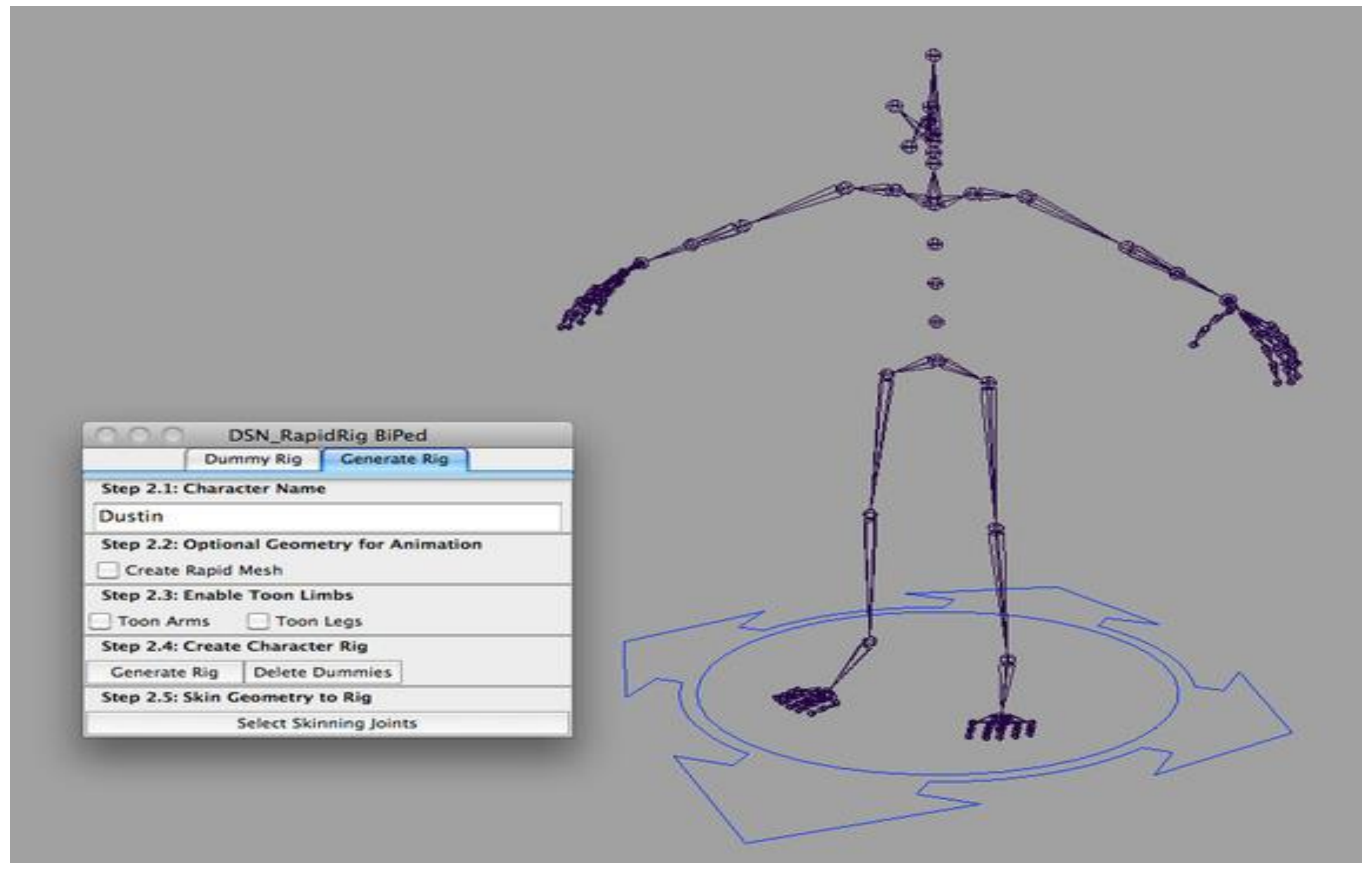
Skinning

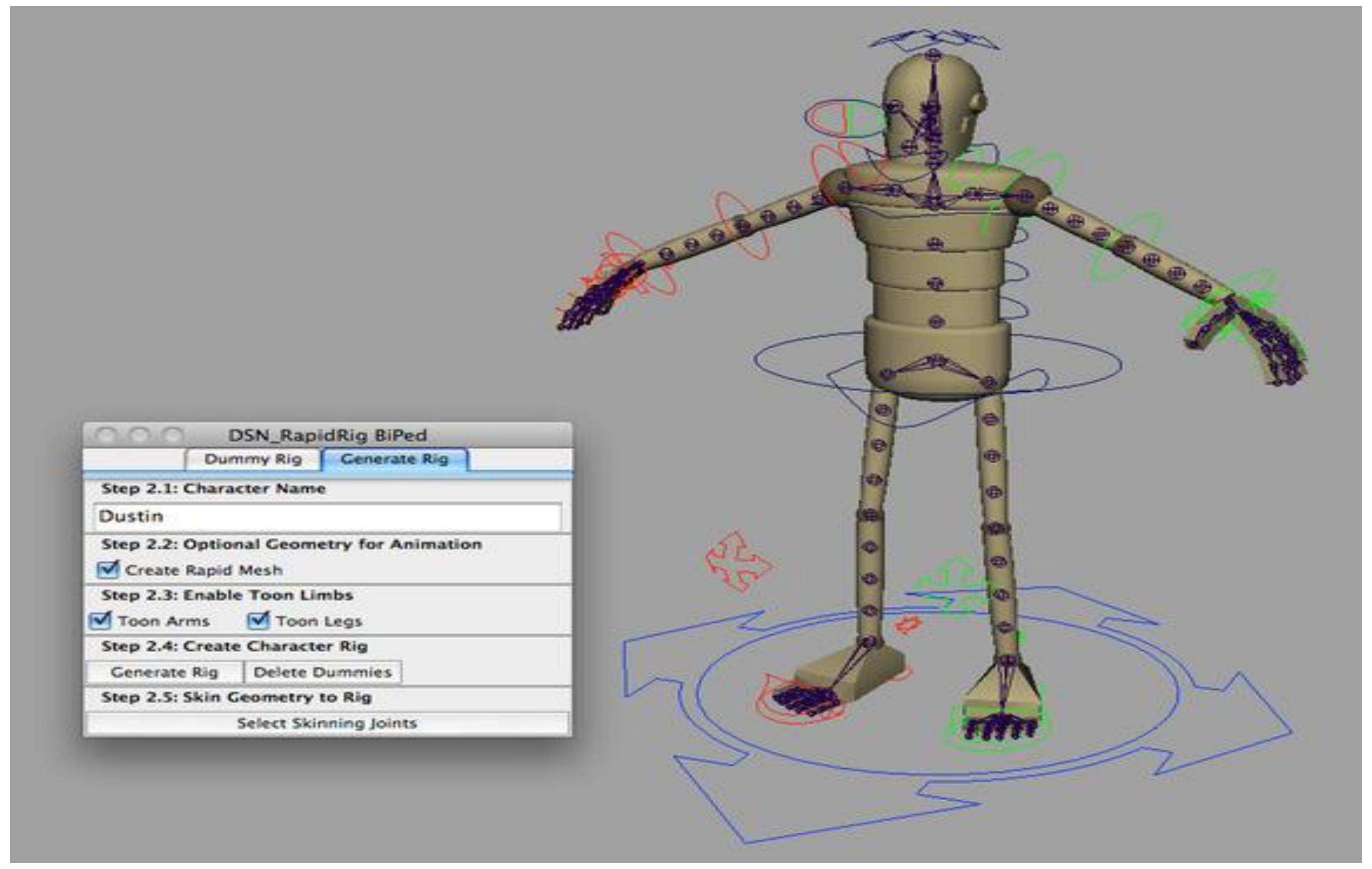
- Skinning.
  - Attach a mesh (“**skin**”) to the skeletal system of the character.
- The **skin** is represented as a polygon mesh, e.g., a set of vertices, or a parametric surface



Skinning: *The Skin*

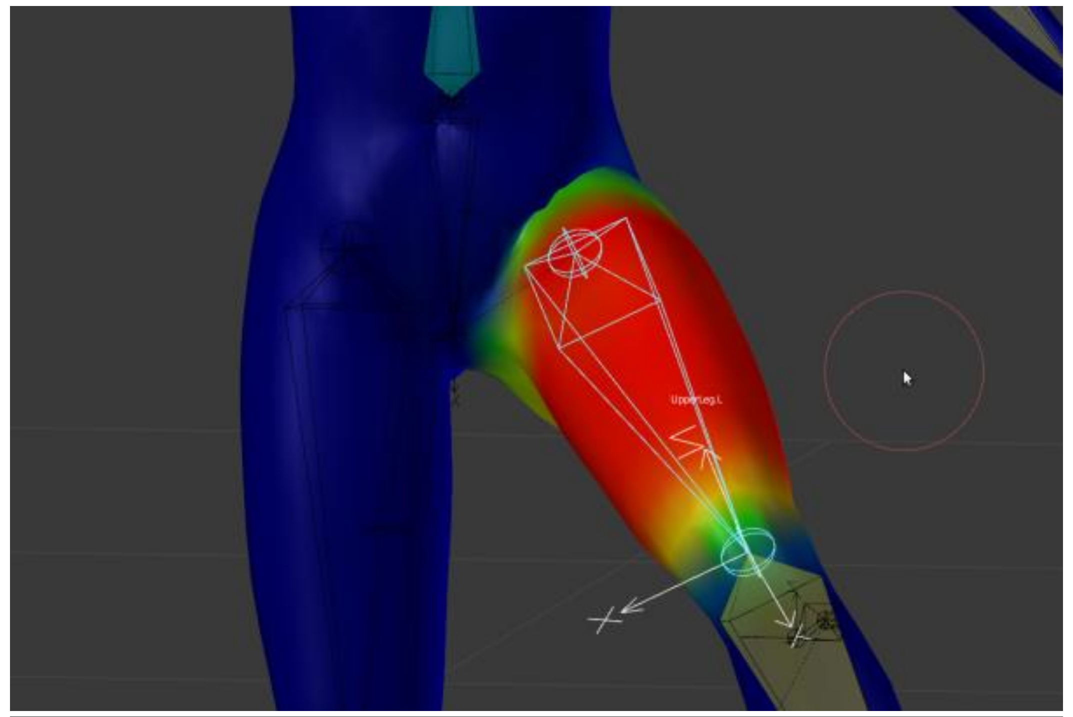
- We **bind** the skeleton to the mesh when we first associate them.
  - The **T-pose** (or “**bind pose**”) refer to the initial transformation matrices of the rig and skin when they are first associated.
  - The T-pose defines a coordinate system used later when animating the skin via the skeleton.
  - The T-pose is a convention used because:
    - modeling the mesh and the skeleton is easier, using symmetry.
    - rigging is much easier when the limbs are spread apart.

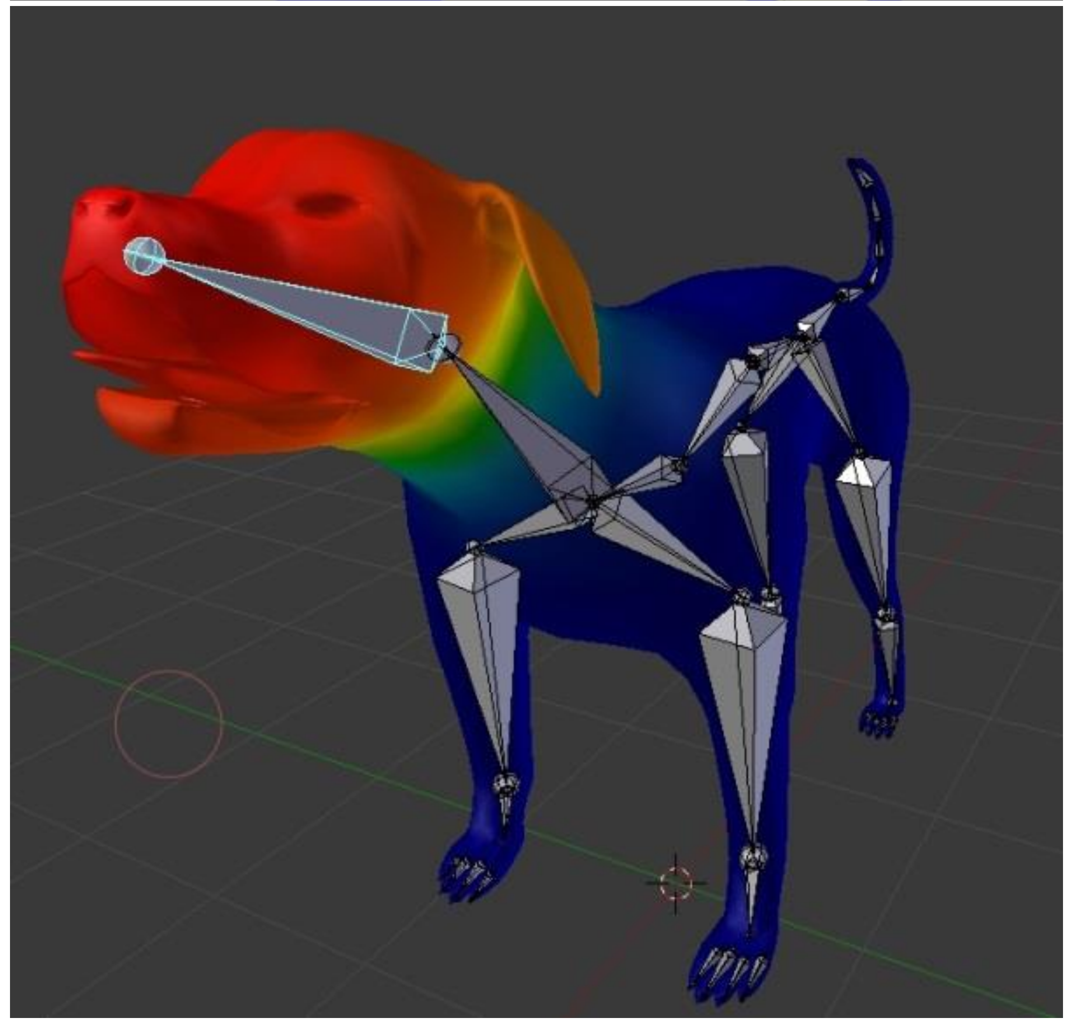




Skinning: *The Skin*

- Each vertex is associated with a bone in the skeleton, and moves relative to that bone.
- Each vertex is multiplied by several “weighted” transformation matrices that provide the influence factor each bone has to the vertex, and the results are added together.
  - The skin’s vertices can then be assigned weights.
    - Rigid skinning: 1 bone per vertex (weight = 1.0)
    - Smooth skinning: Multiple bones per vertex (weights != 1.0)





Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Texture

Skeletal Animation

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What 3D character animation involves?

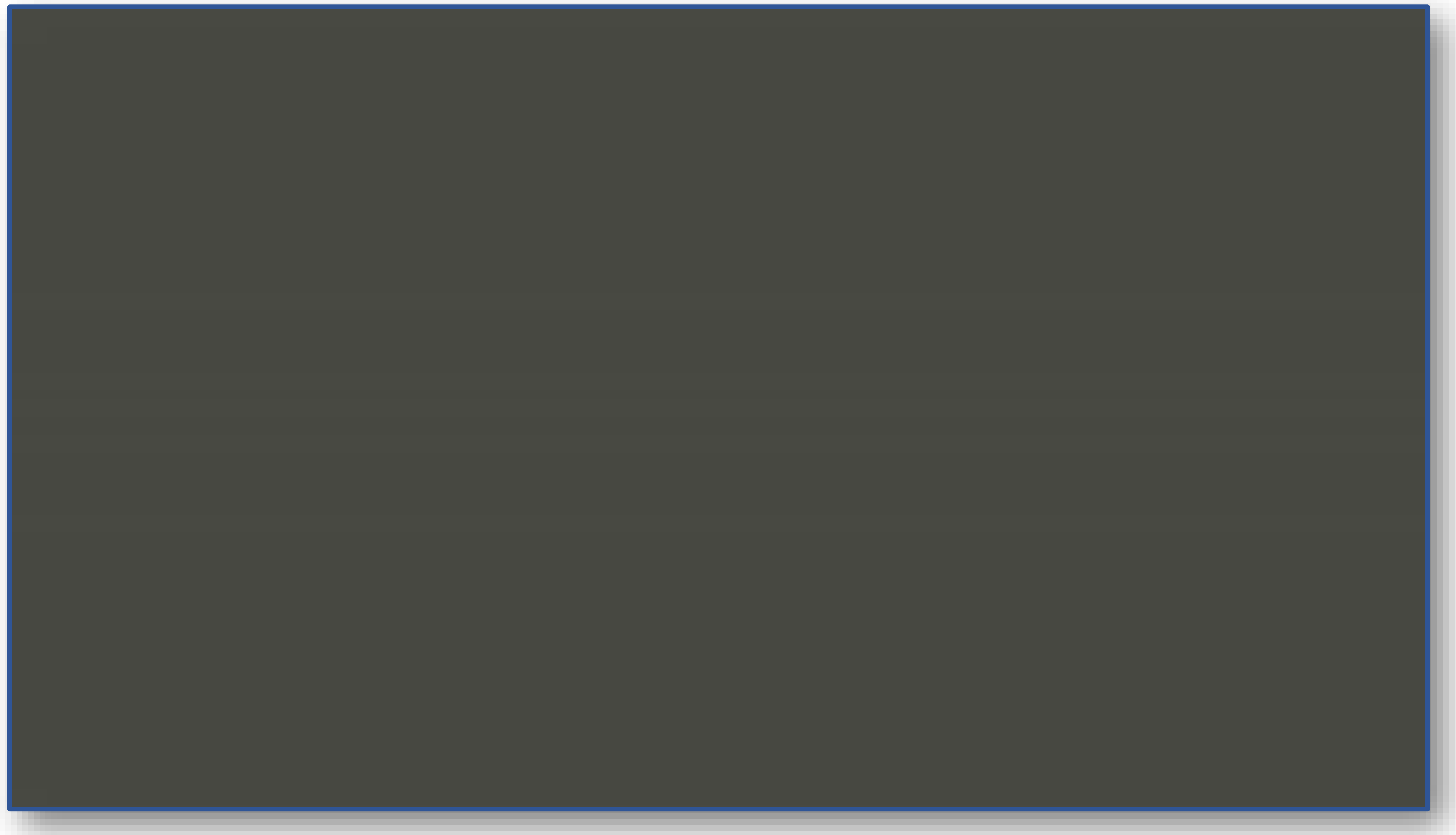
Animating characters can be broken down to:

- **Skeletal animation** – animating their main body parts.

What 3D character animation involves?

Animating characters can be broken down to:

- **Skeletal animation** – animating their main body parts.
- **Facial animation** – animating their facial features.



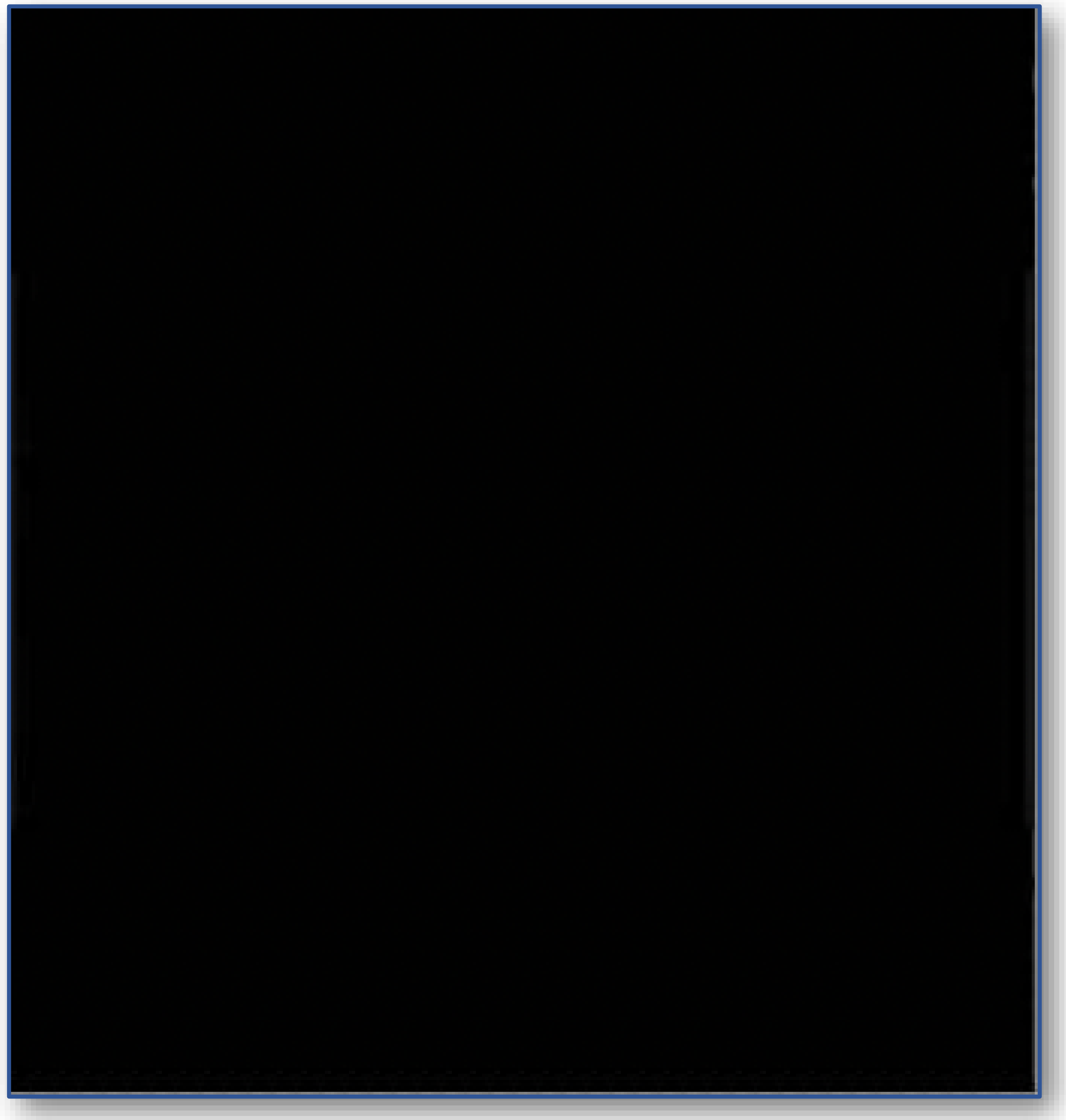
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What 3D character animation involves?

Animating characters can be broken down to:

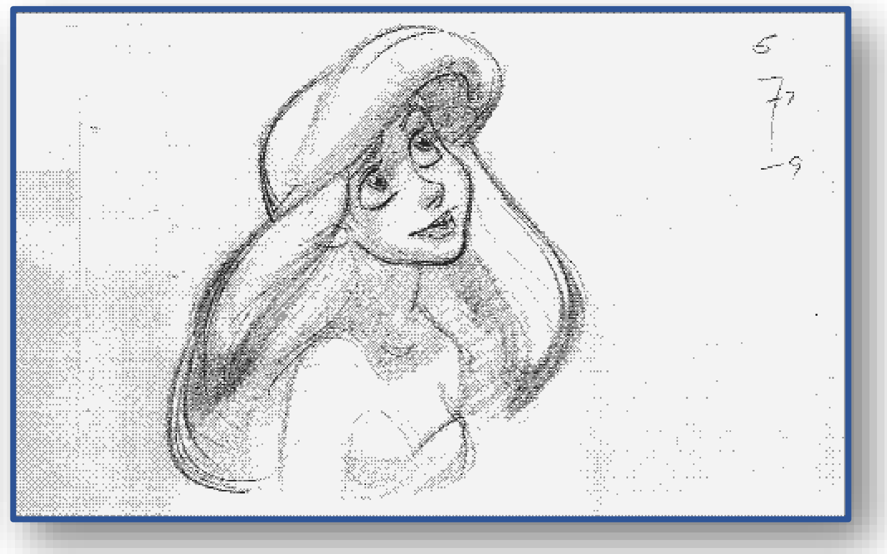
- **Skeletal animation** – animating their main body parts.
- **Facial animation** – animating their facial features.
- **Hair** (and **fur)** animation



Co-financed by the European Union
Connecting Europe Facility

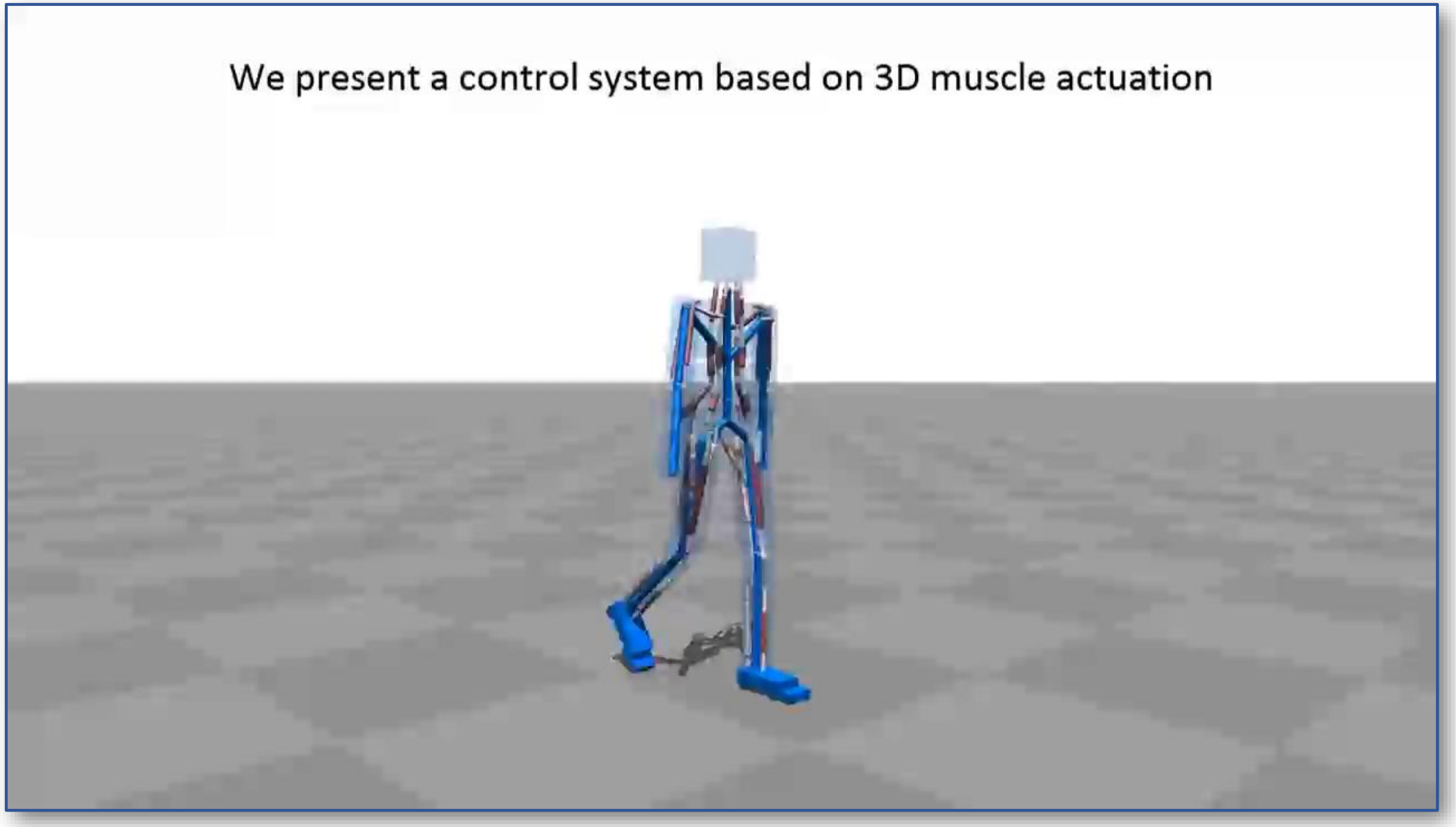
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

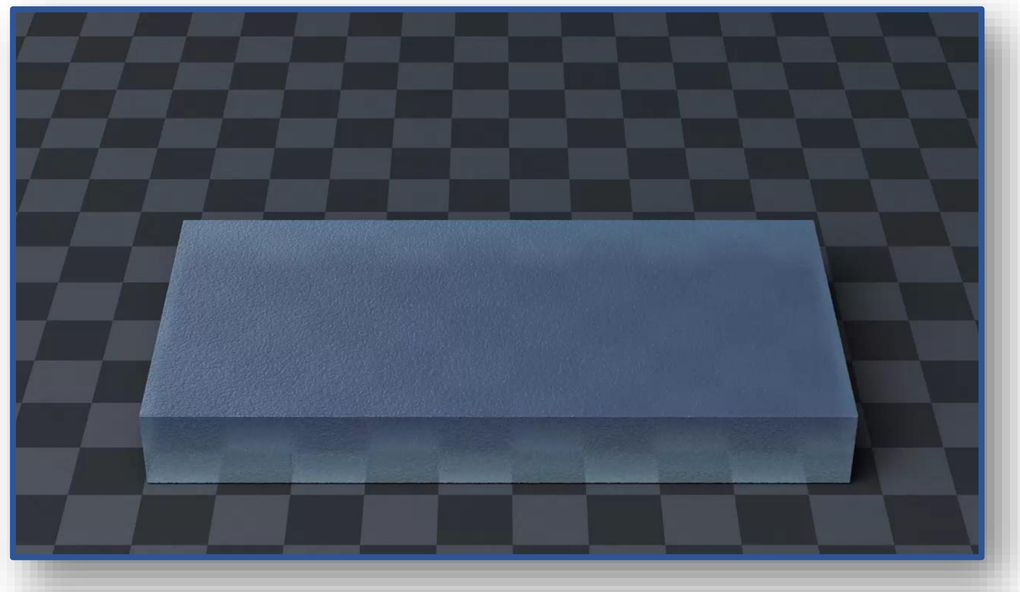
MAI4CAREU | Master programmes in Artificial Intelligence 4 Careers in Europe

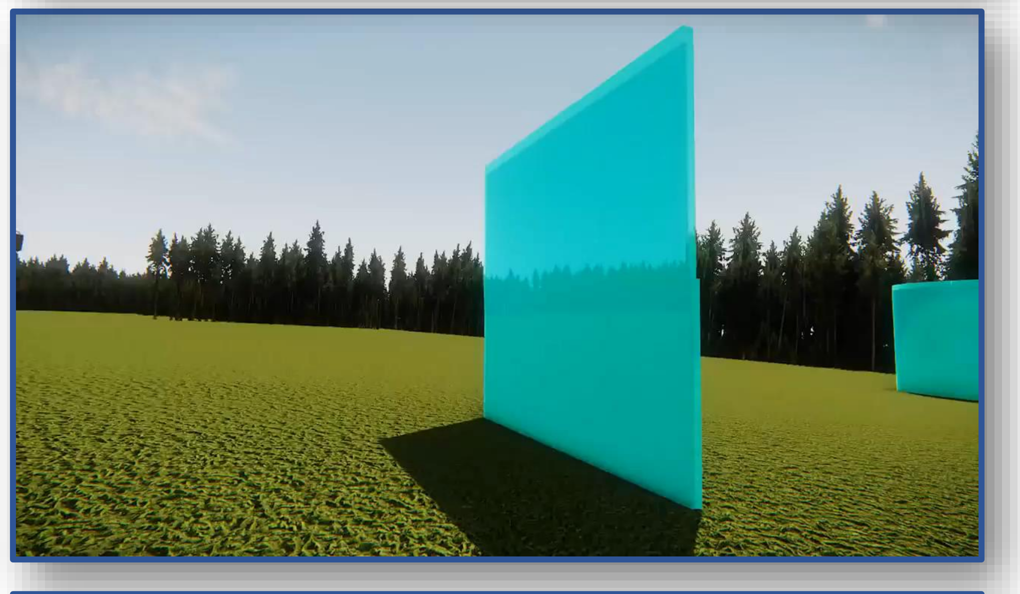


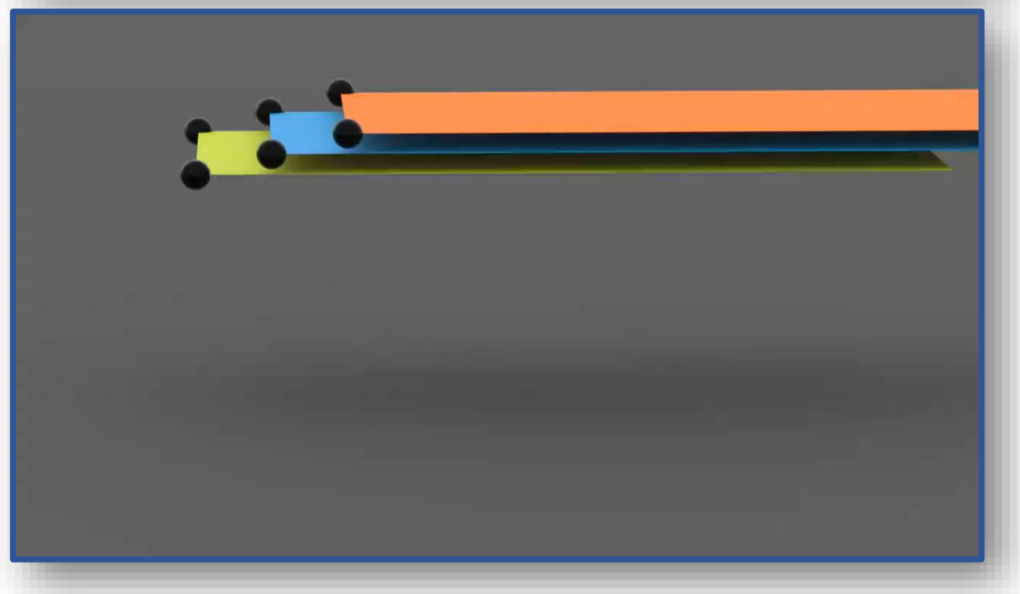
Skeletal Animation: *Keyframing*

Skeletal Animation: *Physics-based animation*





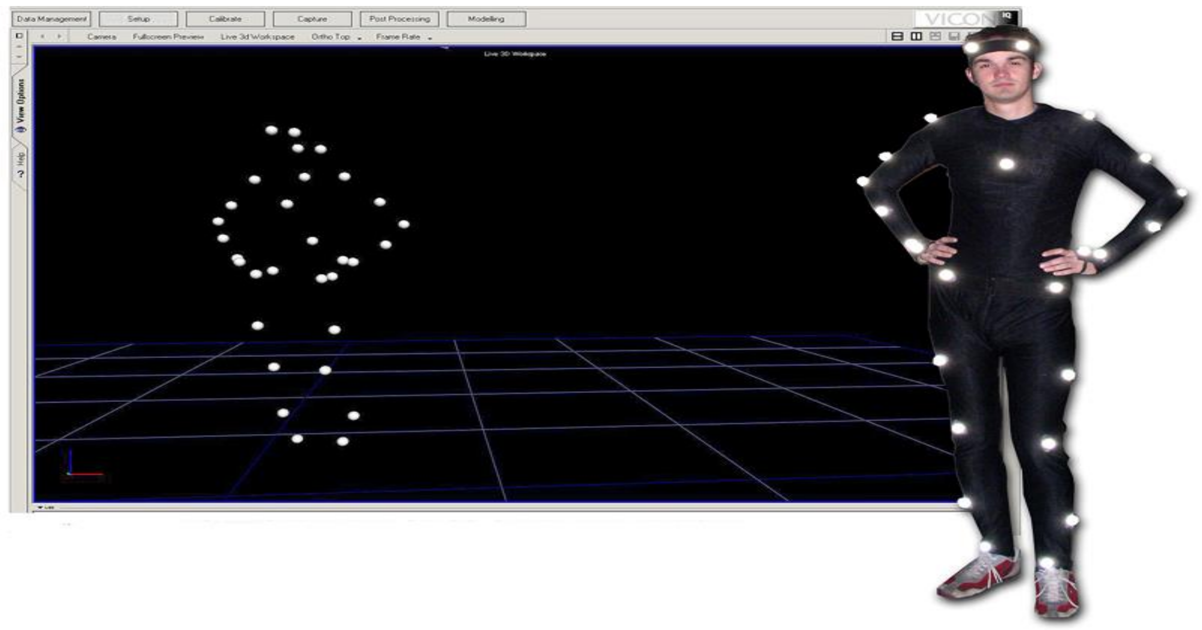


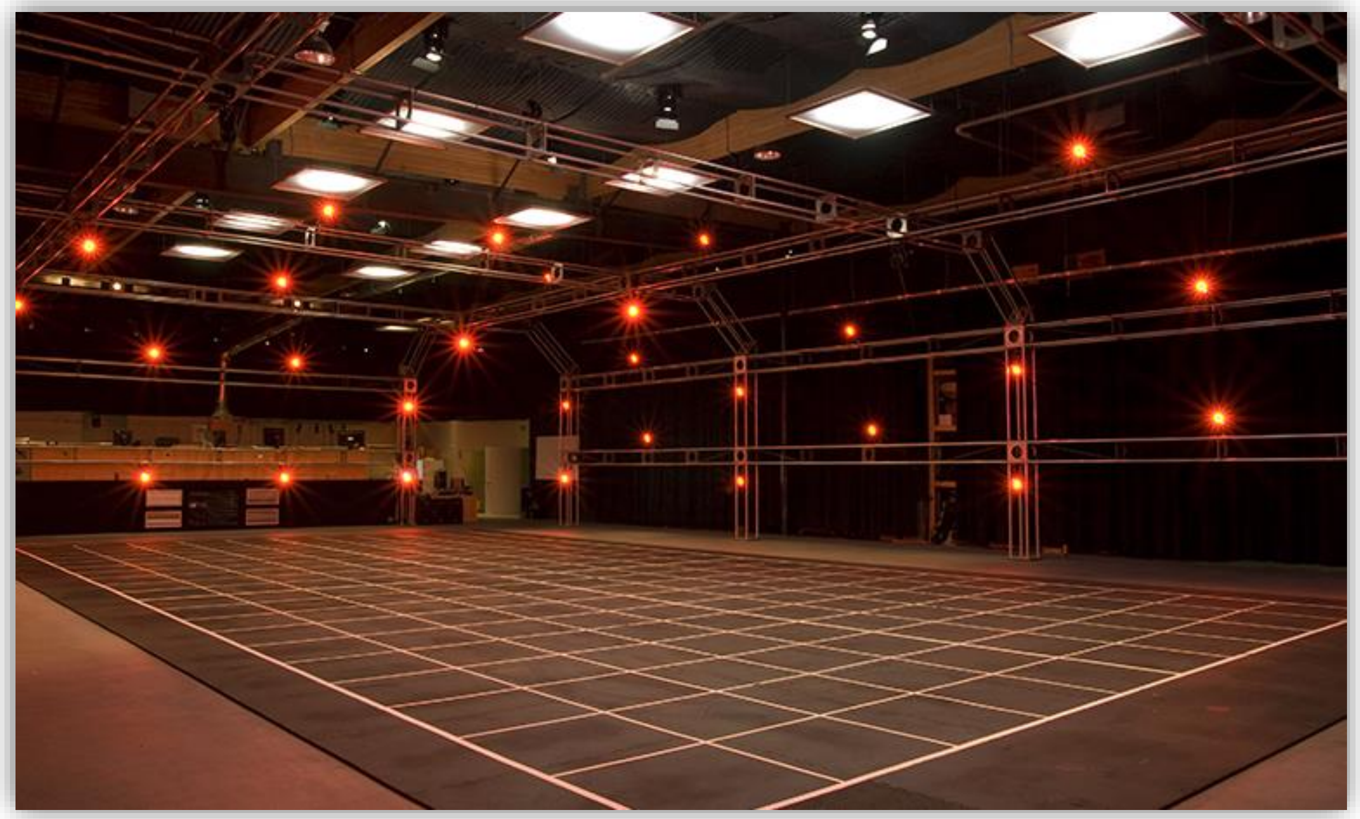


Computer generated animation: *Motion Capture*

Optical Motion Capture

- The system combines the information of the tracked markers to describe the 3D position of the object
  - Repeat this operation several times per second the system can provide us the volumetric trajectory of the marker according to time and space (usually from 30Hz to 960Hz)
- Great naturalness and realism in the captured movements.
  - High quality recording
  - Capturing of both main and secondary movements
  - Ease of use (Skeletal geometry is given)
- Capture volume is the physical space where the cameras can combine their fields of view





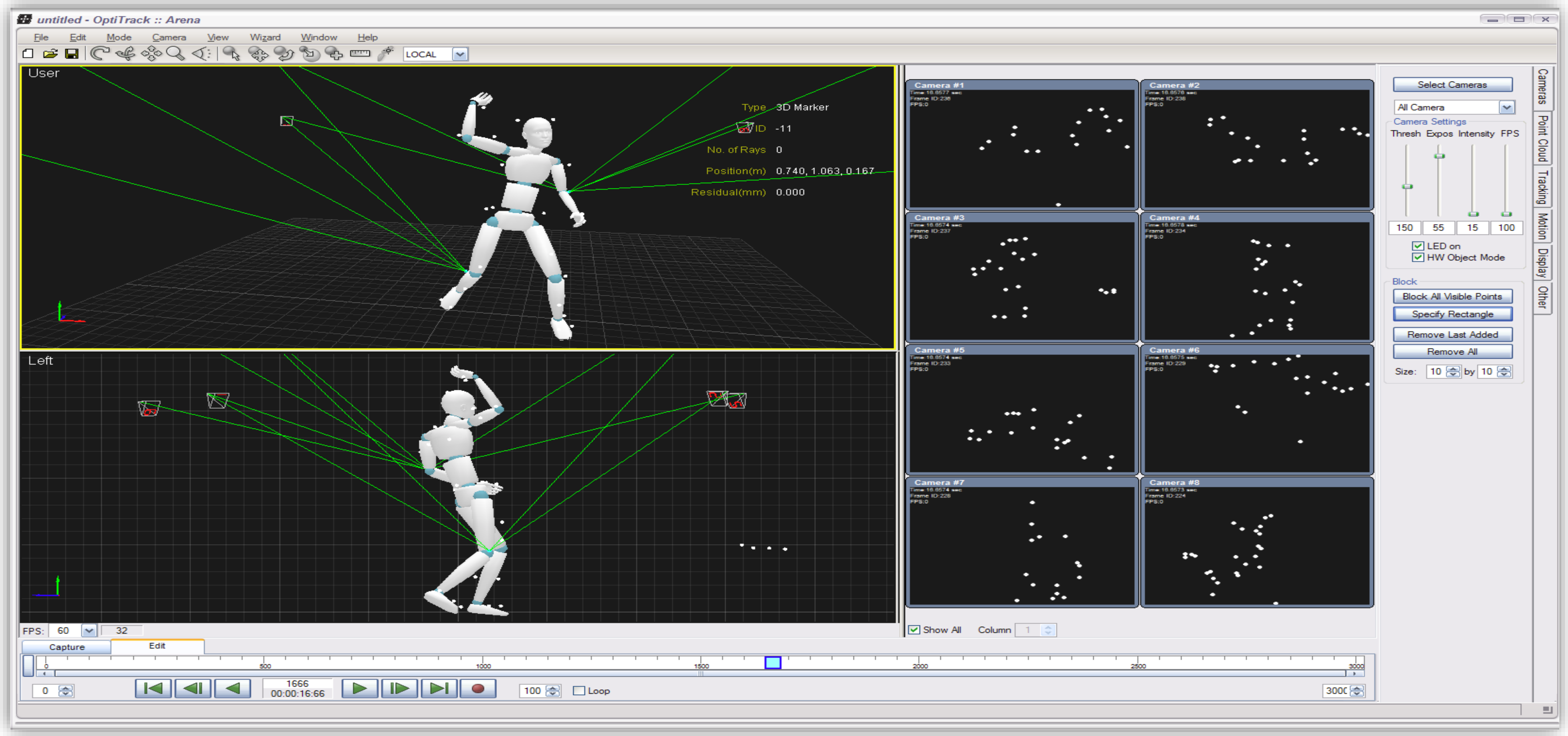
Co-financed by the European Union
Connecting Europe Facility

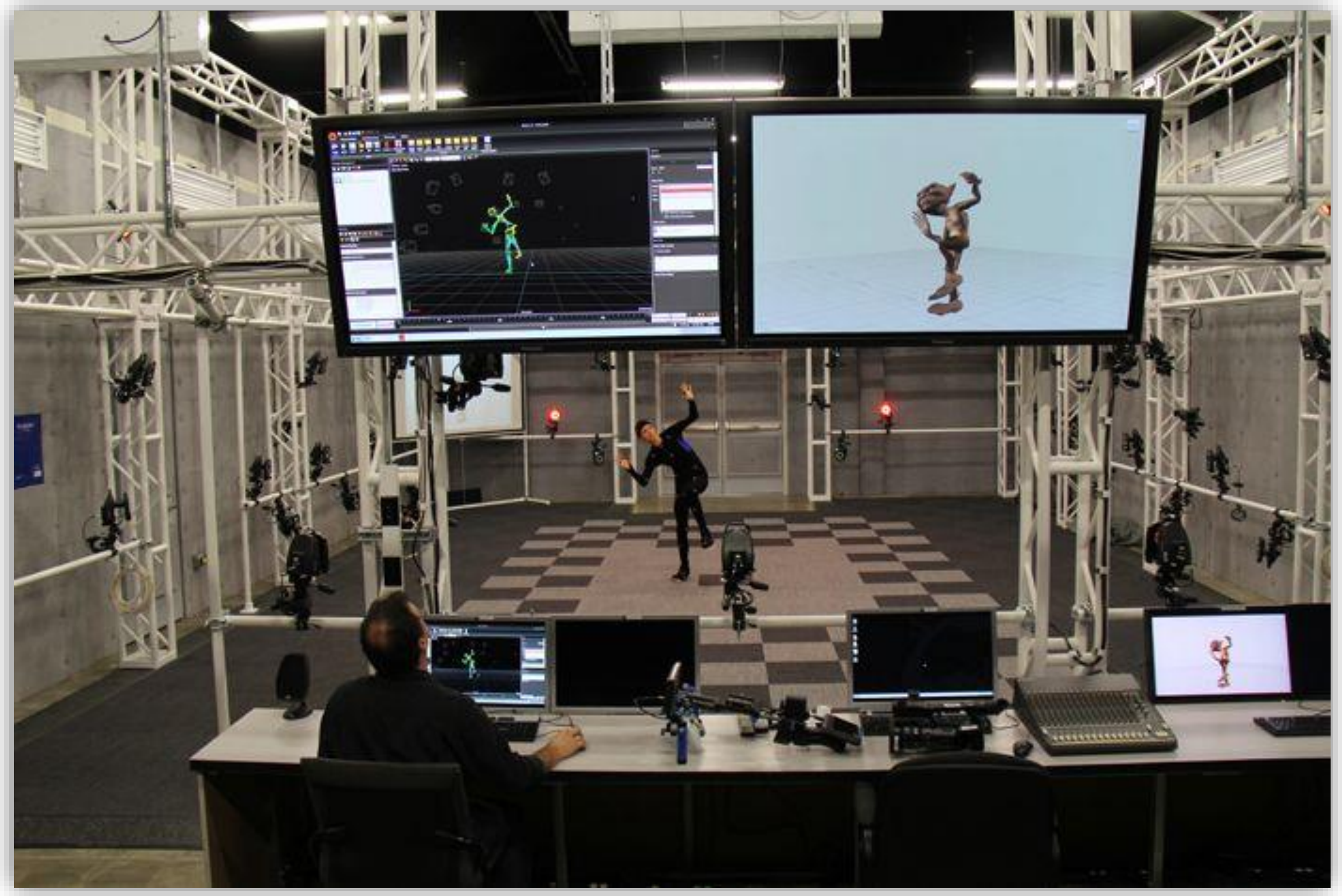
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

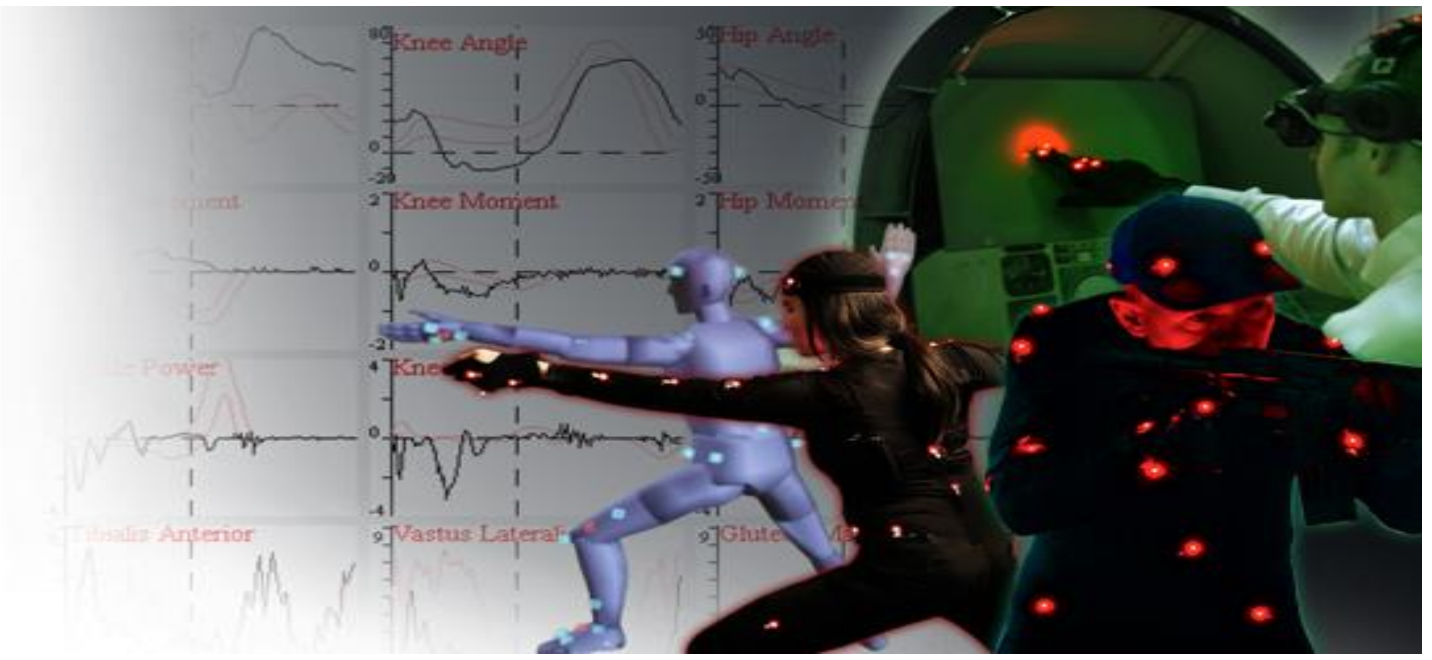
Computer generated animation: *Motion Capture*

Optical Motion Capture

- Each person wears a suit with markers attached.
- Enters a space that is surrounded with cameras.
- Divided into two main categories: passive and active



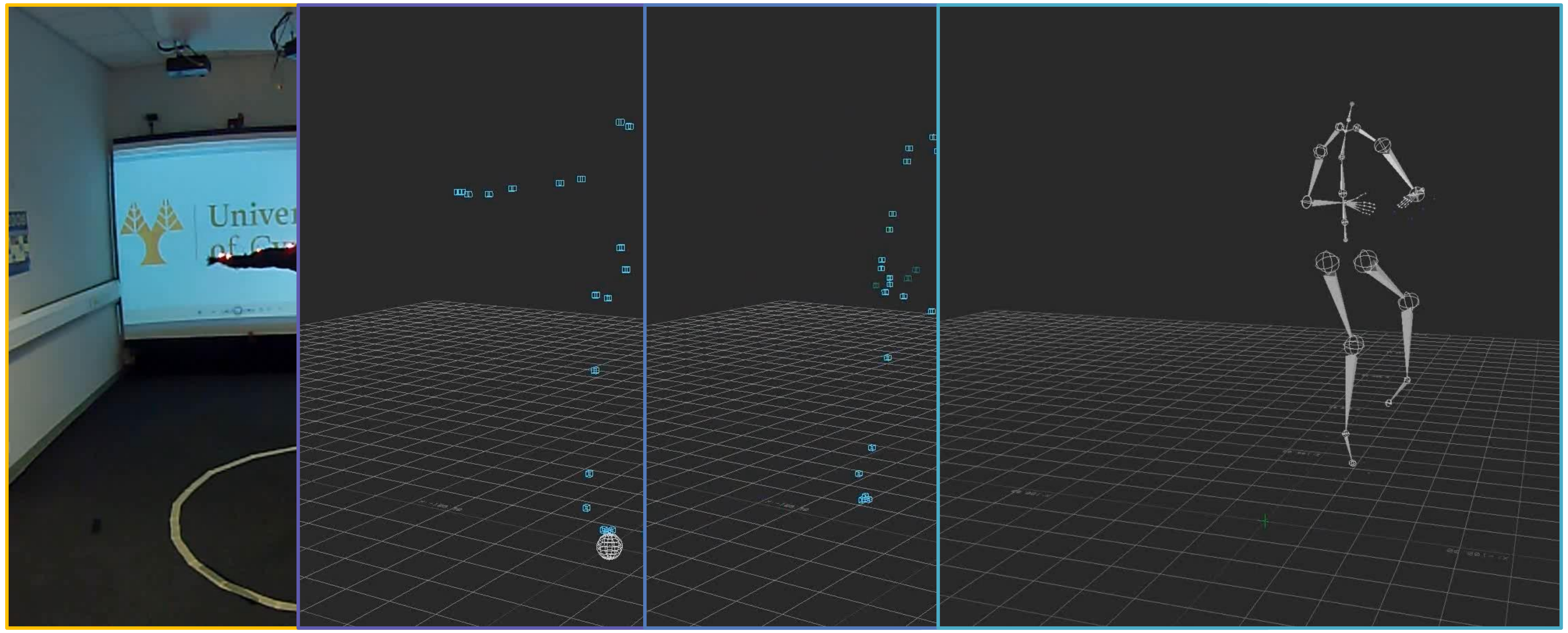




Motion Capture pipeline

Motion Capture pipeline

Marker-based motion acquisition → Label Markers → Marker Data Clean-up → Convert to Joint Angles



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Other popular motion capture systems

Inertial Markers

- Micro-inertial sensors, biomechanical models and sensor fusion algorithms.
- Use a number of gyroscopes and accelerometers to measure rotational rates.
- These rotations are translated to a skeleton model.

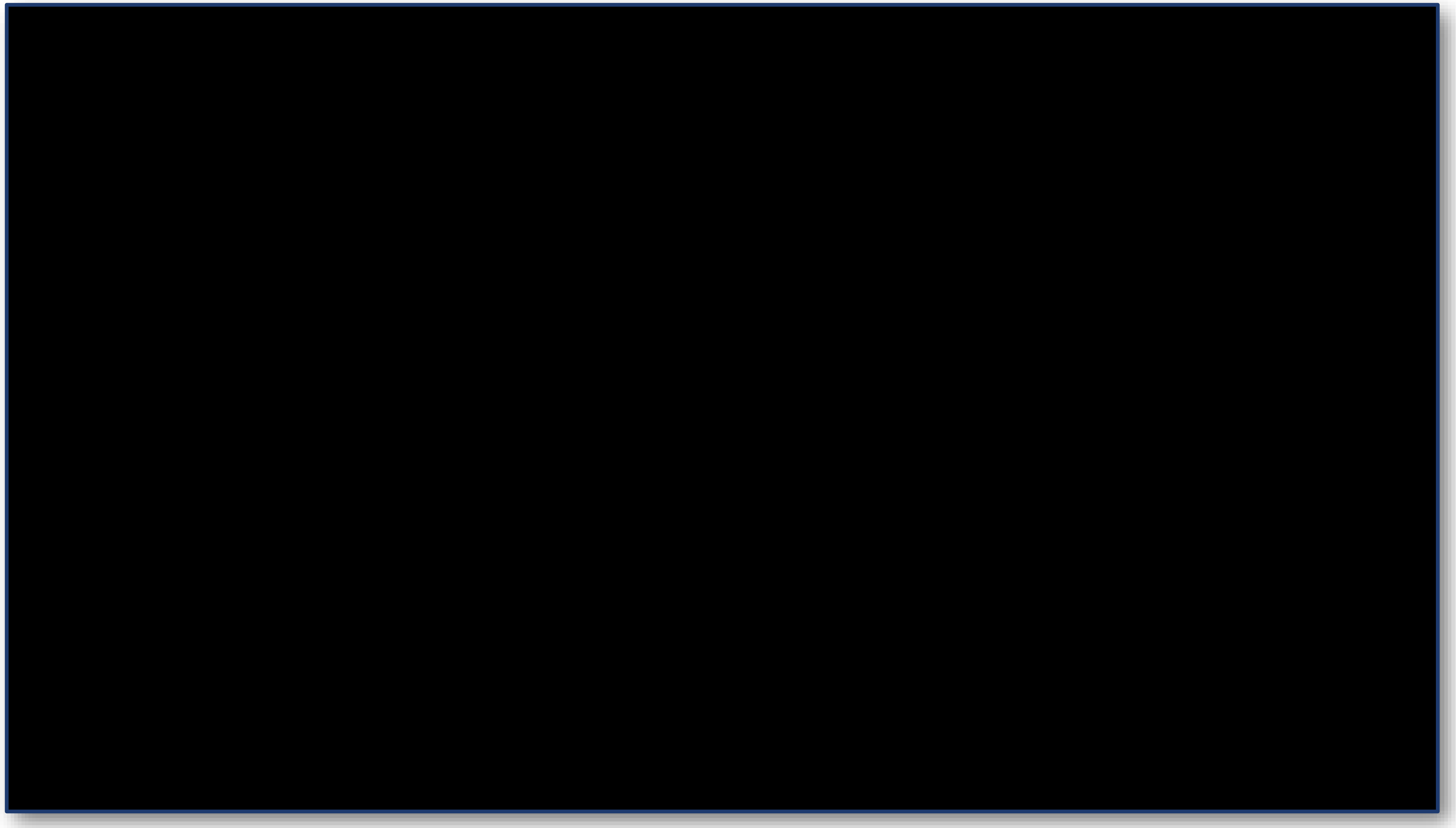
Depth-Based

- Use a combination of color cameras and depth sensors.
  - the subject's silhouette is captured from multiple angles.
- Reconstruct the object's volume (mesh) from the point clouds.
- Fit a skeleton into the 3D model to estimate motion.

Vision-Based

- Use a singe or multiple RGB cameras
- Mainly based on deep-learning methods
- Use large amount of training motion data

Motion Capture: *Current technological trends*



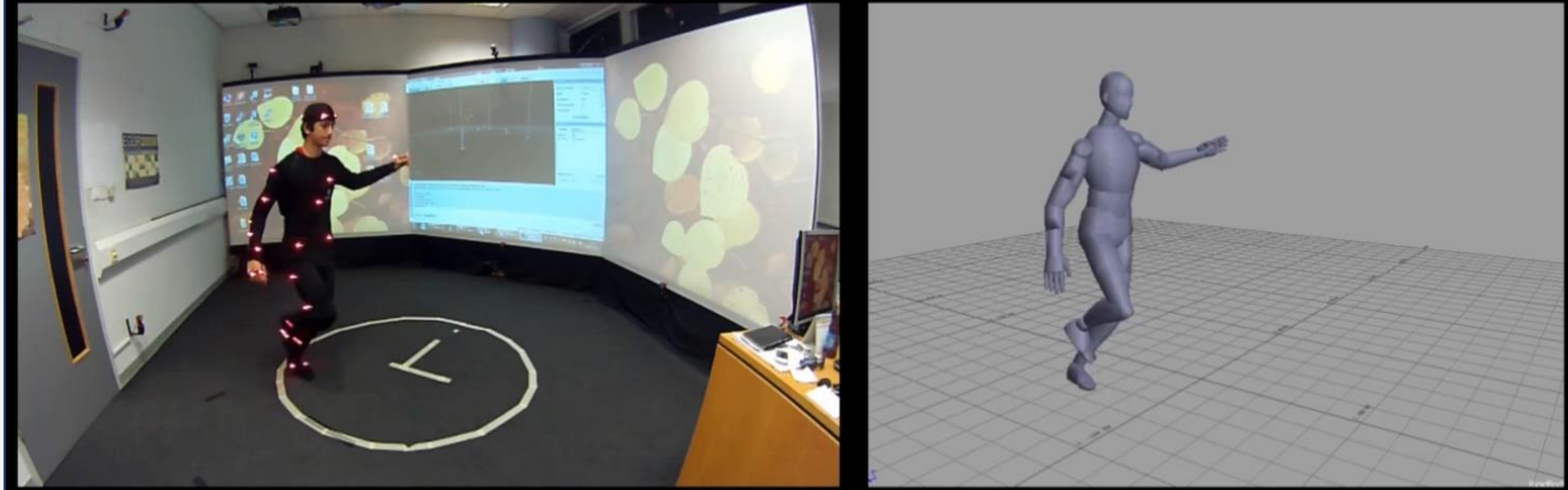
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

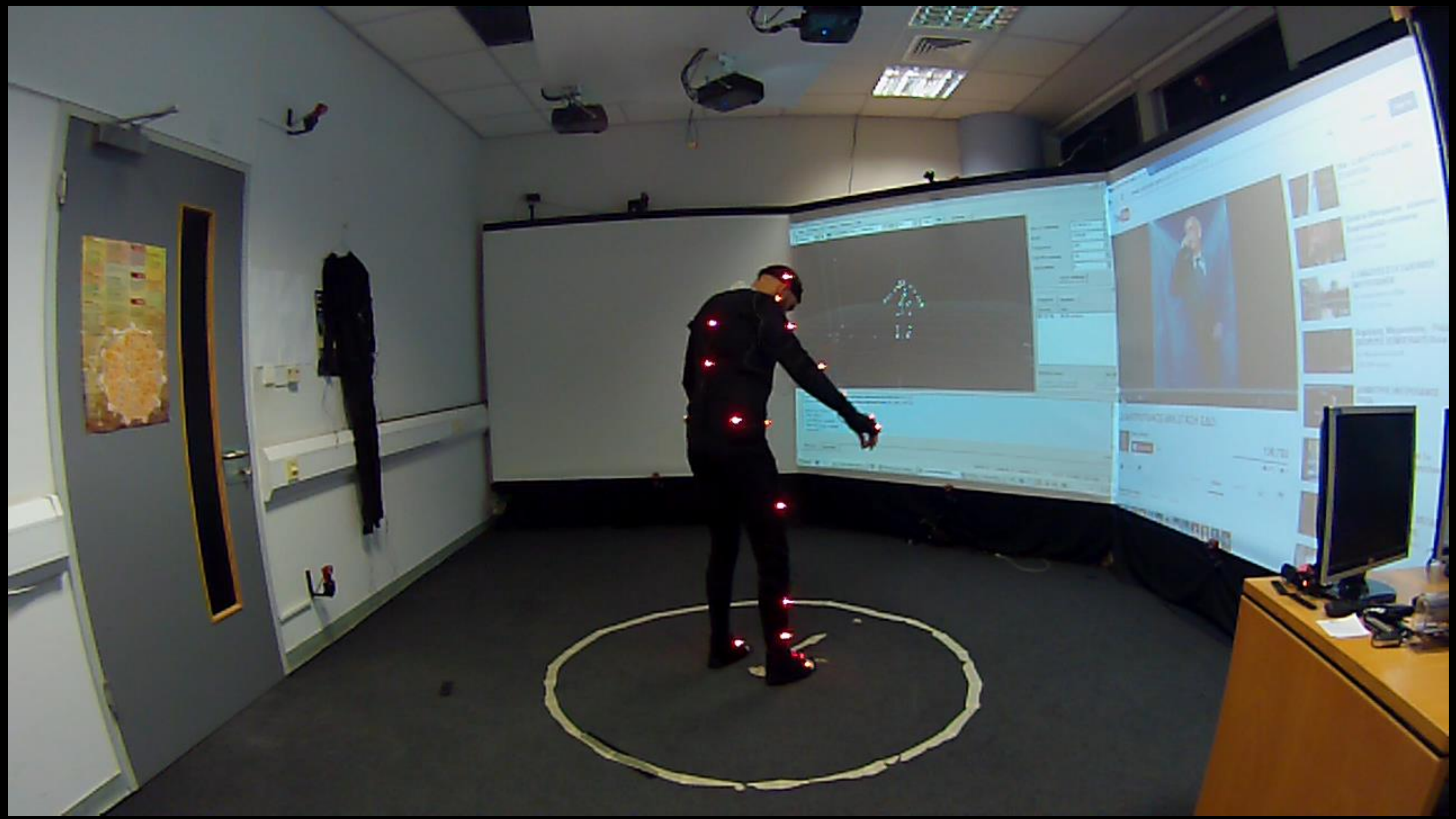
Why 3D?



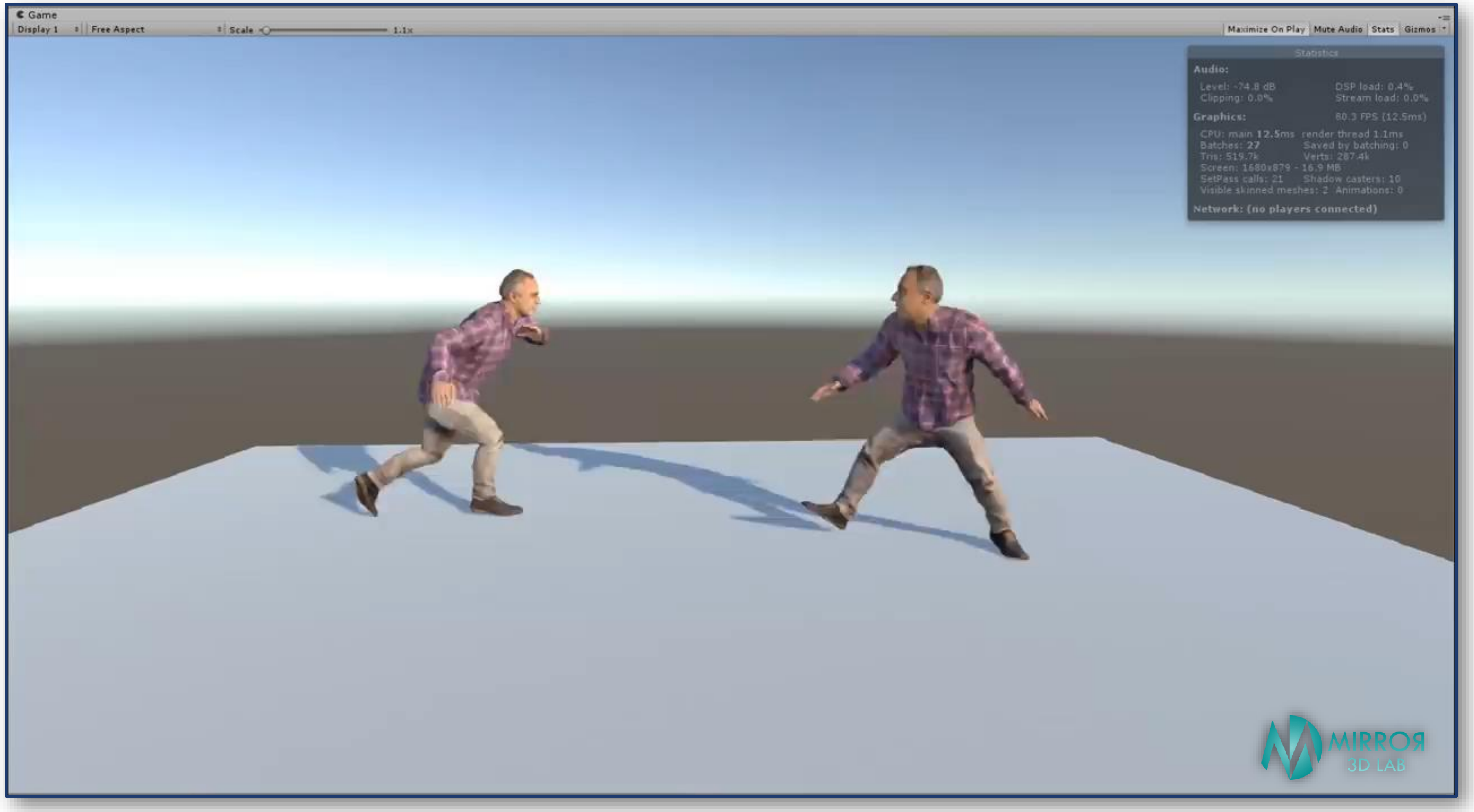
© Virtual Reality Lab, Department of Computer Science, University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

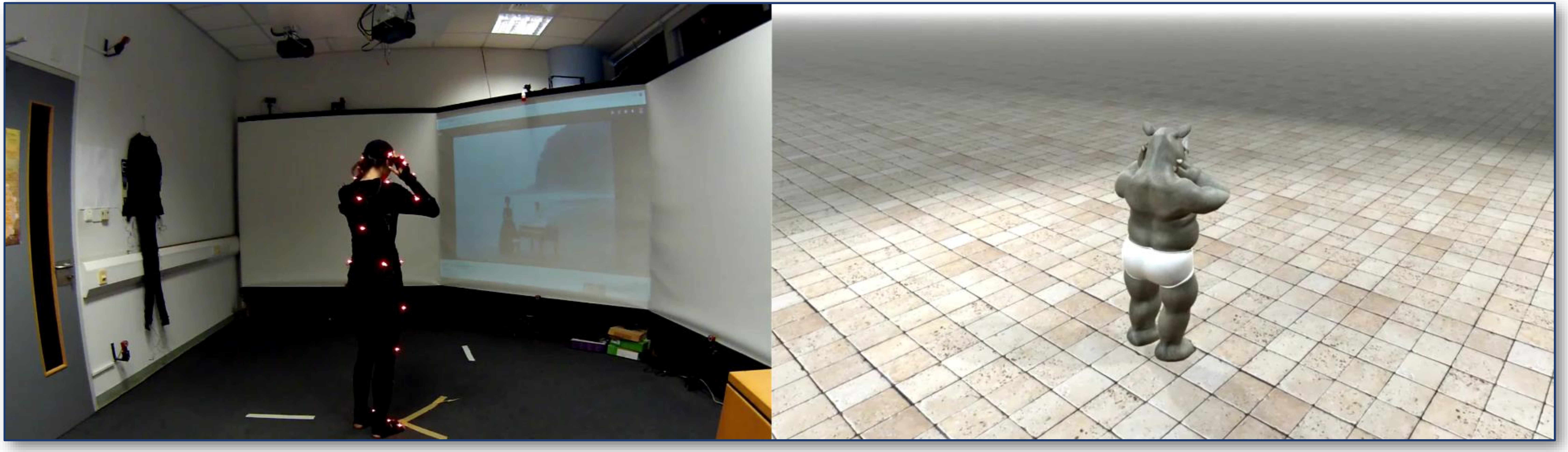
3D scanning by MIRROЯ 3D LAB



3D scanning and animation



More Examples



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

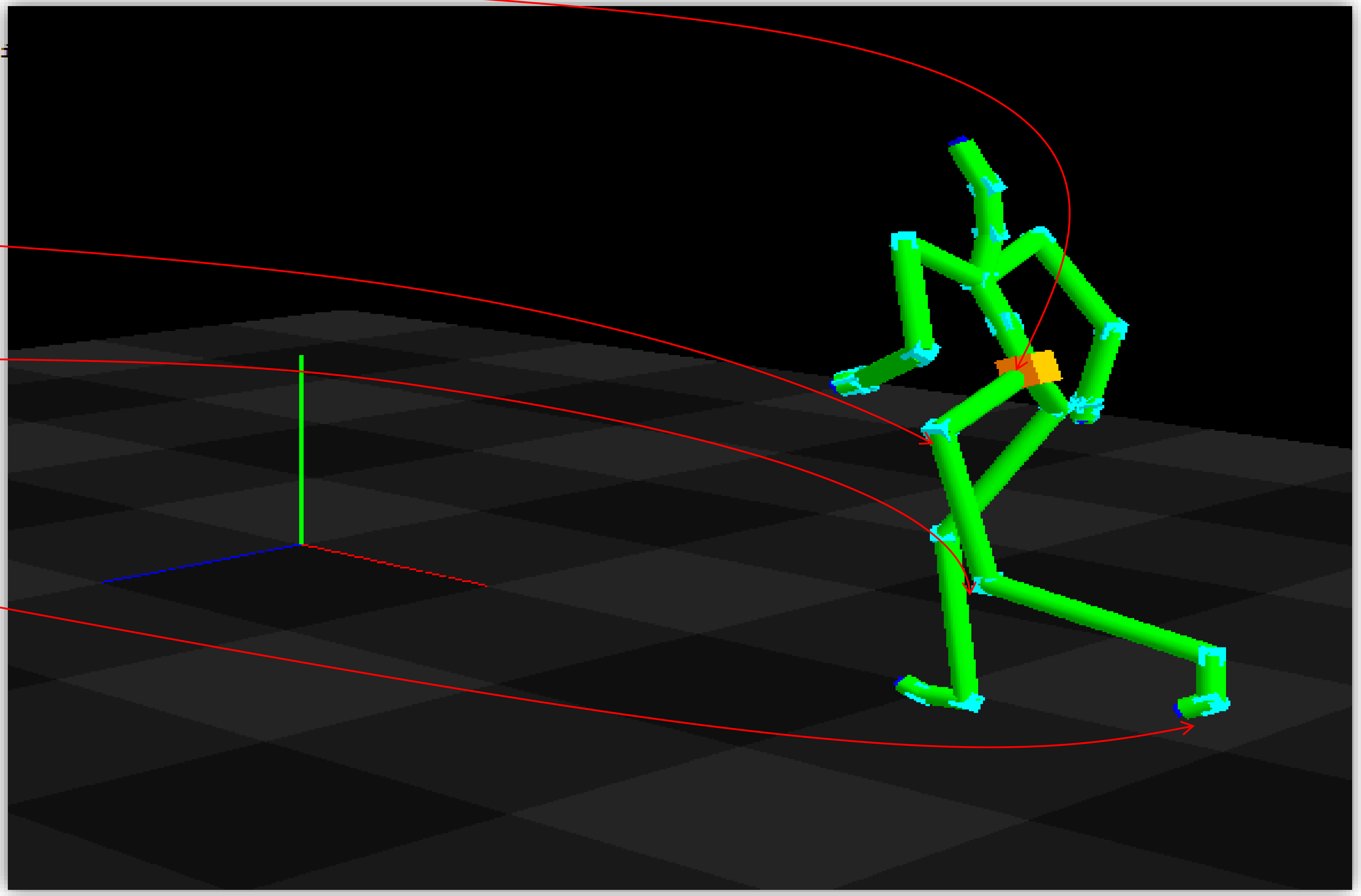
Motion Capture Data

- Depending on the sensors used
- Popular file formats:
  - ASF/AMC (Acclaim's skeleton and motion capture files)
  - BVH (BioVision Hierarchy)
  - C3D (Coordinate 3D – biomechanics – C3D.org)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Motion Capture Data: *BVH format*

```
HIERARCHY
ROOT Hips
{
    OFFSET 0.00000 0.00000 0.00000
    CHANNELS 6 Xposition Yposition Zposition Zrotation Yrotation Xrotati
    JOINT LHipJoint
    {
        OFFSET 0 0 0
        CHANNELS 3 Zrotation Yrotation Xrotation
        JOINT LeftUpLeg
        {
            OFFSET 3.13874 -1.57224 1.49786
            CHANNELS 3 Zrotation Yrotation Xrotation
            JOINT LeftLeg
            {
                OFFSET 2.10955 -5.79594 0.00000
                CHANNELS 3 Zrotation Yrotation Xrotation
                JOINT LeftFoot
                {
                    OFFSET 2.41843 -6.64458 0.00000
                    CHANNELS 3 Zrotation Yrotation Xrotation
                    JOINT LeftToeBase
                    {
                        OFFSET 0.04713 -0.12948 1.66229
                        CHANNELS 3 Zrotation Yrotation Xrotation
                        End Site
                        {
                            OFFSET 0.00000 -0.00000 0.85167
                        }
                    }
                }
            }
        }
    }
    JOINT RHipJoint
    {
        OFFSET 0 0 0
        CHANNELS 3 Zrotation Yrotation Xrotation
```



Motion Capture Data: *BVH format*

```
JOINT LeftUpLeg
{
    OFFSET 3.13874 -1.57224 1.49786
    CHANNELS 3 Zrotation Yrotation Xrotation
```

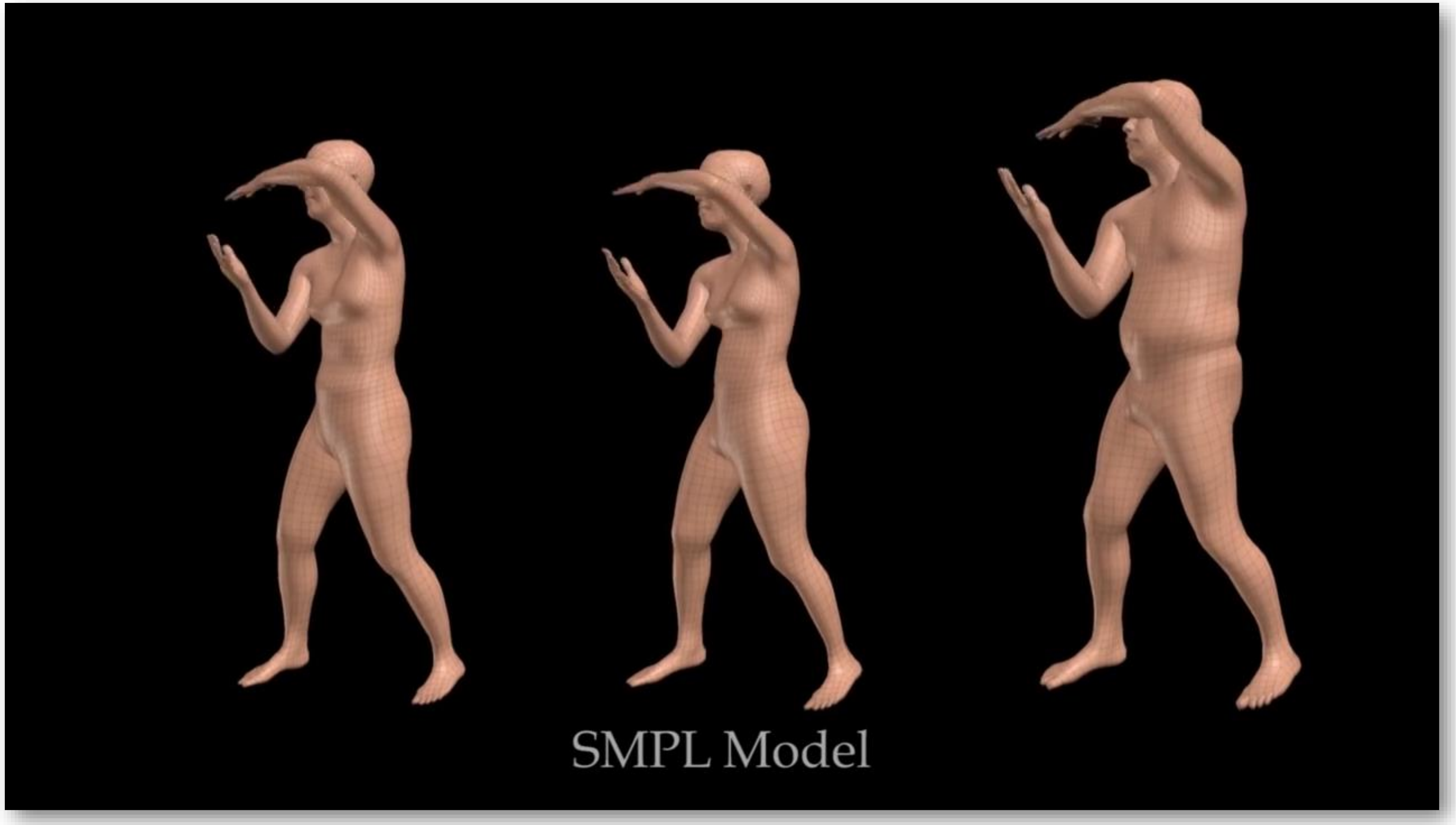
```
MOTION
Frames: 75
Frame Time: .0083333
34.0924 14.783 2.9468 0 0 0 0 0 0 -21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -16 0
34.0924 14.7830 2.9468 17.4229 -87.4657 5.1295 0.0000 0.0000 0.0000 -29.5745 -3.6613 -8.7583 11.0227 17.9462 62.8551 -1
33.3916 14.7552 2.8854 -58.8315 -85.2317 81.4832 0.0000 0.0000 0.0000 -23.5753 -2.3688 -17.8110 14.7006 19.3129 74.333
32.6541 14.7979 2.8438 -71.7590 -84.8325 95.1453 0.0000 0.0000 0.0000 -22.5698 -2.5804 -26.4803 17.9988 19.9037 84.139
31.8948 14.8911 2.7984 -75.6060 -85.3178 99.0013 0.0000 0.0000 0.0000 -23.0057 -5.8807 -33.9743 20.1042 19.9997 90.304
31.1232 14.9864 2.7457 -77.4701 -86.0438 100.8650 0.0000 0.0000 0.0000 -23.1084 -9.9967 -40.9151 21.3634 19.9554 93.98
30.3209 15.0896 2.6880 -79.5441 -86.1688 102.4380 0.0000 0.0000 0.0000 -21.9079 -13.9394 -46.7402 21.2632 19.9617 93.6
29.4932 15.2072 2.6225 -79.2415 -85.8899 101.5390 0.0000 0.0000 0.0000 -19.8826 -17.6174 -51.9266 20.7376 19.9869 92.1
28.6373 15.3433 2.5545 -78.1692 -85.6178 99.9907 0.0000 0.0000 0.0000 -17.8164 -20.7510 -56.5955 19.5517 19.9952 88.68
27.7548 15.4964 2.5055 -70.3414 -85.6867 91.8153 0.0000 0.0000 0.0000 -16.2872 -24.1174 -59.2862 17.5528 19.8558 82.82
26.8446 15.6692 2.4820 -53.5912 -85.7050 74.6419 0.0000 0.0000 0.0000 -15.4411 -27.5409 -59.6380 14.3642 19.2208 73.31
25.9208 15.8551 2.4558 -40.2095 -85.2517 60.8676 0.0000 0.0000 0.0000 -14.4638 -30.5887 -59.1844 11.1103 17.9886 63.13
24.9973 16.0414 2.4157 -33.1182 -84.5586 53.4984 0.0000 0.0000 0.0000 -13.6320 -32.9012 -56.4396 7.5387 15.7692 50.883
24.0737 16.2165 2.3701 -28.7693 -83.7550 48.9431 0.0000 0.0000 0.0000 -12.2060 -35.4136 -53.1281 4.3759 12.6445 38.051
23.1504 16.3653 2.3269 -27.5208 -82.8715 47.6202 0.0000 0.0000 0.0000 -10.6818 -37.6447 -49.7847 2.3540 9.5675 27.5871
22.2301 16.4781 2.2937 -30.5492 -81.9083 50.5828 0.0000 0.0000 0.0000 -8.0945 -39.8894 -46.9860 1.2734 7.1519 20.1663 1
21.3241 16.5402 2.2649 -36.9704 -80.9346 56.90
20.4356 16.5464 2.2313 -45.5334 -79.6014 65.34
19.5406 16.4651 2.1998 -54.0566 -79.7785 73.03
18.6511 16.3197 2.1717 -66.6258 -79.5556 84.58
17.7787 16.1118 2.1800 -65.2687 -78.8534 82.201
16.9761 15.8614 2.1727 -68.5351 -80.2276 84.0458 0.0000 0.0000 0.0000 -14.8237 -28.2761 -30.8201 1.5729 7.9131 22.4504
16.1768 15.6329 2.1620 -70.6336 -81.7219 84.3520 0.0000 0.0000 0.0000 -18.6407 -28.2134 -28.0709 2.4239 9.6982 28.0049
15.4307 15.5150 2.1499 -2.2680 -87.7380 16.3926 0.0000 0.0000 0.0000 -29.9127 -28.7976 -24.5424 2.8264 10.4085 30.3101
14.6696 15.4962 2.1445 58.2182 -85.7858 -44.5680 0.0000 0.0000 0.0000 -36.1801 -25.4197 -19.5626 2.5435 9.9166 28.7070
13.9284 15.5314 2.1254 52.6266 -86.4840 -38.5054 0.0000 0.0000 0.0000 -35.7337 -22.3260 -17.1207 2.2923 9.4500 27.2135
```

```
ROOT Hips
{
    OFFSET 0.00000 0.00000 0.00000
    CHANNELS 6 Xposition Yposition Zposition Zrotation Yrotation Xrotation
```


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Motion Capture Data: *SMPL format*



Motion Capture: *Advantages*

- Great naturalness and realism in the movements that have been recorded.
  - High-quality recording
  - Recording of both primary and secondary movements
- Recording at a very high frequency
  - Up to 980 samples per second (e.g., birds)
- Ease of use
  - Geometry is a given.
  - Freedom of movement for users

Motion Capture: *Limitations*

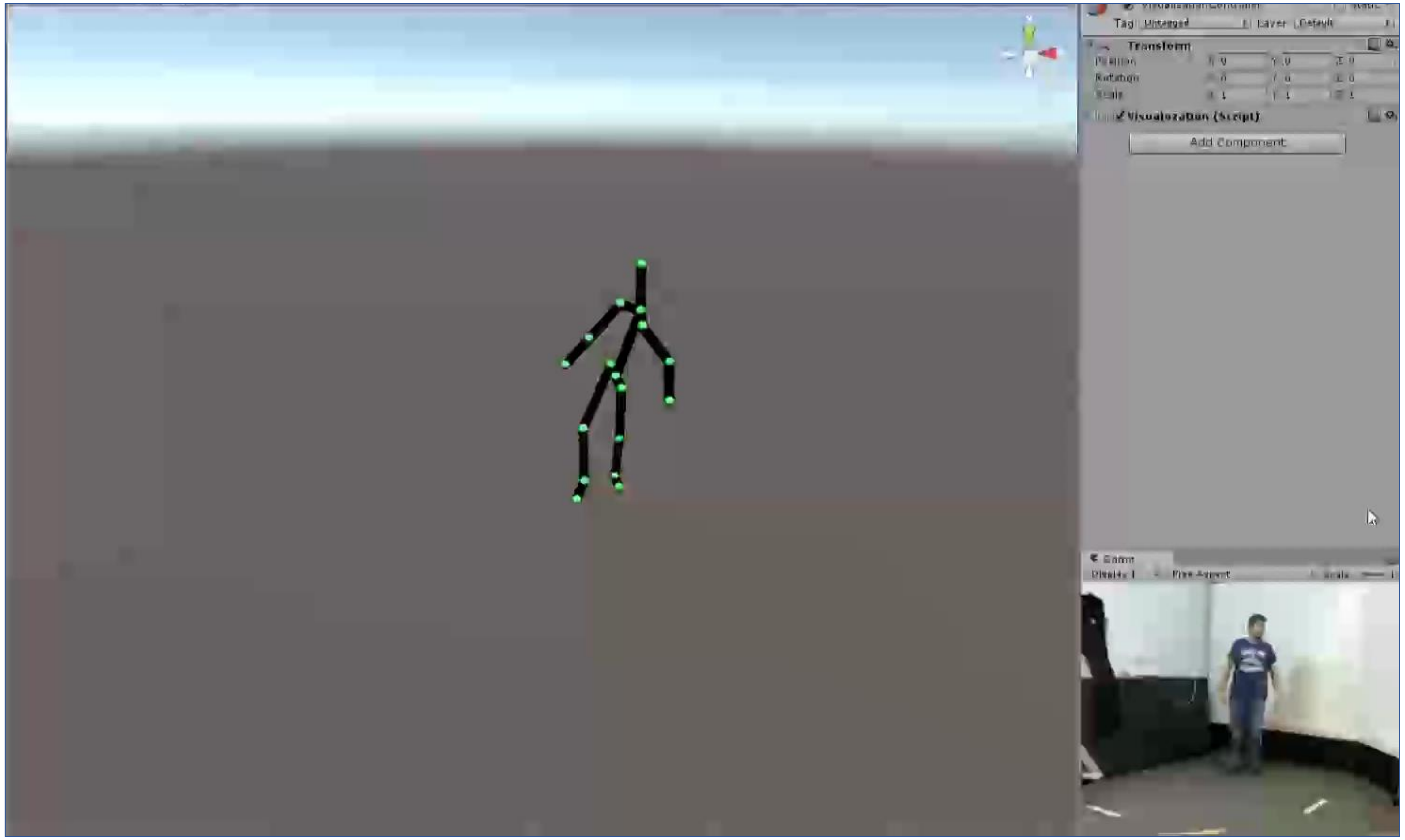
- Only realistic motion captured (movement that does not follow the laws of physics cannot be captured).
  - Cartoony or superhero animations are not possible to be captured.
- WYSIWYG (what you see is what you get).
  - Can't add more expression.
  - Continually need to recapture motion.
- What about muscles?

Animators could use more than 750 controls to create Shrek's performance. Some controlled one joint or muscle, others controlled groups of several.

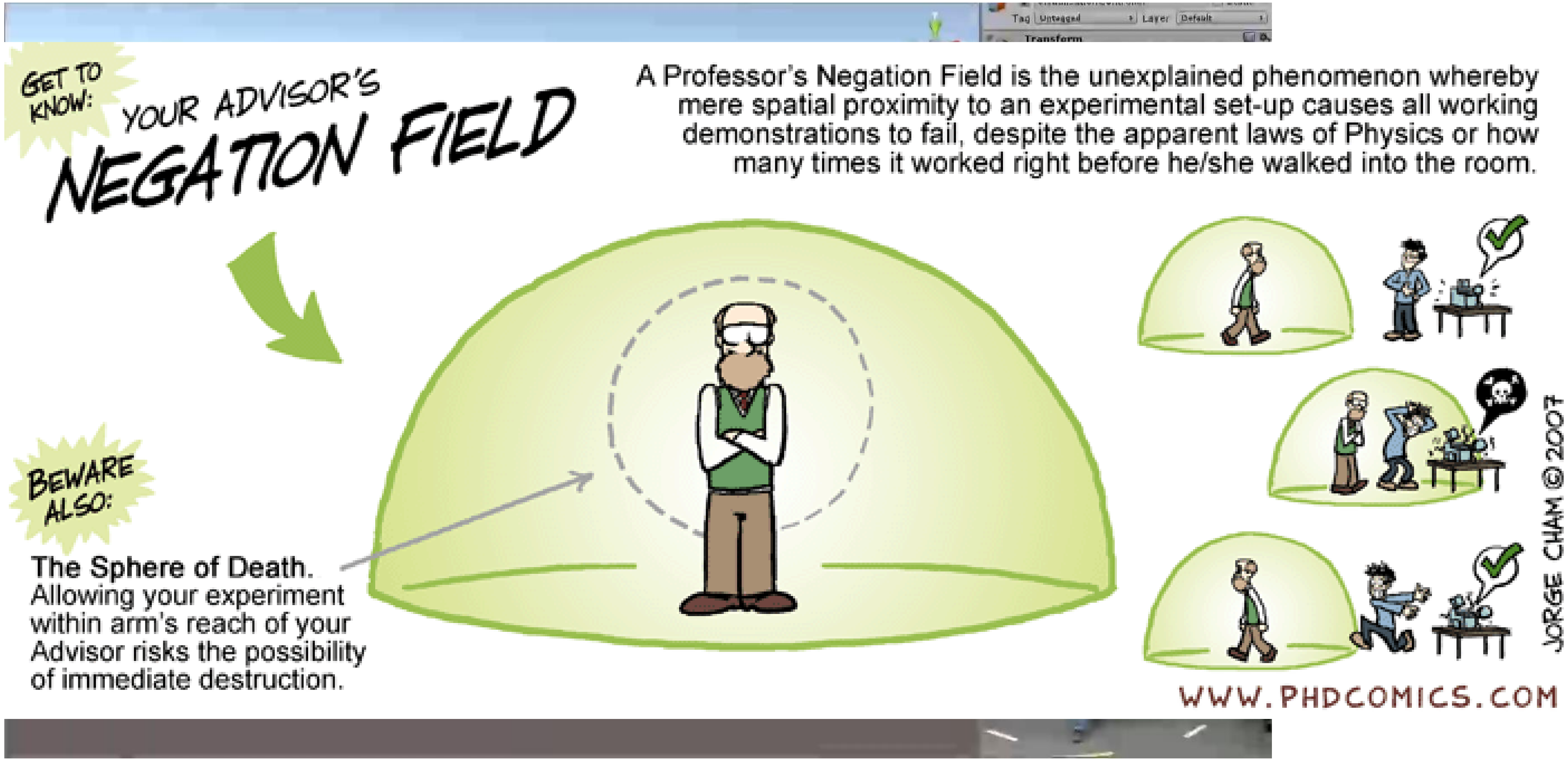
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

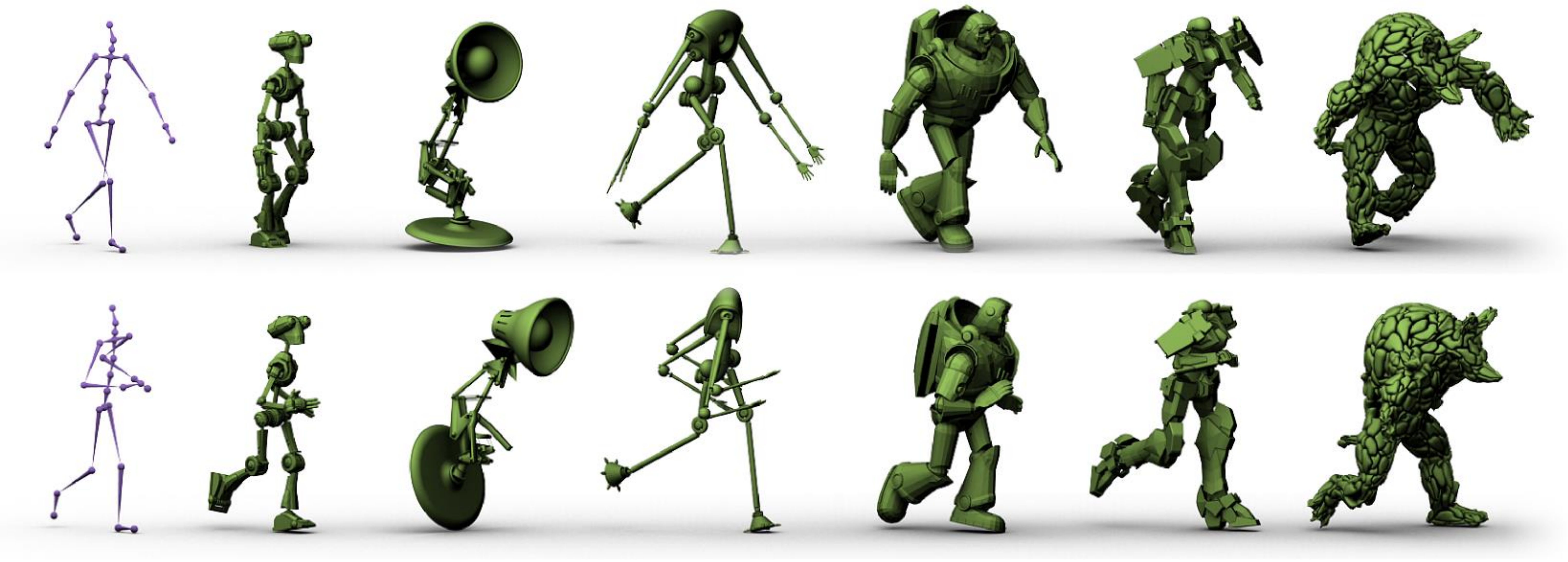
Some home-built motion capture systems



Some home-built motion capture systems



Other Challenges: *Motion Retargeting*



Other Challenges: *Motion Retargeting*

What is motion retargeting?

- A method to retarget animations onto models with different morphologies.
- A way to remap animations onto characters with very different animation-specific structures.

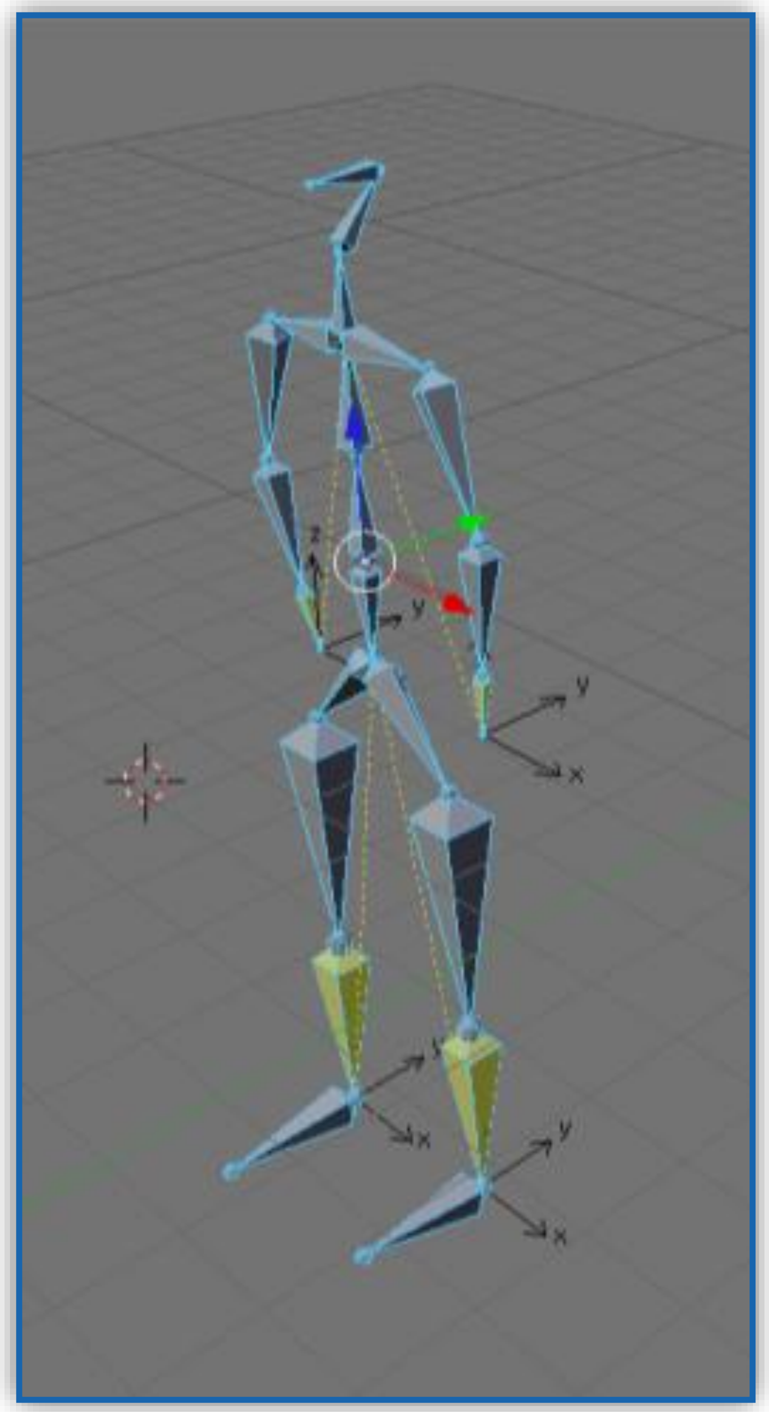
Co-financed by the European Union
Connecting Europe Facility

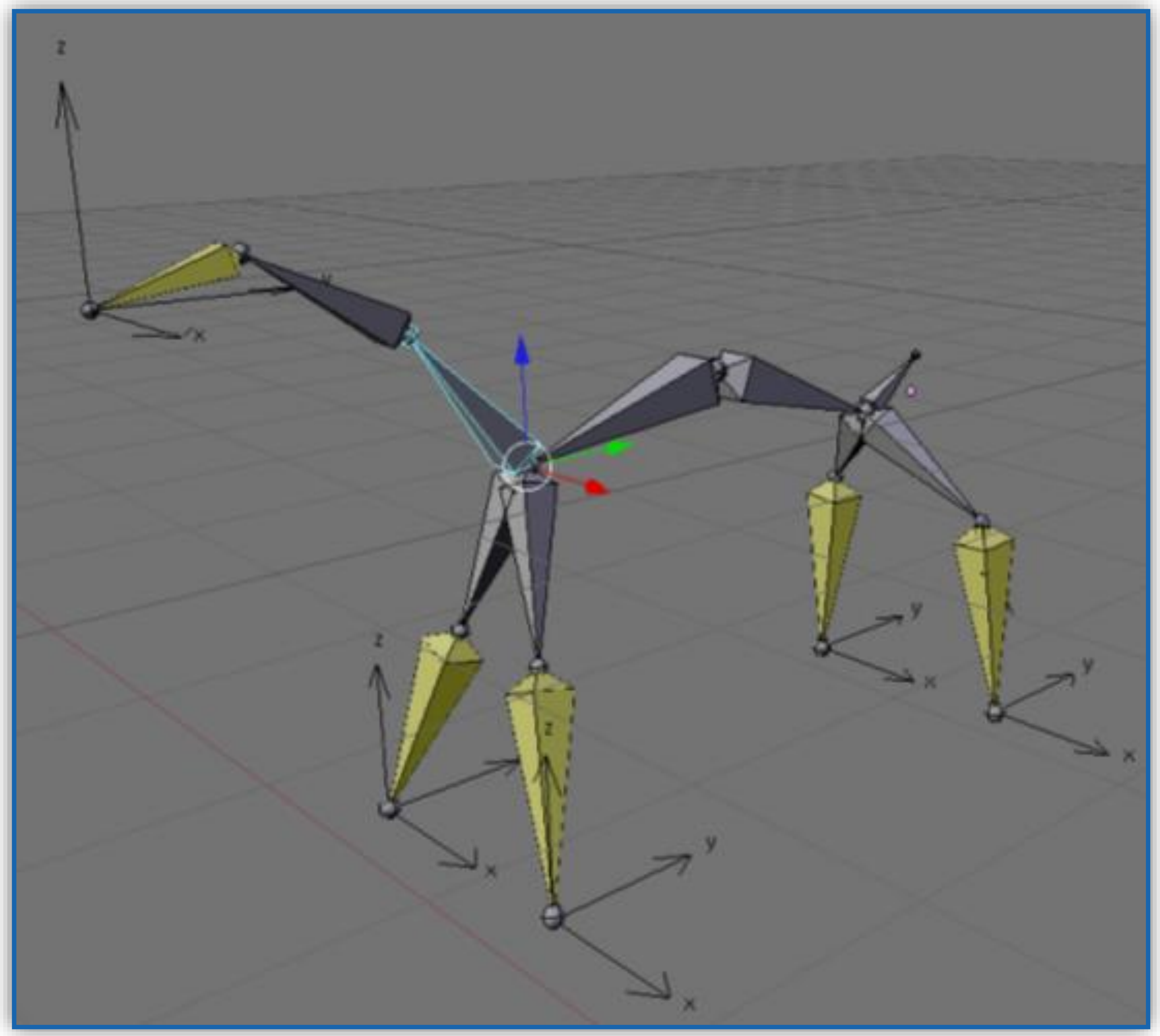
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Other Challenges: *Motion Retargeting*

Why Motion Retargeting?

- Improves content reuse.
- Easy integration of procedurally generated animations.
- Sometimes is not possible to motion capture the subject (e.g. animal with human behavior, character does not exist – fiction movies).



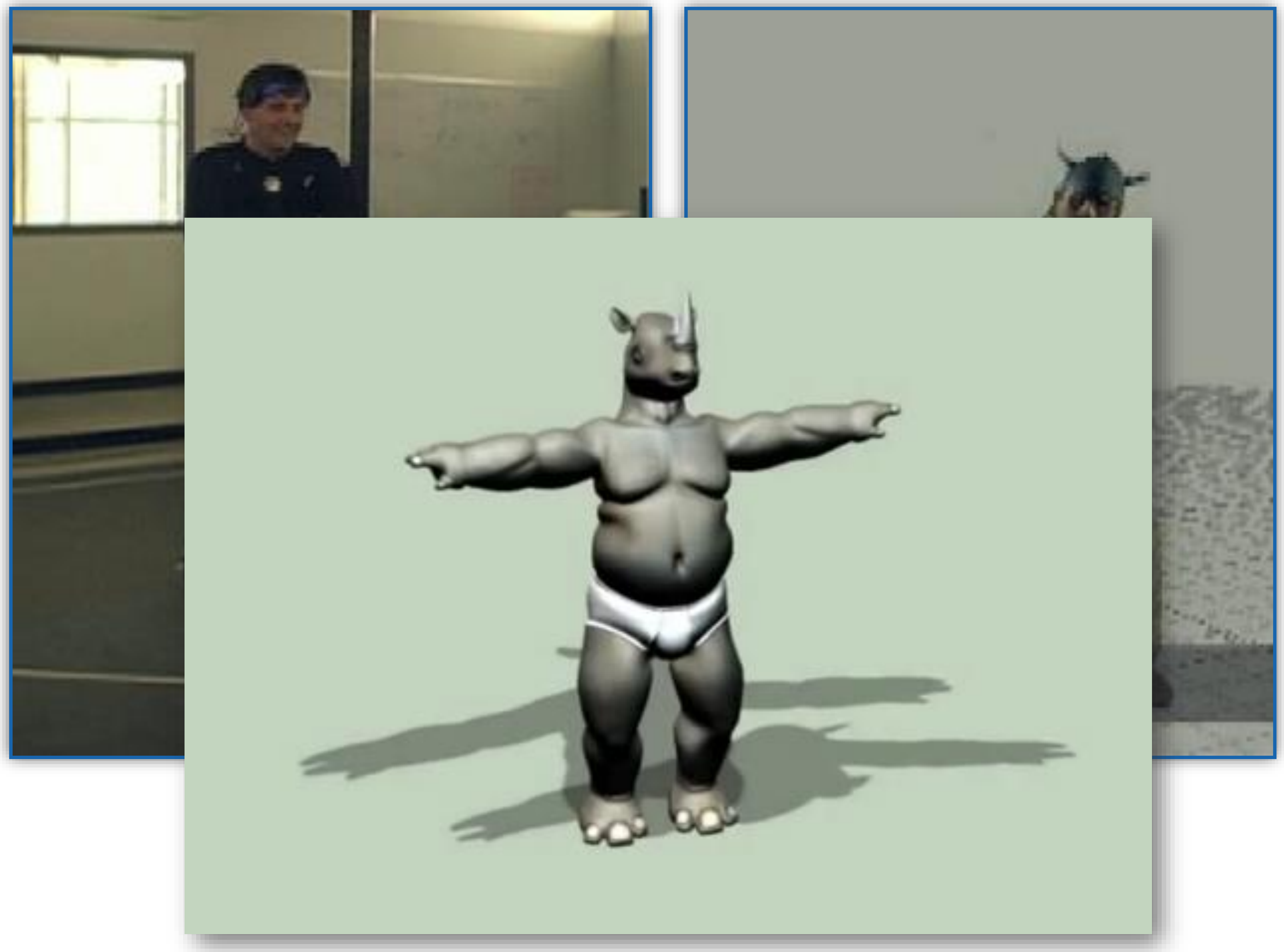


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Other Challenges: *Motion Retargeting*

- Preserve angles or end-effector positions (flying).
- Foot-skating.
- Characters with different proportions may have body penetration.



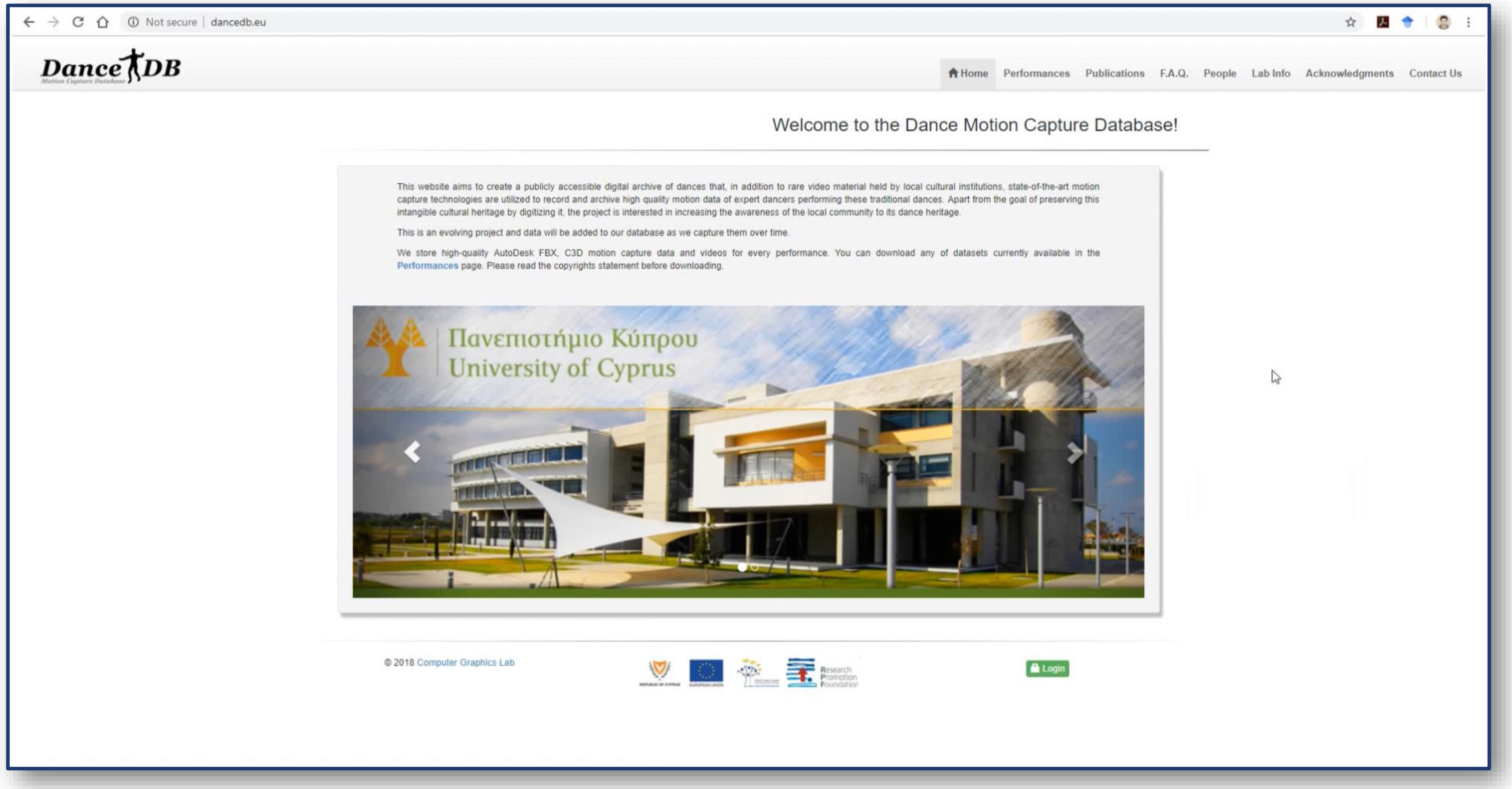
Co-financed by the European Union
Connecting Europe Facility

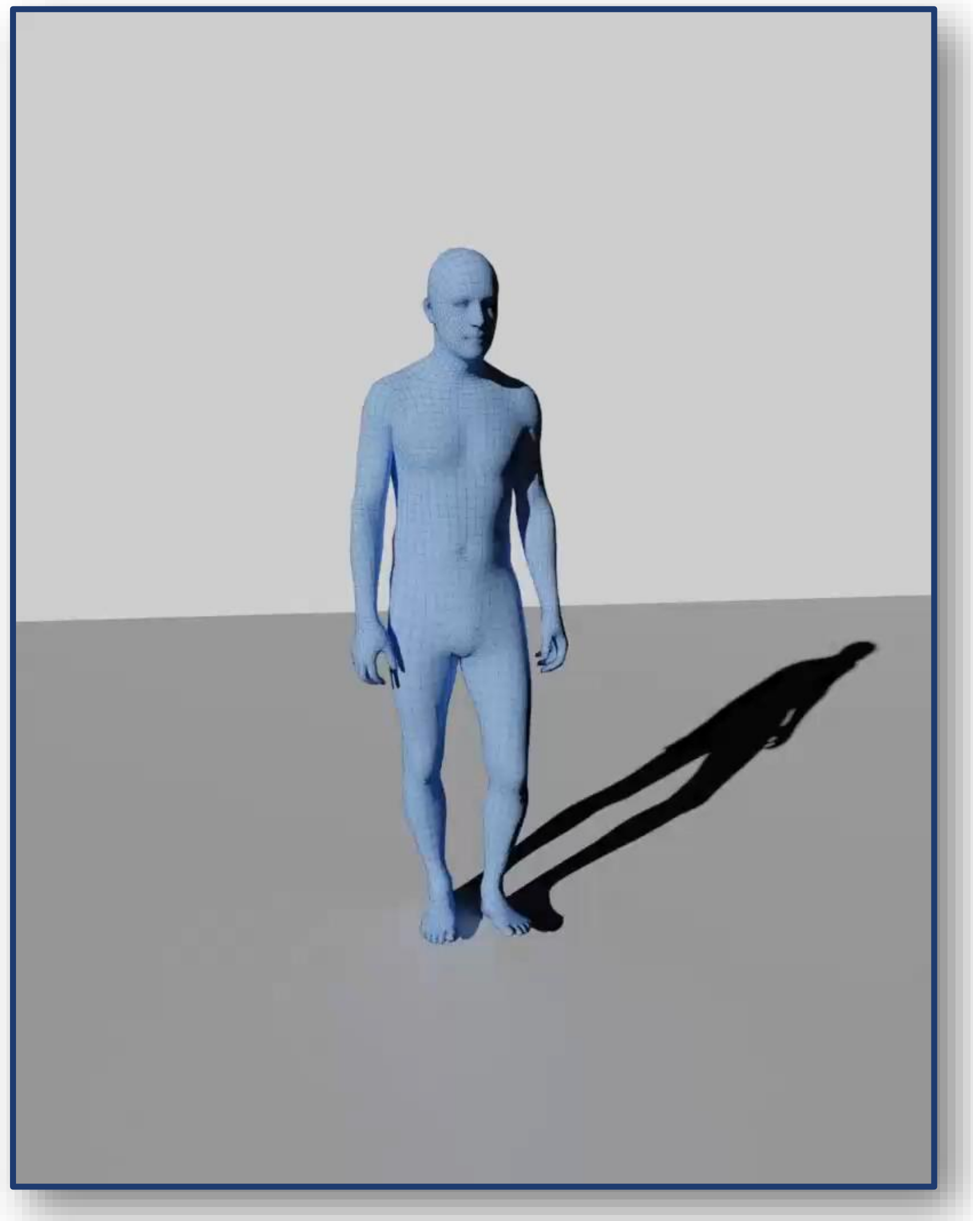
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Alita: Battle Angel, **© WETA & 20th Century FOX**

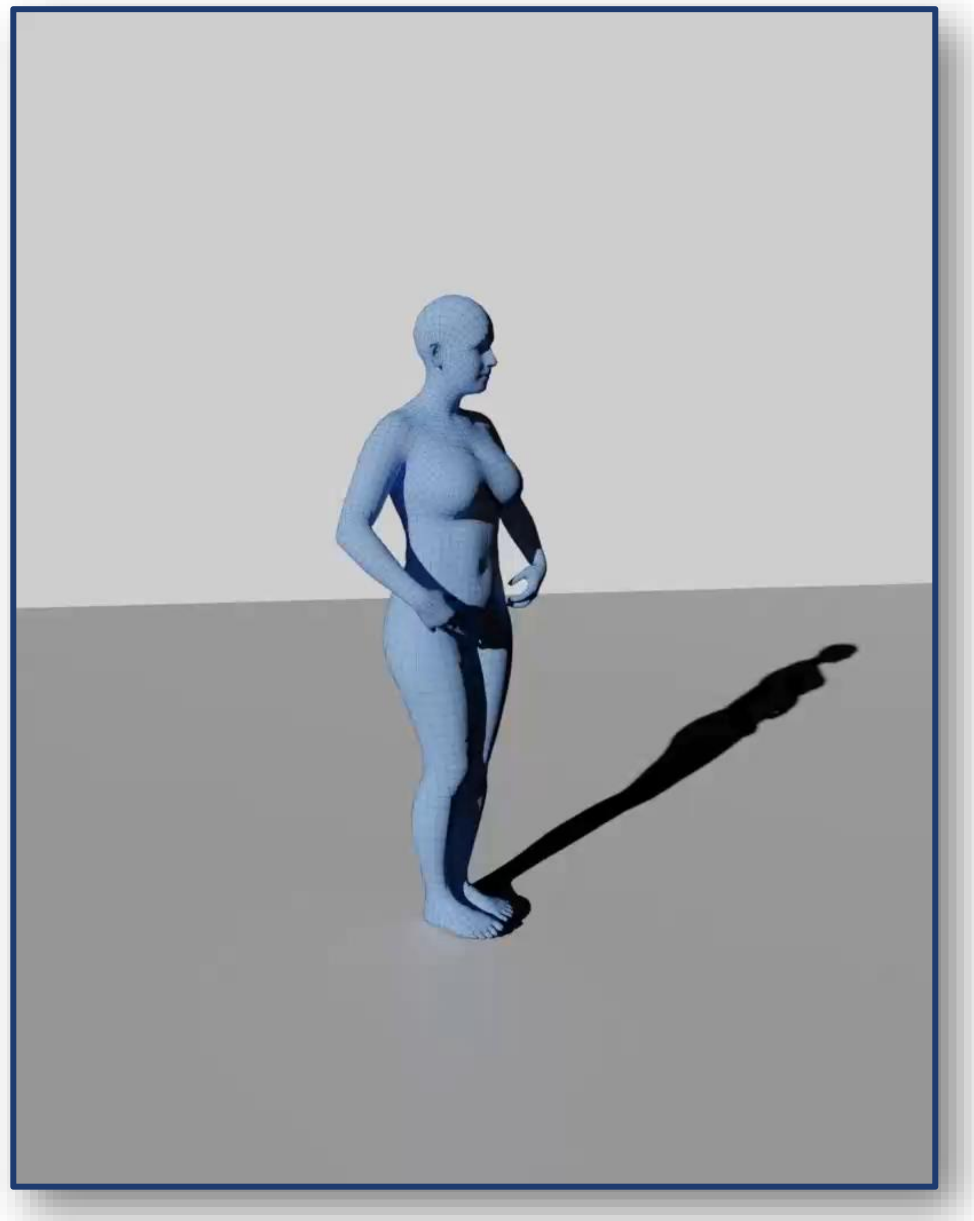
The Lion King, © Disney

Our Dance Motion Capture Database





1st Antikristos

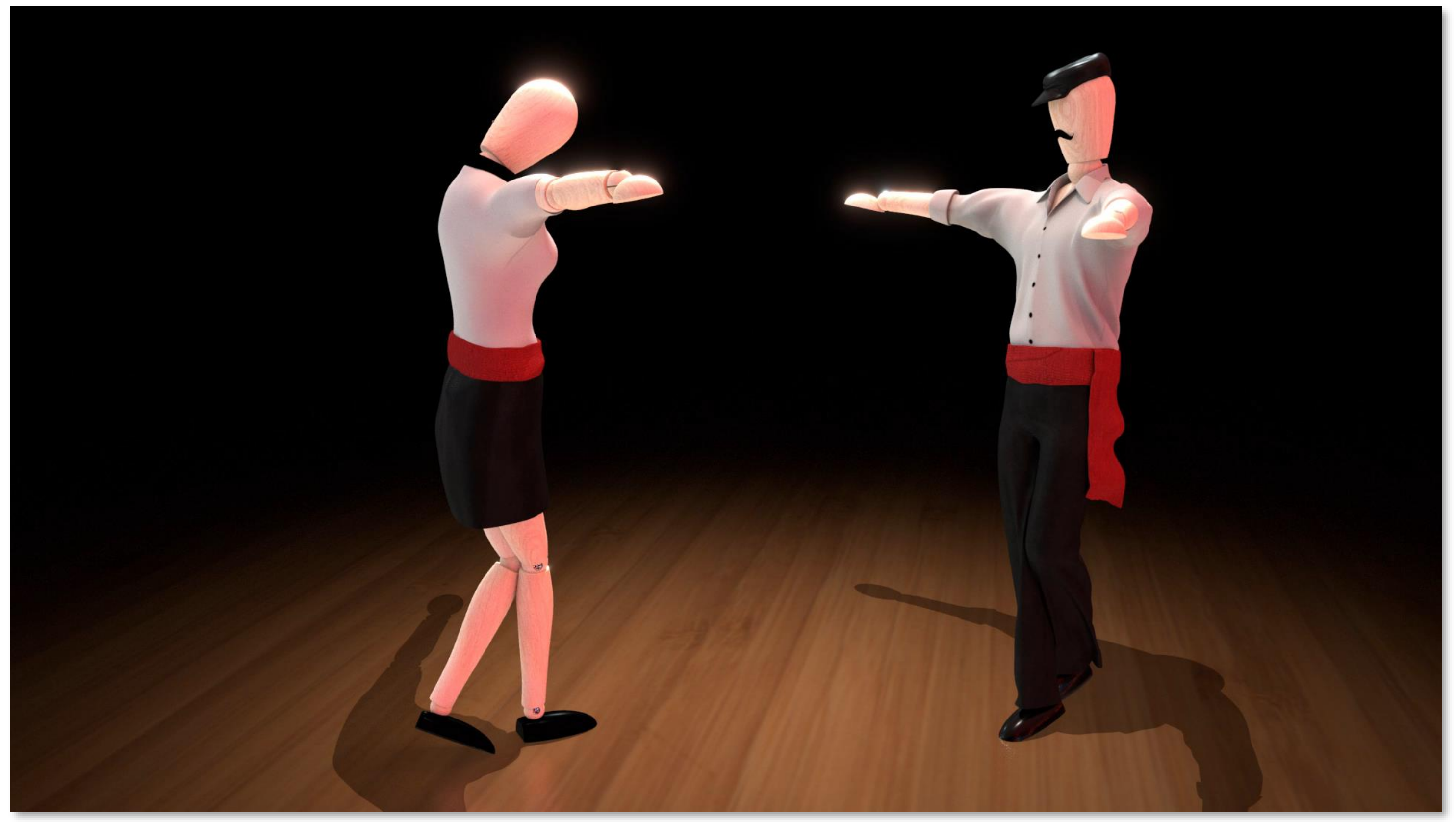


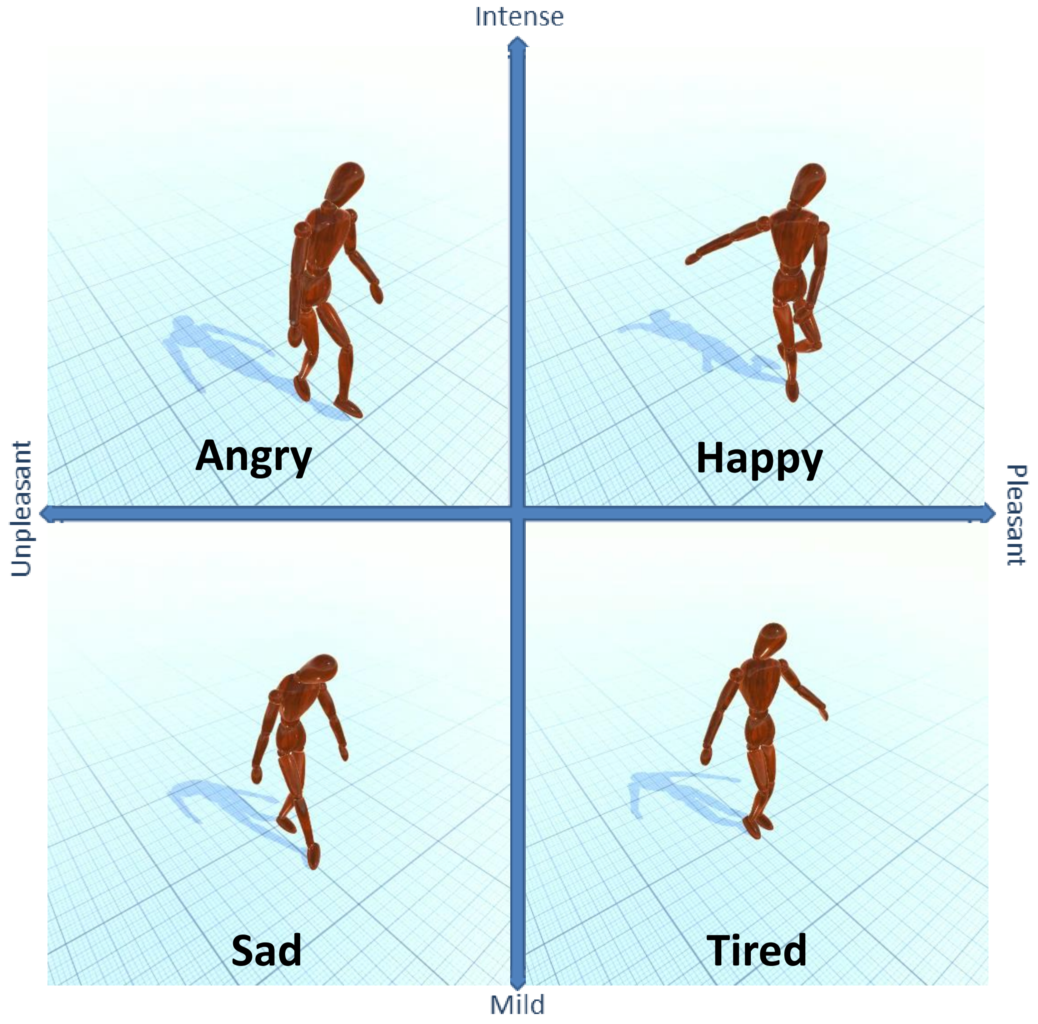
Zumba

Co-financed by the European Union
Connecting Europe Facility

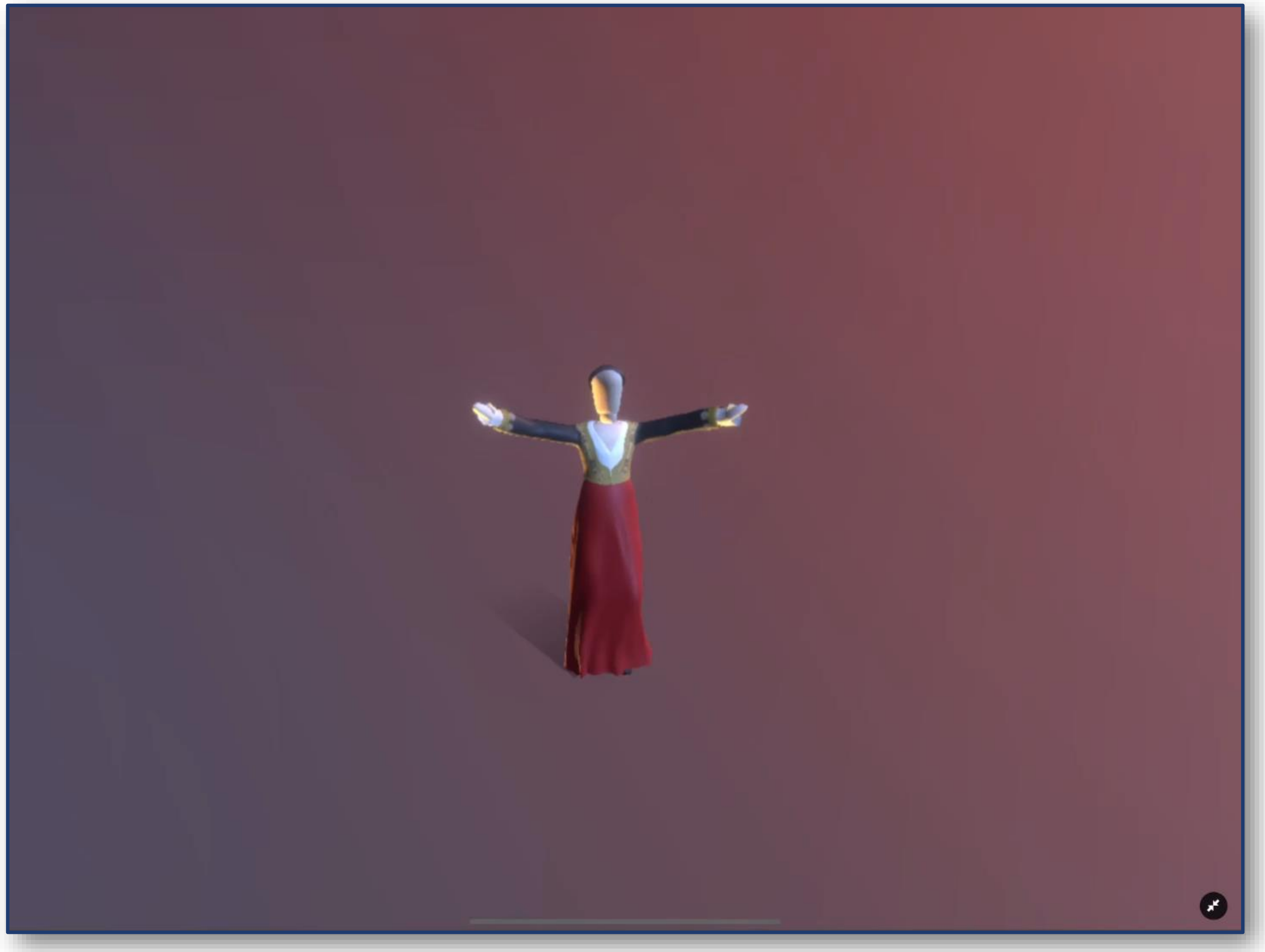
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

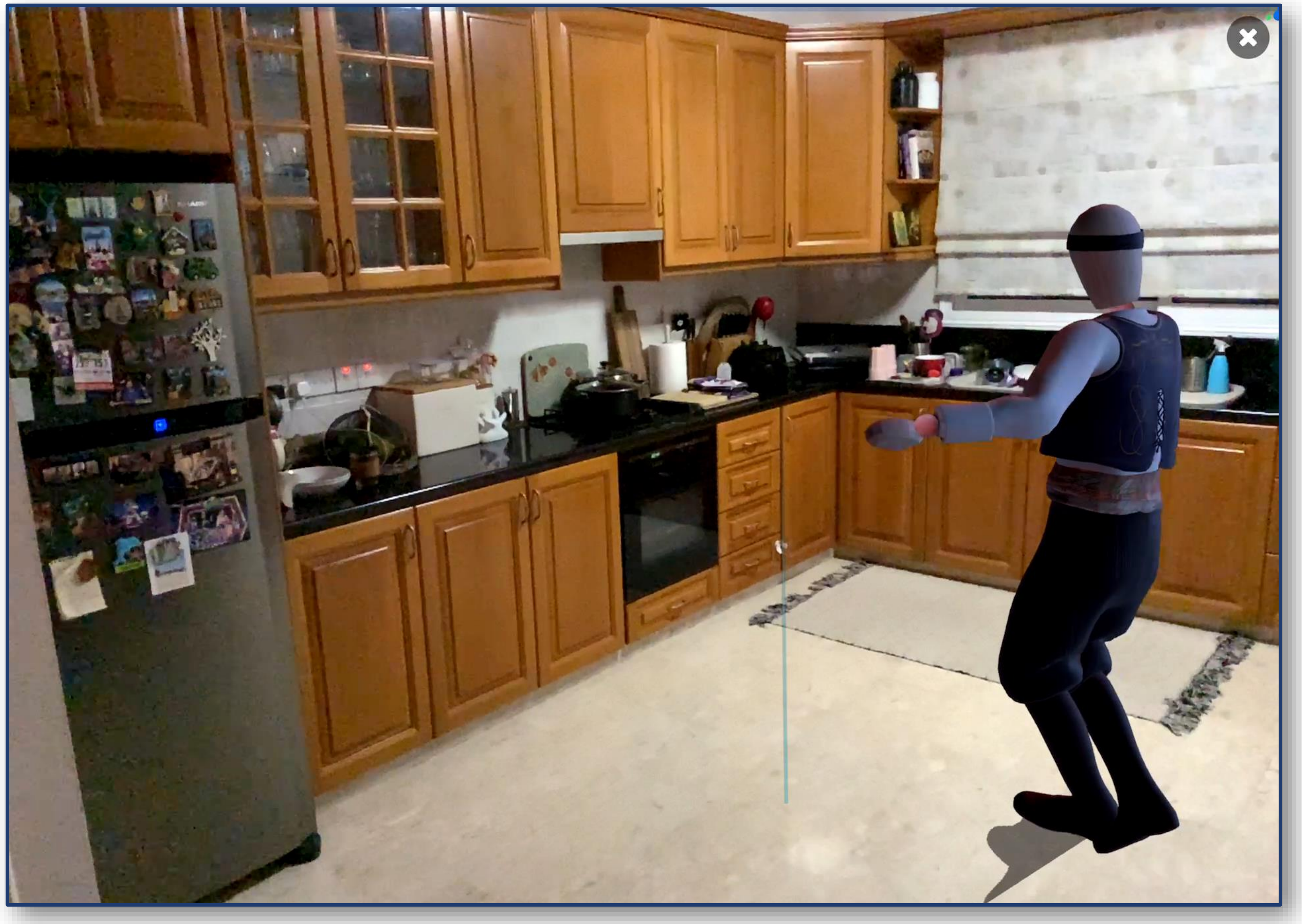
Our Dance Motion Capture Database



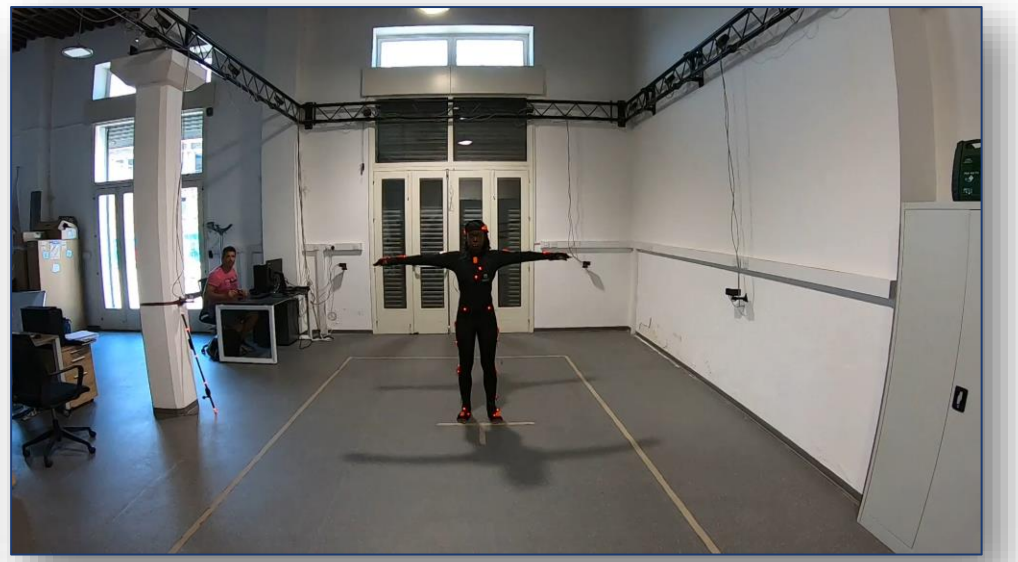


Our Dance Motion Capture Database





Our Dance Motion Capture Database



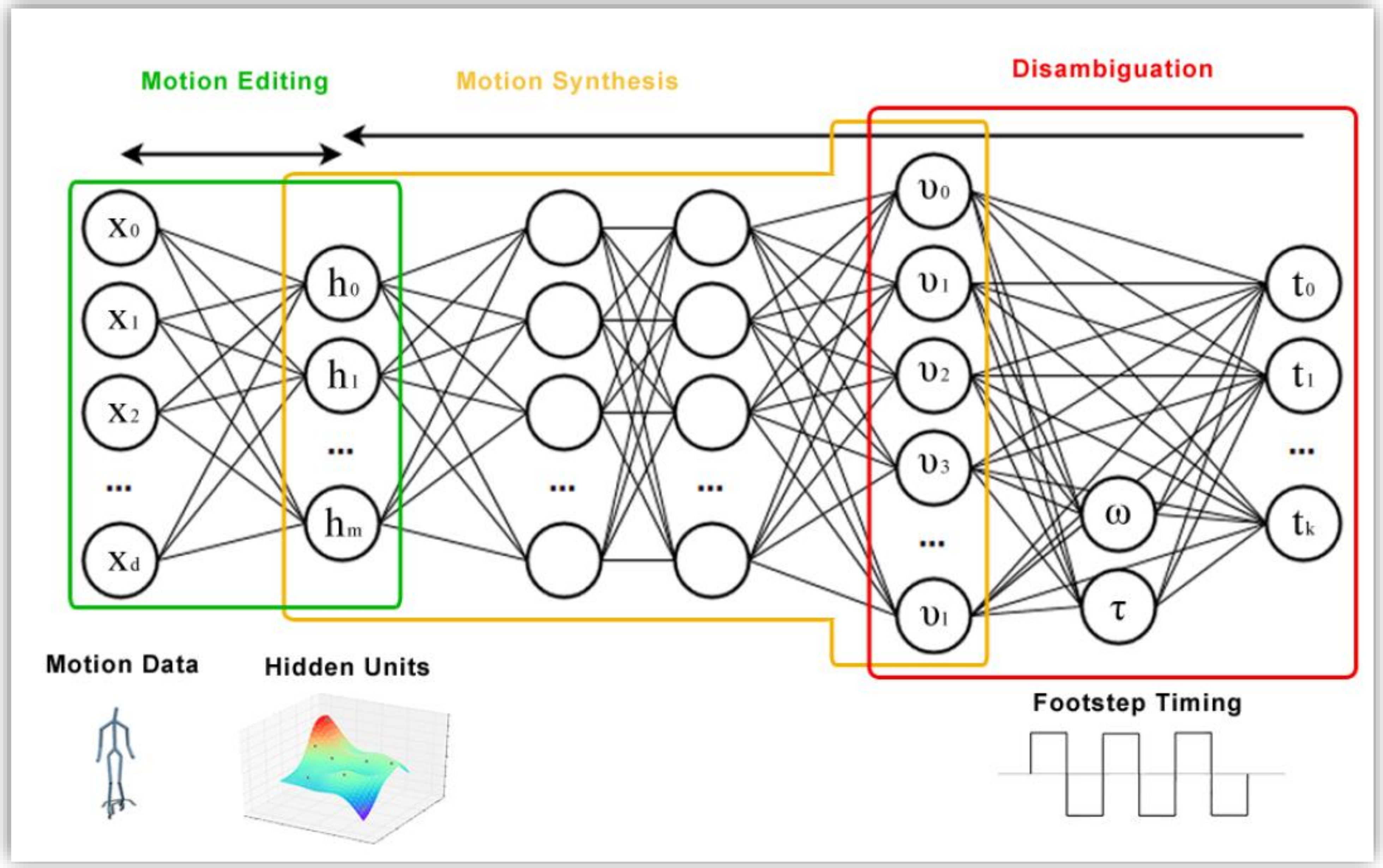
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

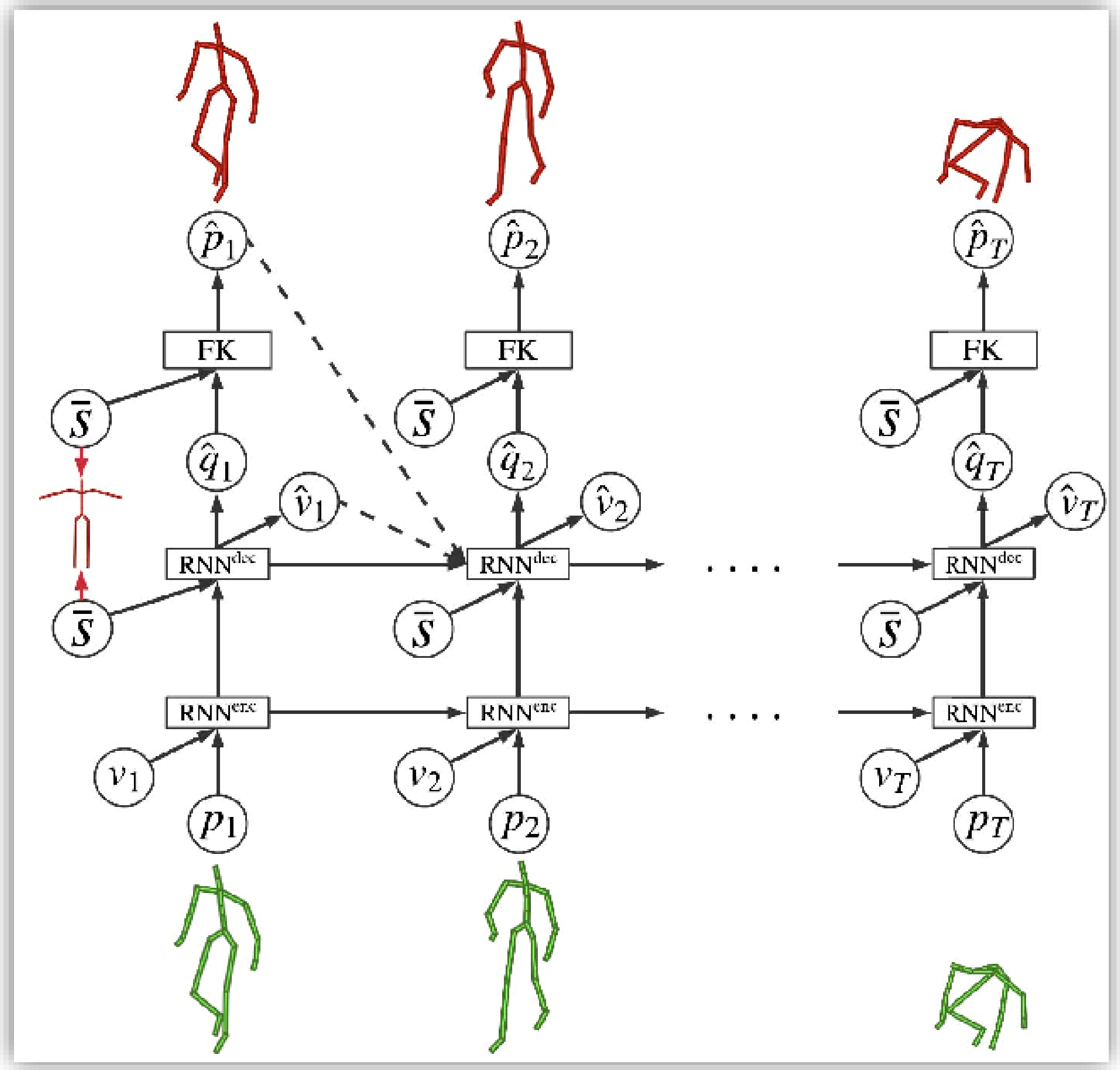
Deep Character Animation

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

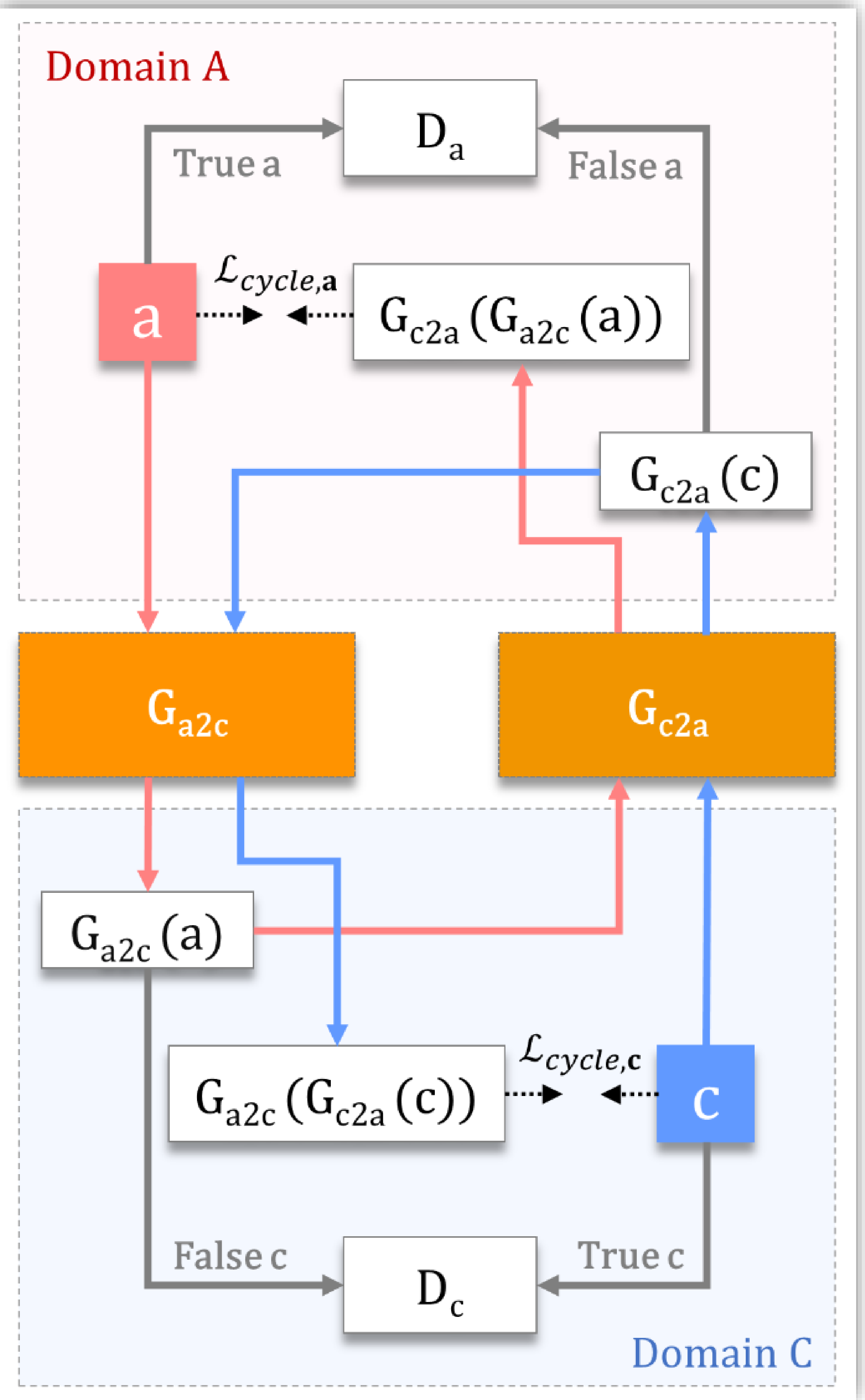
Deep Neural Networks



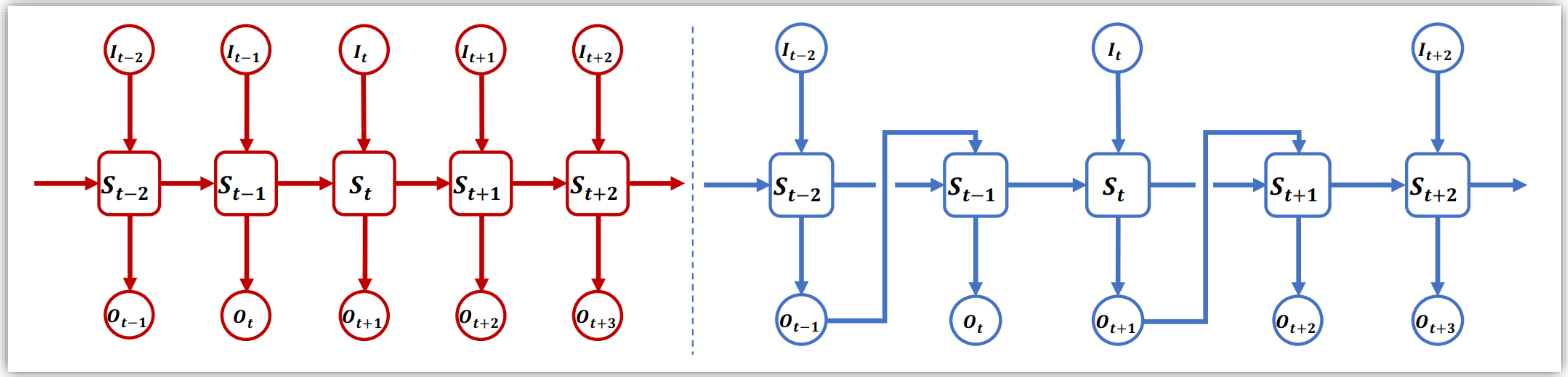
Convolutional Neural Networks (CNNs)



Recurrent Neural Networks (RNNs)



Generative Adversarial Networks (GANs)



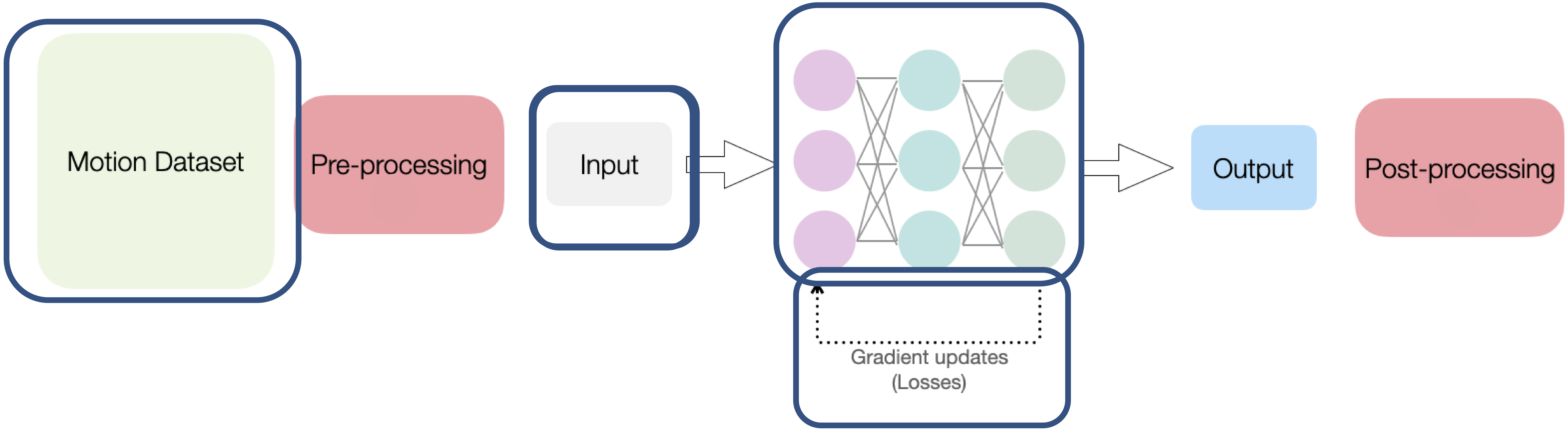
auto-conditional LSTM

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Motivation

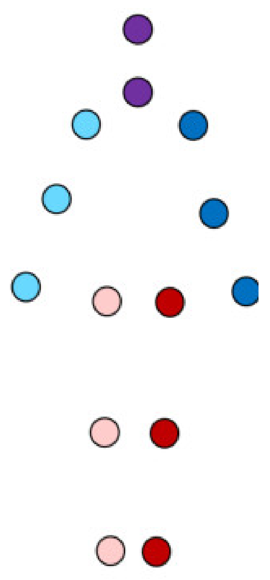
Character Animation with Deep Learning



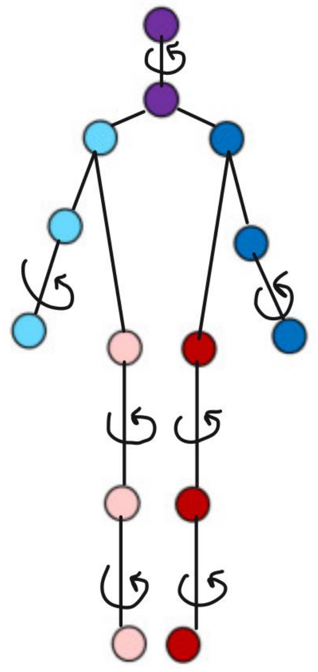
Motivation

Common Pose Representations

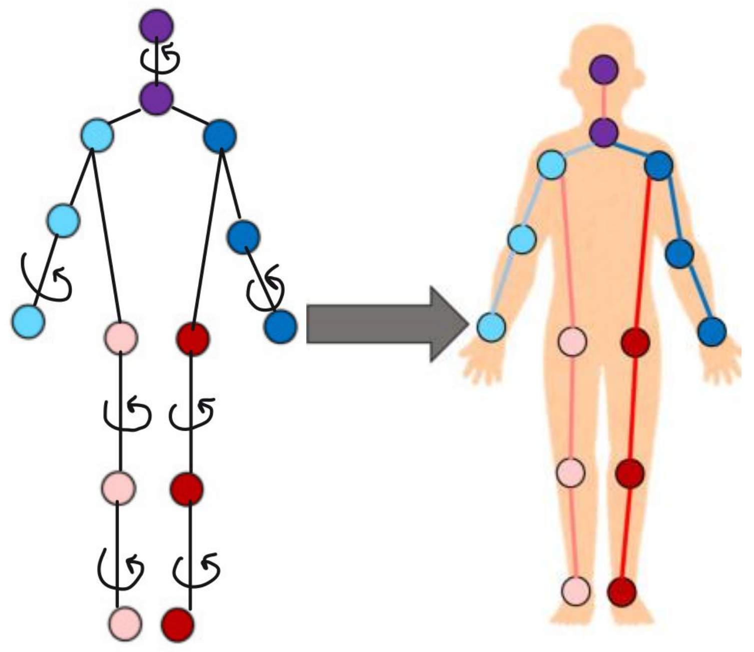
Positional



Angular



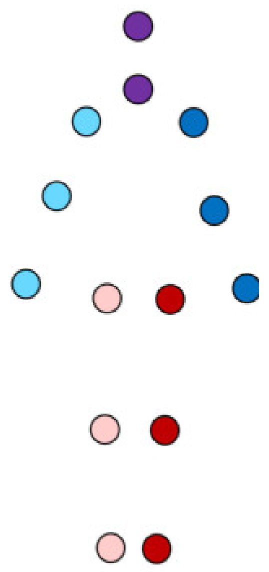
Hybrid



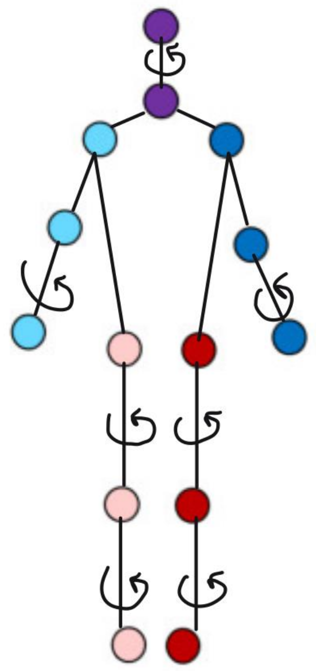
Motivation

Common Pose Representations

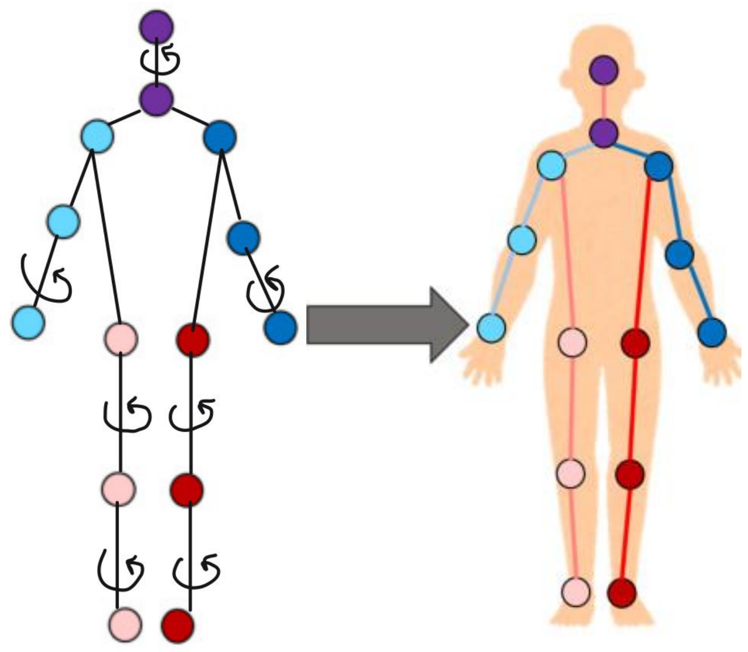
Positional



Angular



Hybrid



- Euclidean joint locations [Zhou et al., 2018]
- Motion Capture markers [Zhang et al., 2020]

- Exponential Maps
- Quaternions
- Euler angles
- Rotation Matrices
- Ortho6D [Zhou et al., 2019]

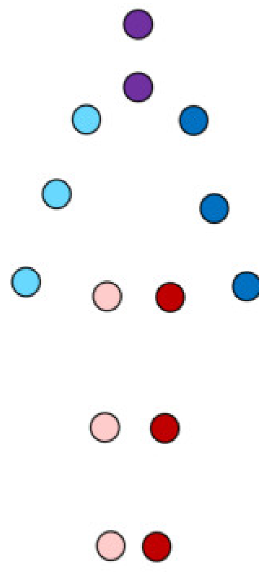
- Positional and angular
- Joint velocities/accelerations [Holden et al., 2017]

* Angular representations with positional losses [Aberman et al., 2020]

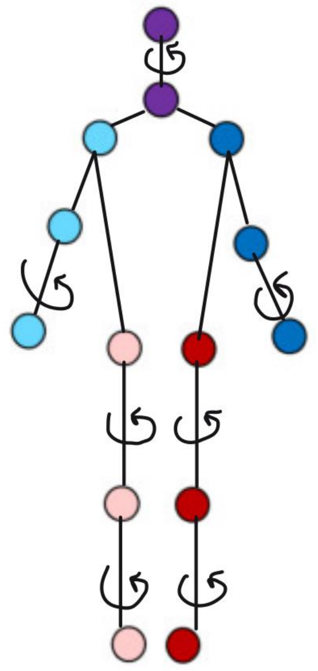
Motivation

Common Pose Representations

Positional



Angular



✓ Intuitive
✓ Visual result
✗ Not straightforward to apply to different characters

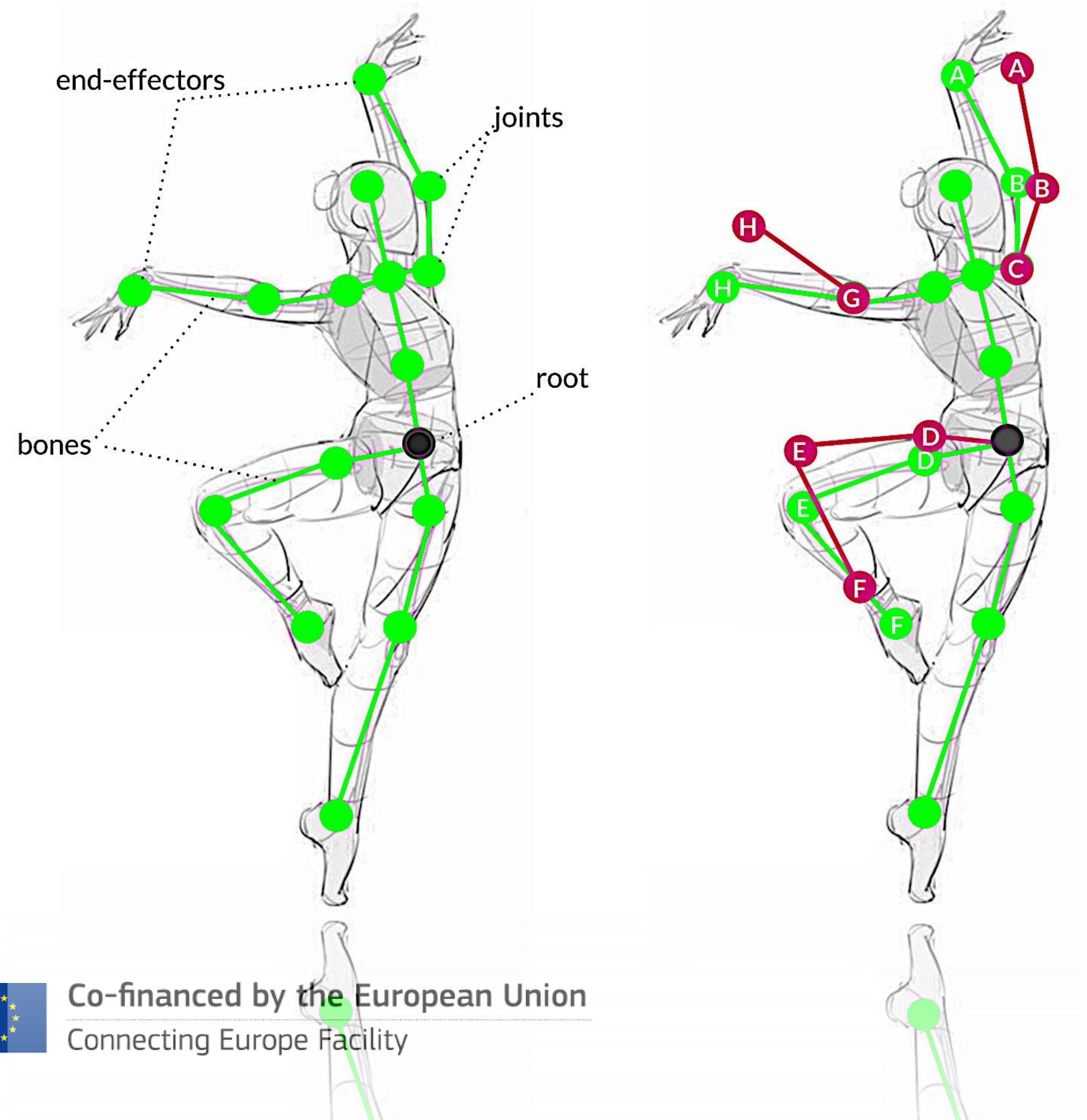
✓ Disentangle shape/skeletal proportions
✓ Convenient to work with
✗ Common rotation representations are discontinuous [Zhou et al., 2019]
✗ Error accumulation [Pavllo et al., 2018]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Method

Error accumulation along kinematic tree



Problem: Error accumulation along chain

- Angular representation causes problems in optimization-based methods

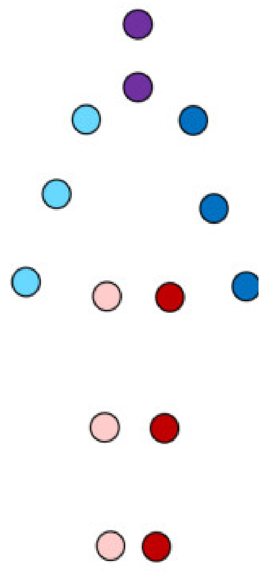
- Angular representations are often paired with loss that averages errors over joints
- Skeleton is a connected graph
- Ignores the fact that prediction errors of different joints have varying impact on qualitative results

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

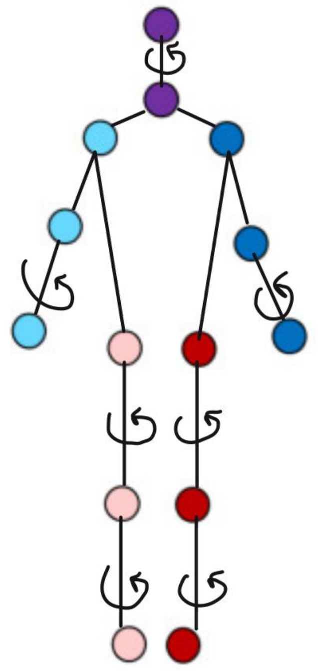
Motivation

Common Pose Representations

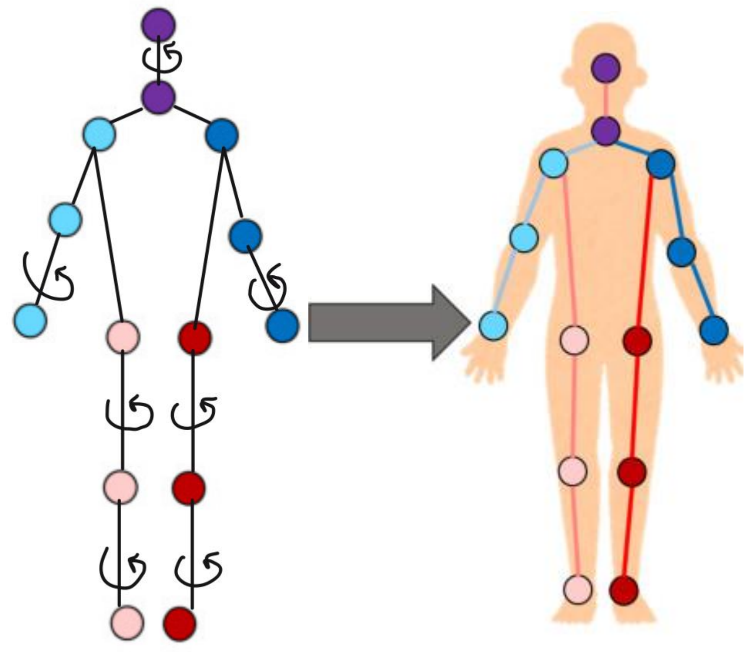
Positional



Angular



Hybrid



√ Intuitive
√ Visual result
✗ Not straightforward to apply to different characters

√ Disentangle shape/skeletal proportions
√ Convenient to work with
✗ Common rotation representations are discontinuous [Zhou et al., 2019]
✗ Error accumulation [Pavllo et al., 2018]

√ Combinations of positional + angular work better
√ Angular representations can be paired with positional losses (requires FK)
✗ Excessive information
✗ Correspondence often ignored
✗ Positional losses hinder the rotational information

Deep Neural Networks

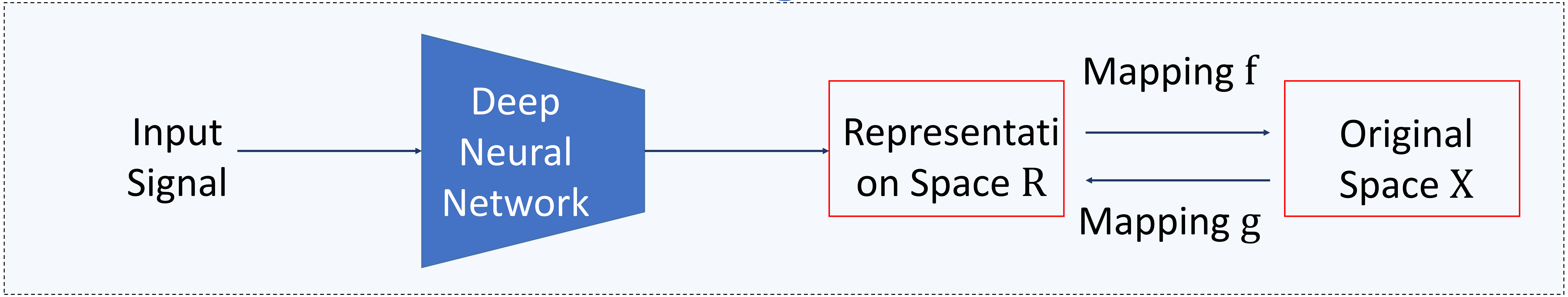
How to get a continuous representation in neural networks?

Let's say that the

- mapping to the original space $\mathbf{f}: \mathbf{R} \rightarrow \mathbf{X}$, and
- mapping to the representation space $\mathbf{g}: \mathbf{X} \rightarrow \mathbf{R}$.

We can say $(\mathbf{f}, \mathbf{g})$ is a good *representation* if for every $\mathbf{x} \in \mathbf{X}$; $\mathbf{f}(\mathbf{g}(\mathbf{x})) = \mathbf{x}$, that is, $\mathbf{f}$ is a left inverse of $\mathbf{g}$.

We can say the representation is continuous if $\mathbf{g}$ is continuous.



*more details in [Zhou et al. 2018]

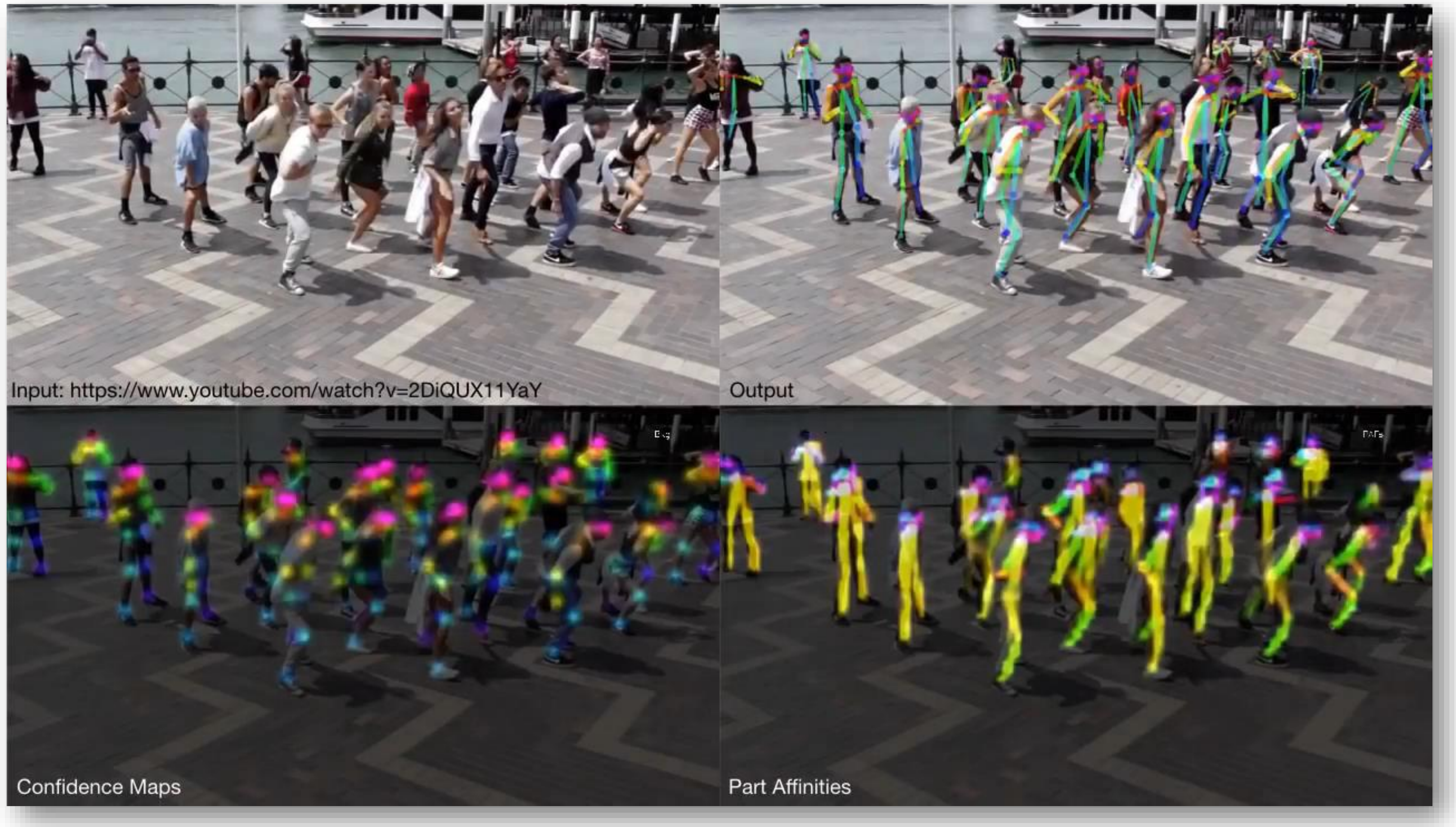
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

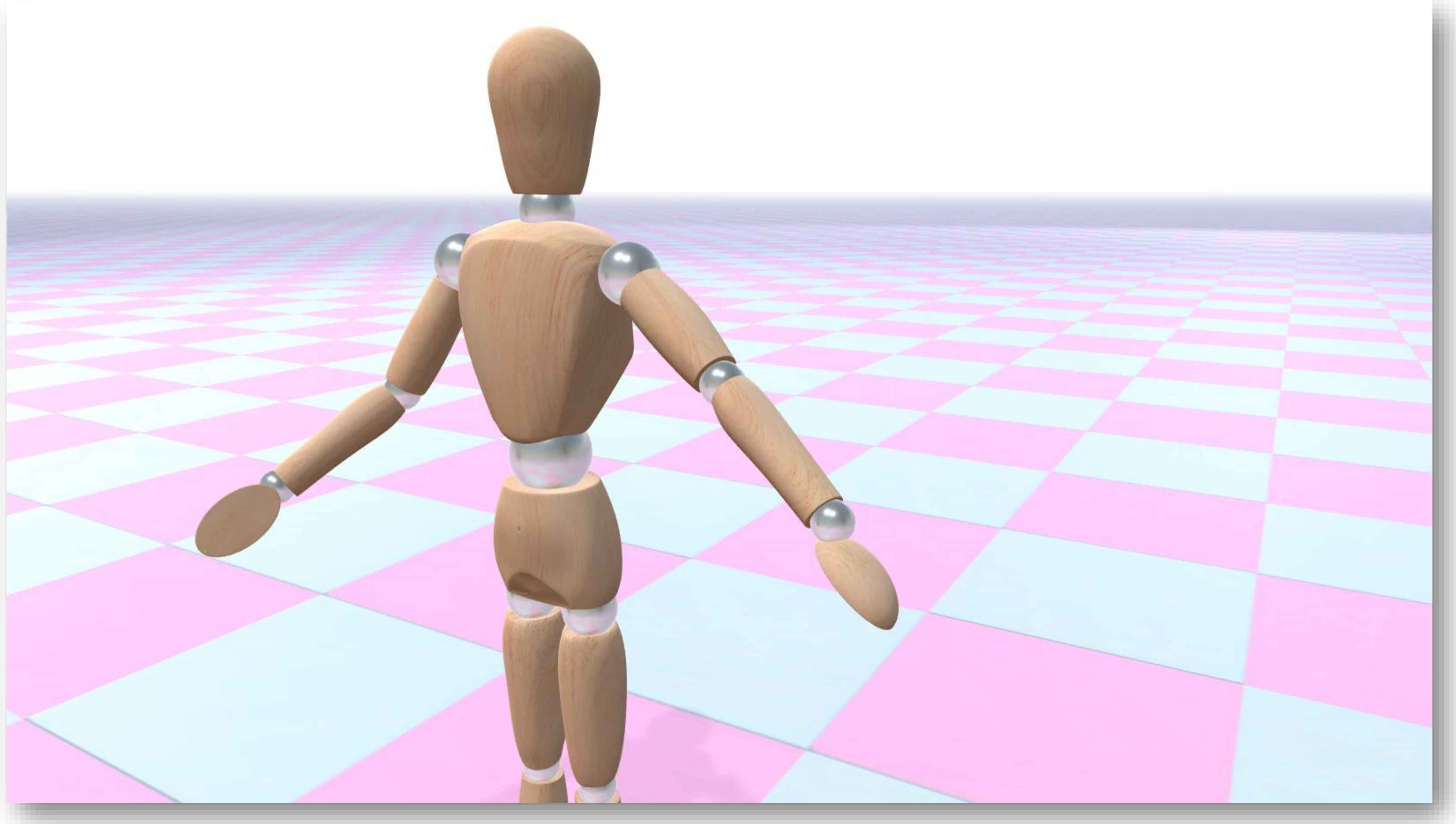
Positional Data

Have been used on early machine learning approaches

- **Advantages:** Good in continuity
- **Disadvantages : (a)** Ambiguity problems → cannot describe the full human motion articulation, **(b)** Skeletal model violations



[Cao et al. 2018]

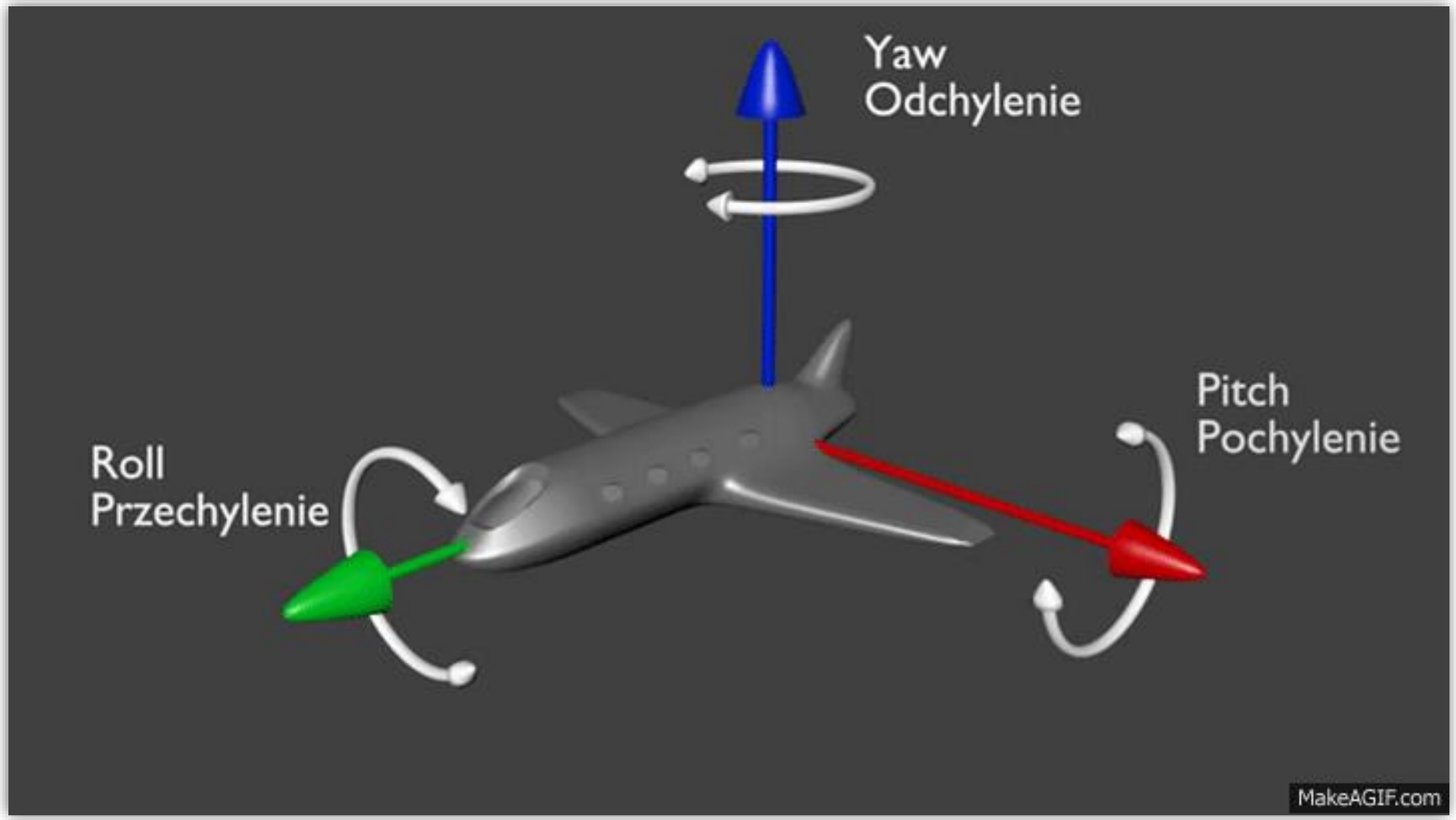


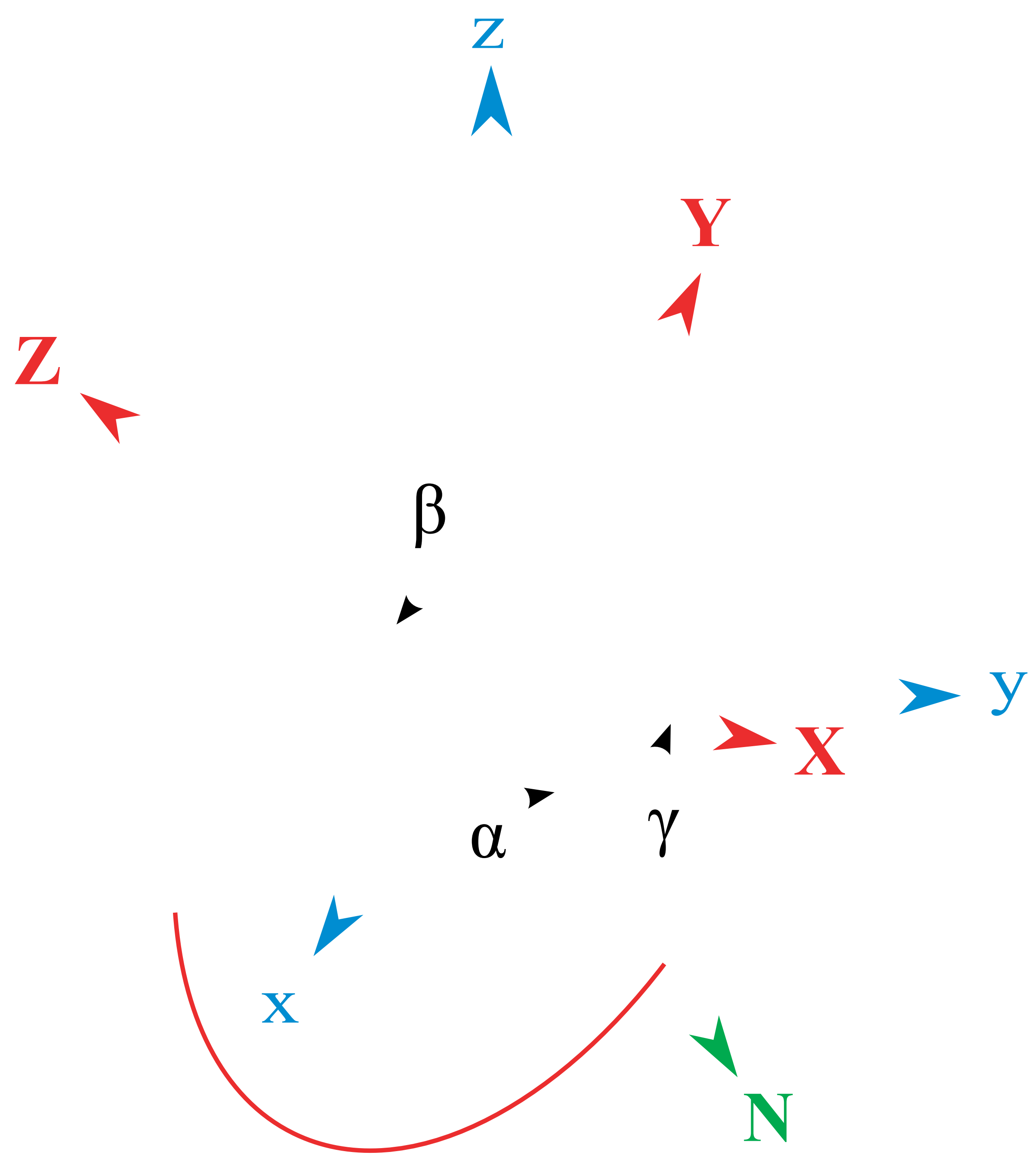
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Euler Angles

Rotate the angles of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ along the $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ axes from the reference frame.



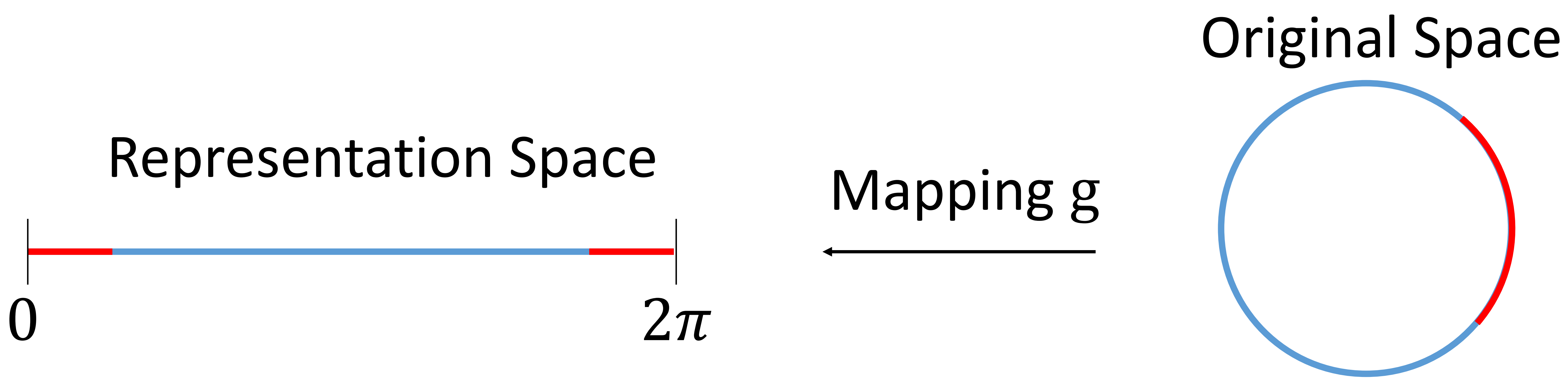


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Euler Angles: *Limitations*

- Gimbal Lock
- Discontinuity
- Singularities that cause learning problems



Disconnected Set of Angular Representations in $[0, 2\pi]$

Connected Set of Rotations

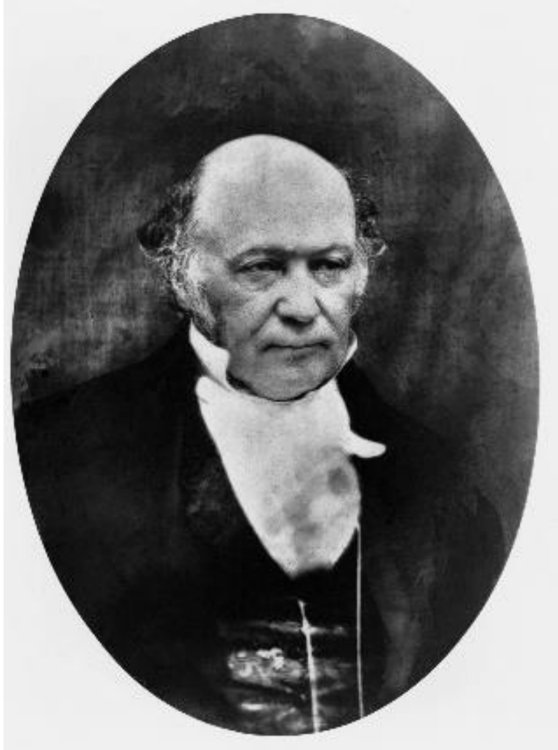
Co-financed by the European Union
Connecting Europe Facility

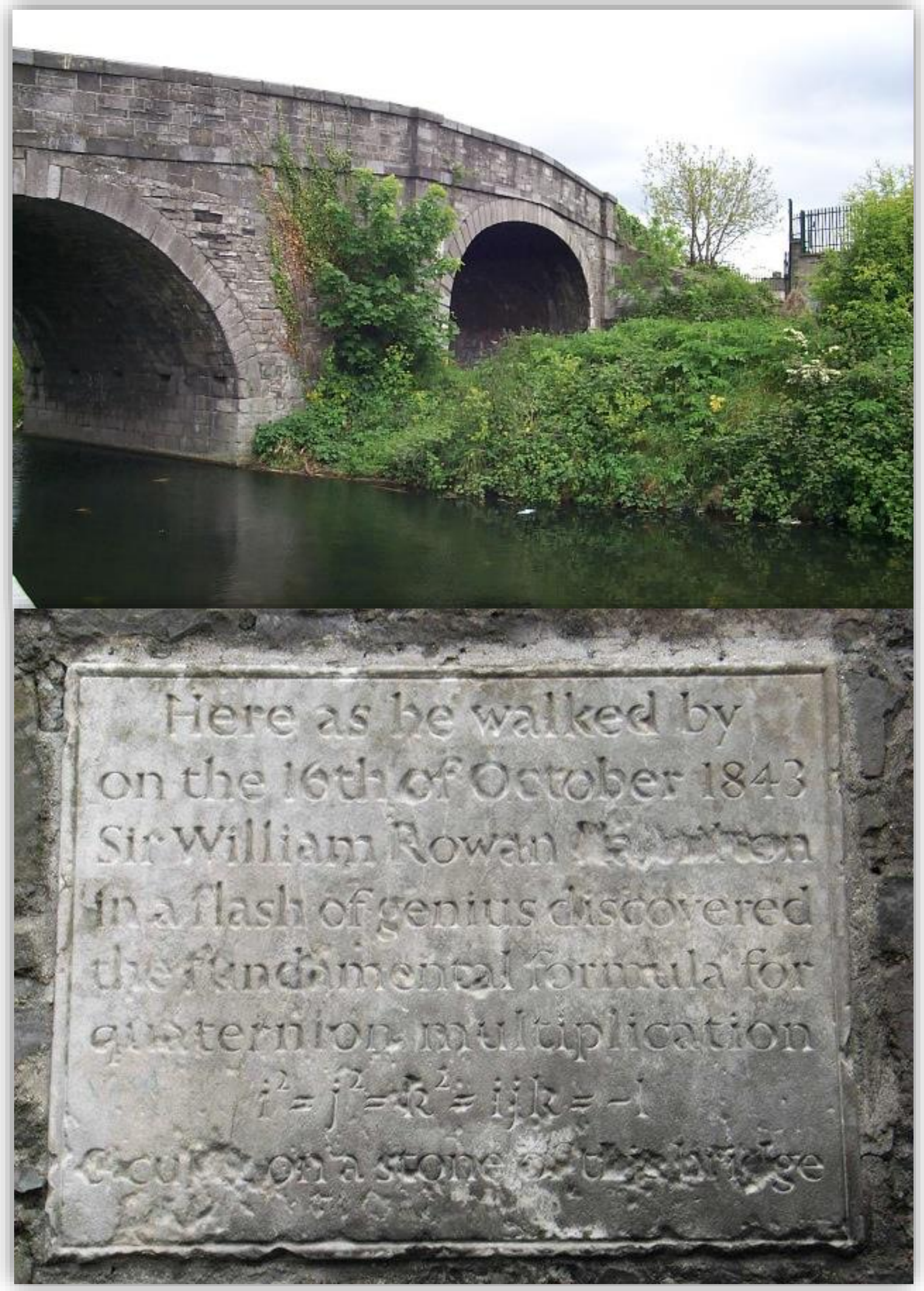
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Quaternions

- Mathematical abstractions alternative to Euler Angles
- Revised and Formulated by Sir William R. Hamilton in 1843
- 4-D complex numbers
  - With one real axis
  - And three imaginary axes, the basis vectors

$$\mathbf{i}, \mathbf{j}, \mathbf{k}$$





Hamilton Math Inst., Trinity College

How are quaternions represented?

$$\mathbf{q} = (w, \mathbf{V}) = w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

or

$$\mathbf{q} = (q_0, \mathbf{V}) = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$$

Co-financed by the European Union
Connecting Europe Facility

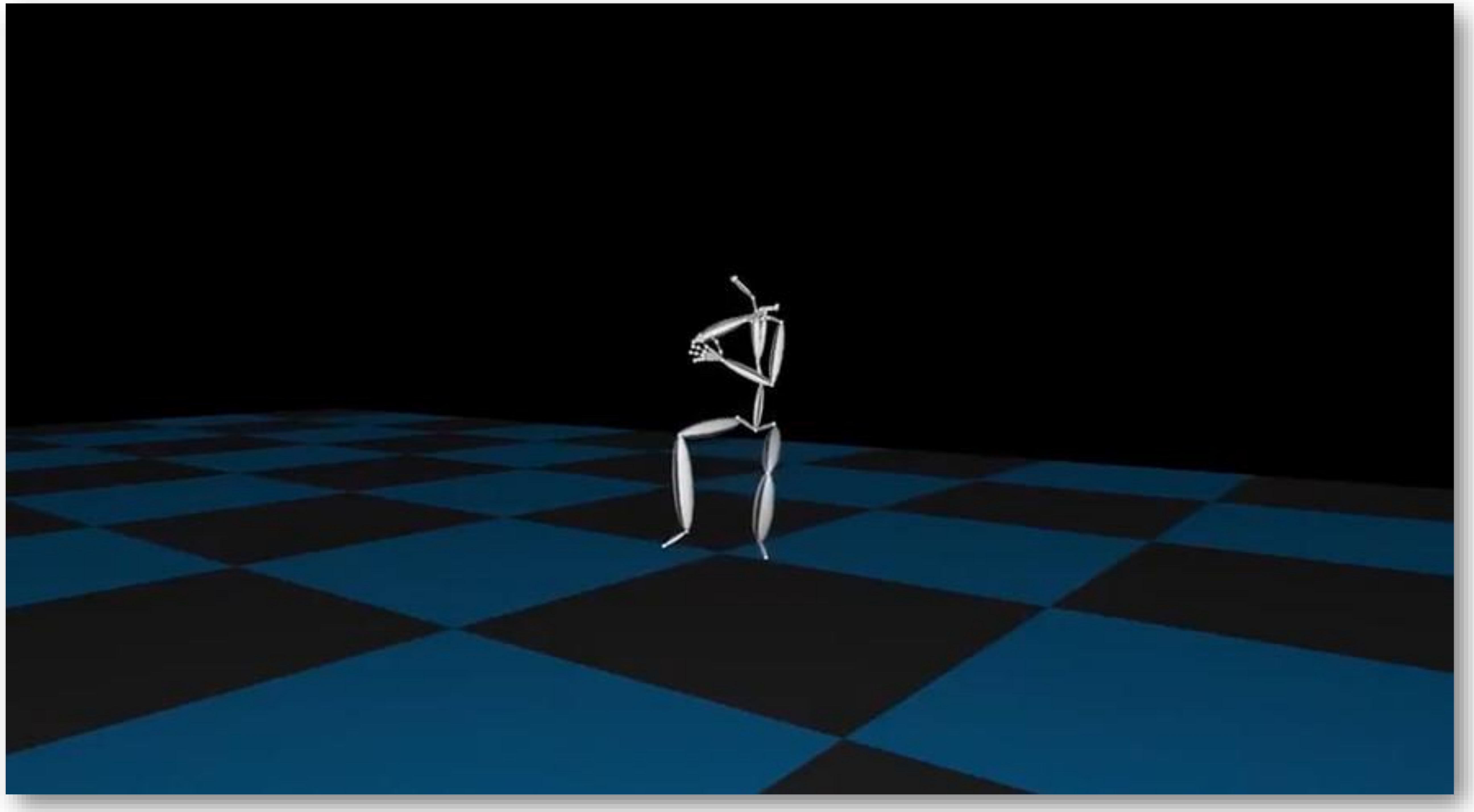
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Motion representation in popular works

In an attempt to overcome these limitations, the character animation community proposed some alternatives/improvements:

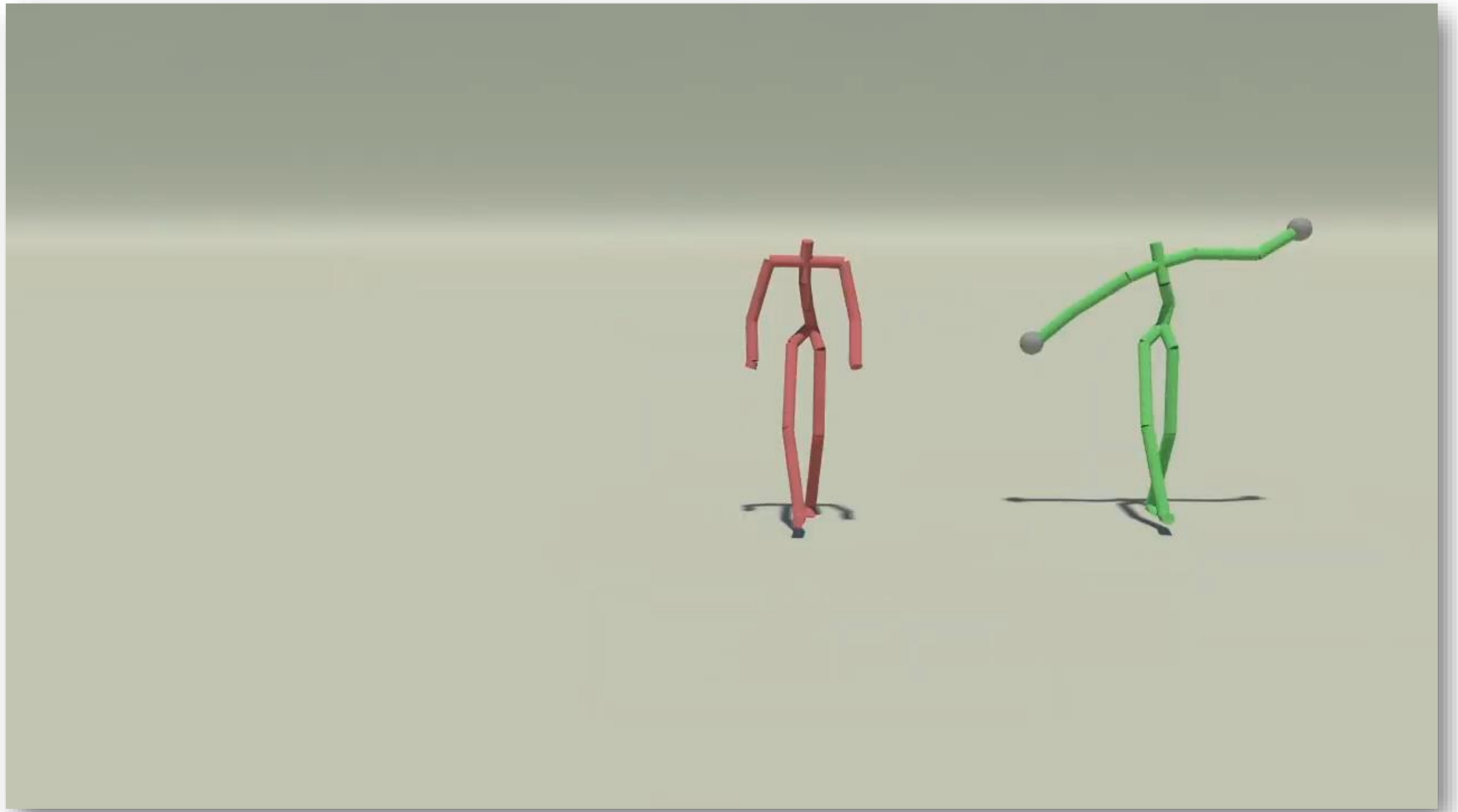
Training using only positional data:

- Zhou et al. 2018. Auto-Conditioned Recurrent Networks for Extended Complex Human Motion Synthesis. International Conference on Learning Representations



Positional data, with bone length constraints:

- Holden et al. 2016. A deep learning framework for character motion synthesis and editing. ACM Trans. Graphics.
- Wang et al. 2021. Spatio-temporal manifold learning for human motions via long-horizon modelling. IEEE Trans. Visualization and Computer Graphics.



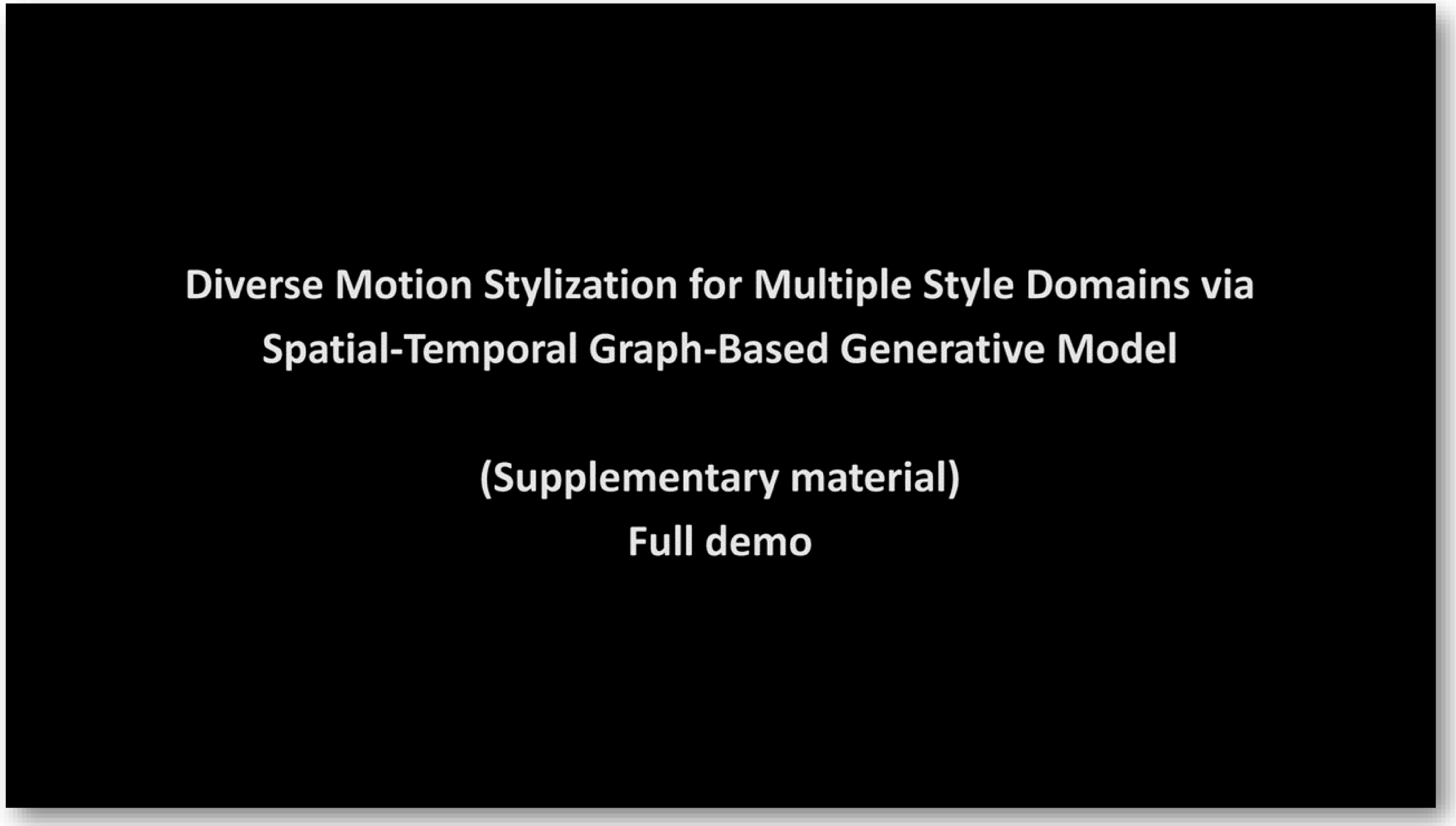
Quaternions, with a Forward Kinematic layer so as to add a positional loss:

- Harvey et al. 2020. Robust Motion In-betweening. ACM Trans. Graphics.
- Aberman et al. 2020. Skeleton-Aware Networks for Deep Motion Retargeting. ACM Trans. Graphics.

Pavllo et al. 2018. QuaterNet: A Quaternion-based Recurrent Model for Human Motion. British Machine Vision Conference

Quaternions, amended with positional data:

- Park et al. 2021. Diverse Motion Stylization for Multiple Style Domains via Spatial-Temporal Graph-Based Generative Model. ACM Comput. Graph. Interact. Tech.

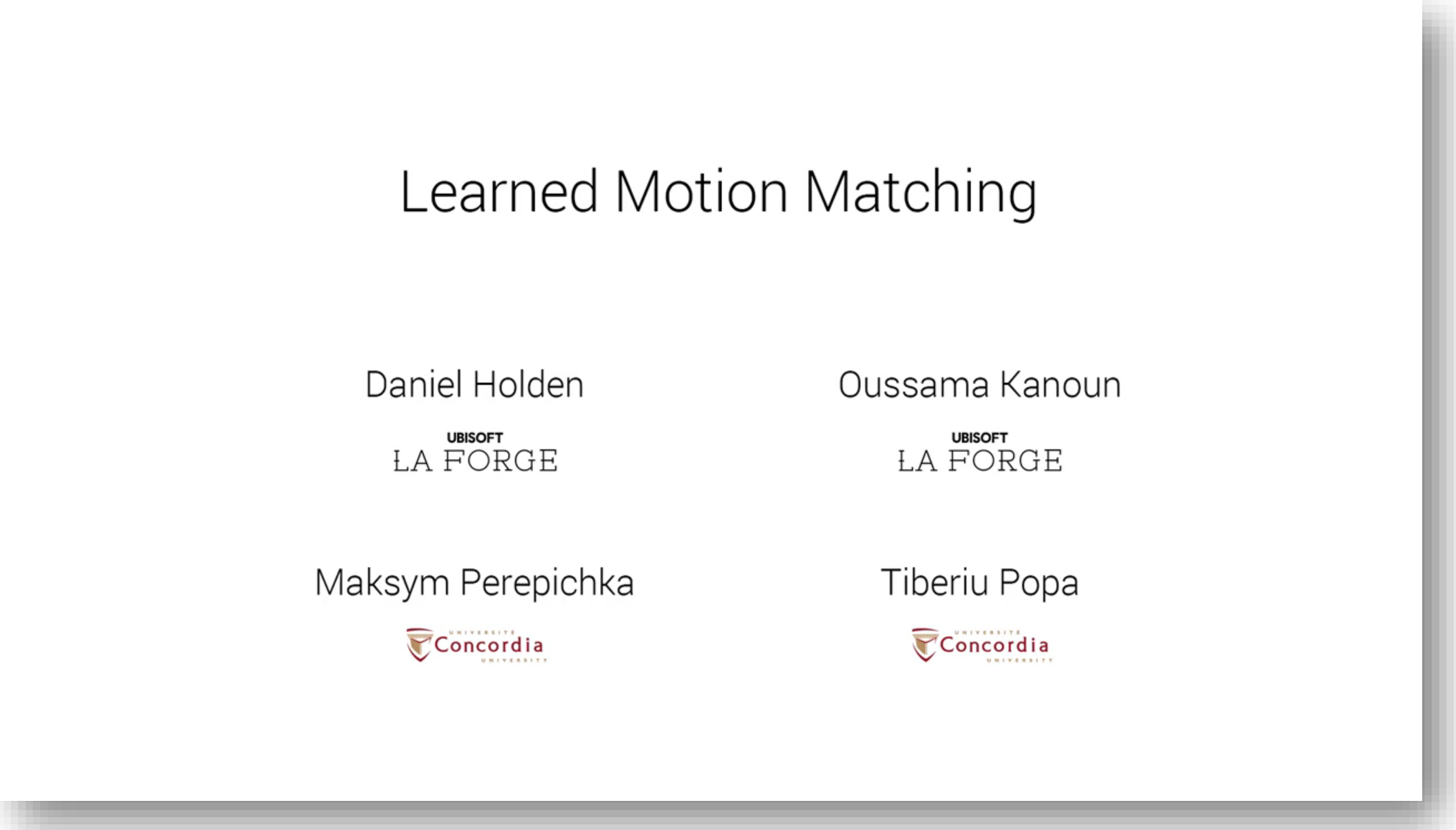


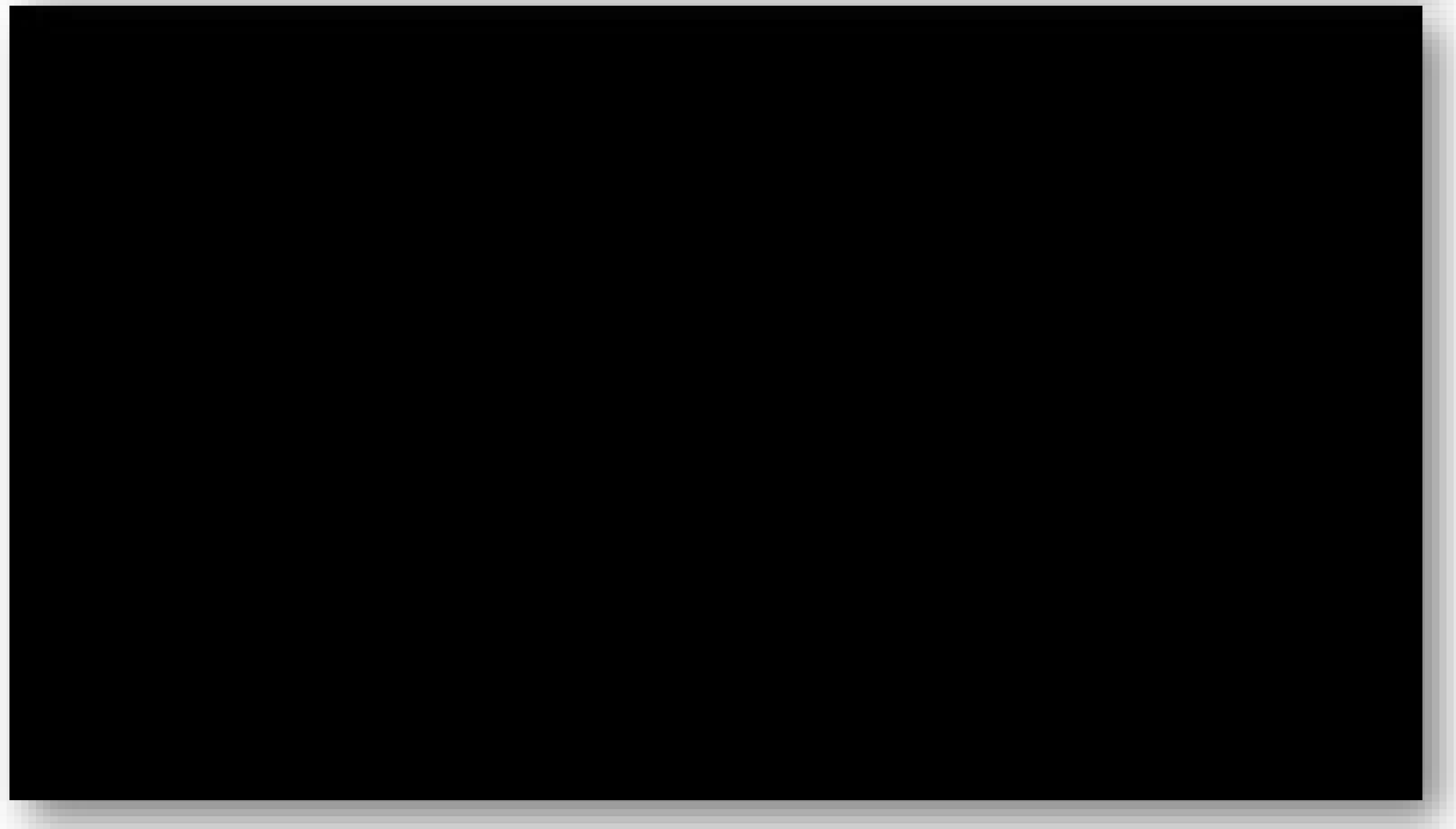
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Quaternions, amended with positional data, and motion dynamics:

- Holden et al. 2021. Learned Motion Matching. ACM Trans. Graphics
- Starke et al. 2021. Neural Animation Layering for synthesizing martial arts movements. ACM Trans. Graphics
- Starke et al. 2021. Neural state machine for character-scene interactions. ACM Trans. Graphics





This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Method

Dual Quaternion Representation

- Hybrid representation based on Dual Quaternions
- Unified entity

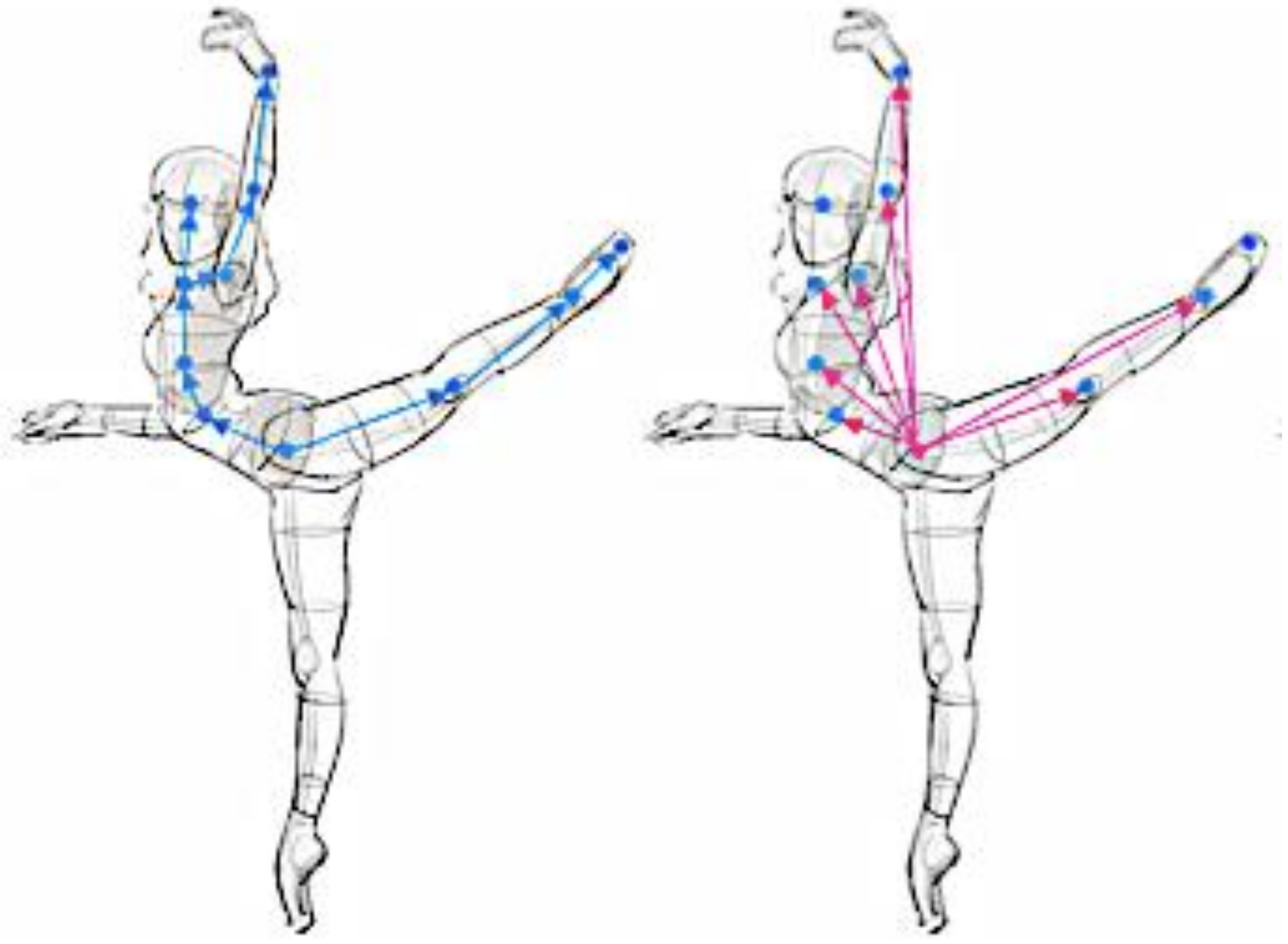
$$q = \underbrace{q_r}_{\text{rotation}} + \epsilon \underbrace{q_d}_{\text{translation}} \text{ where } \epsilon^2 = 0$$

- More compact than homogeneous transformation matrix (8 values per joint) and efficient [Kenwright et al., 2012]
- Well-established mathematical properties

Method

Dual Quaternion Representation

- Can be defined in a root-centered coordinate system mitigating common problems such as error accumulation along the kinematic chain [Pavllo et al., 2018]

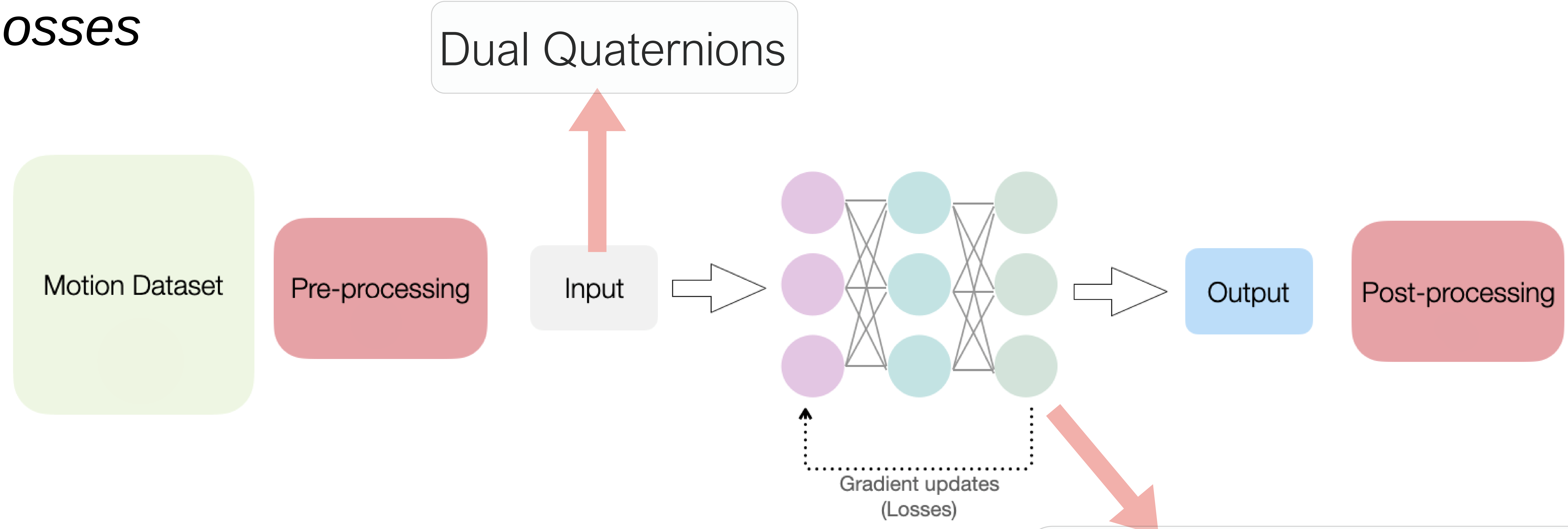


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

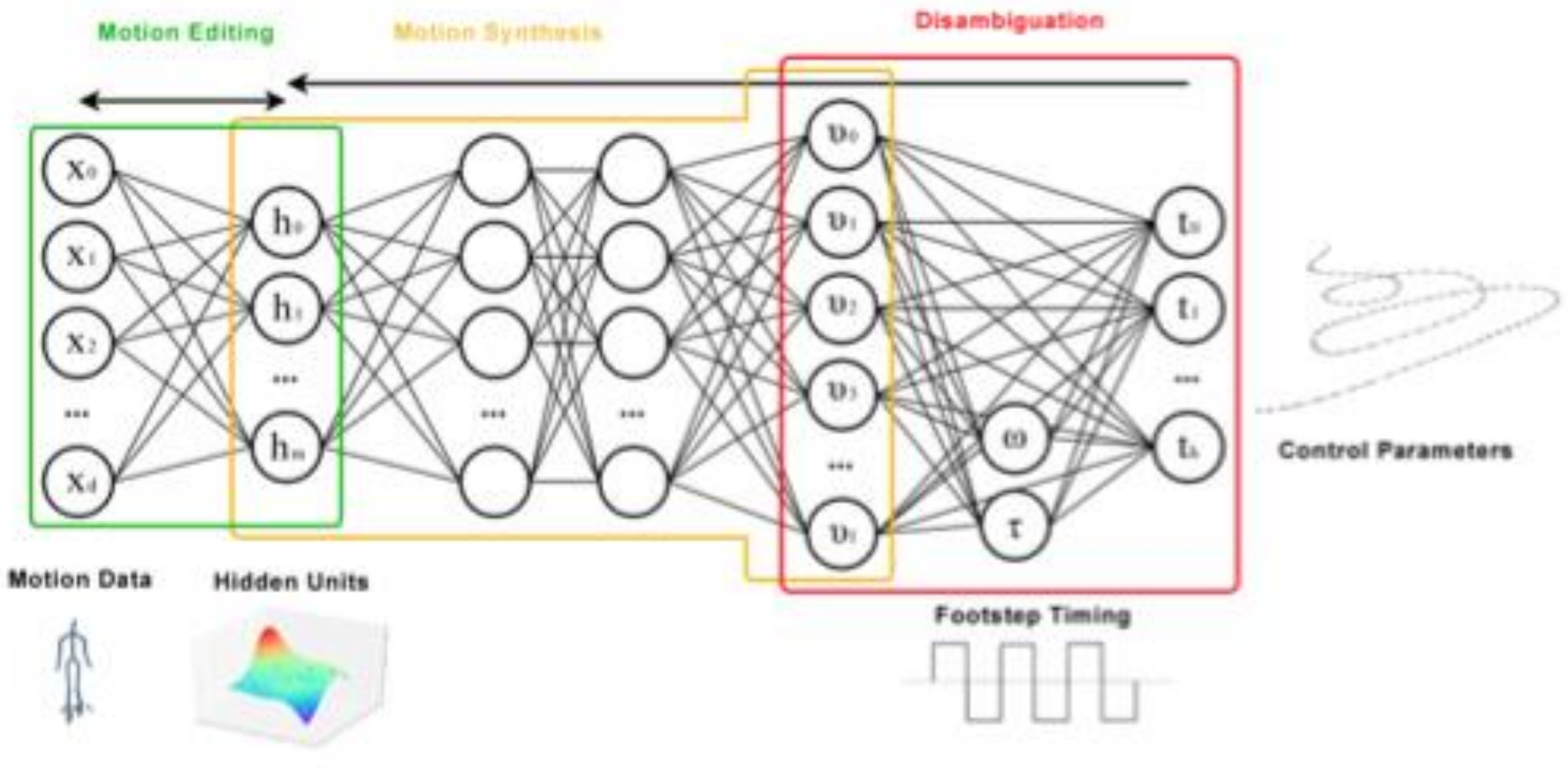
Method

Losses

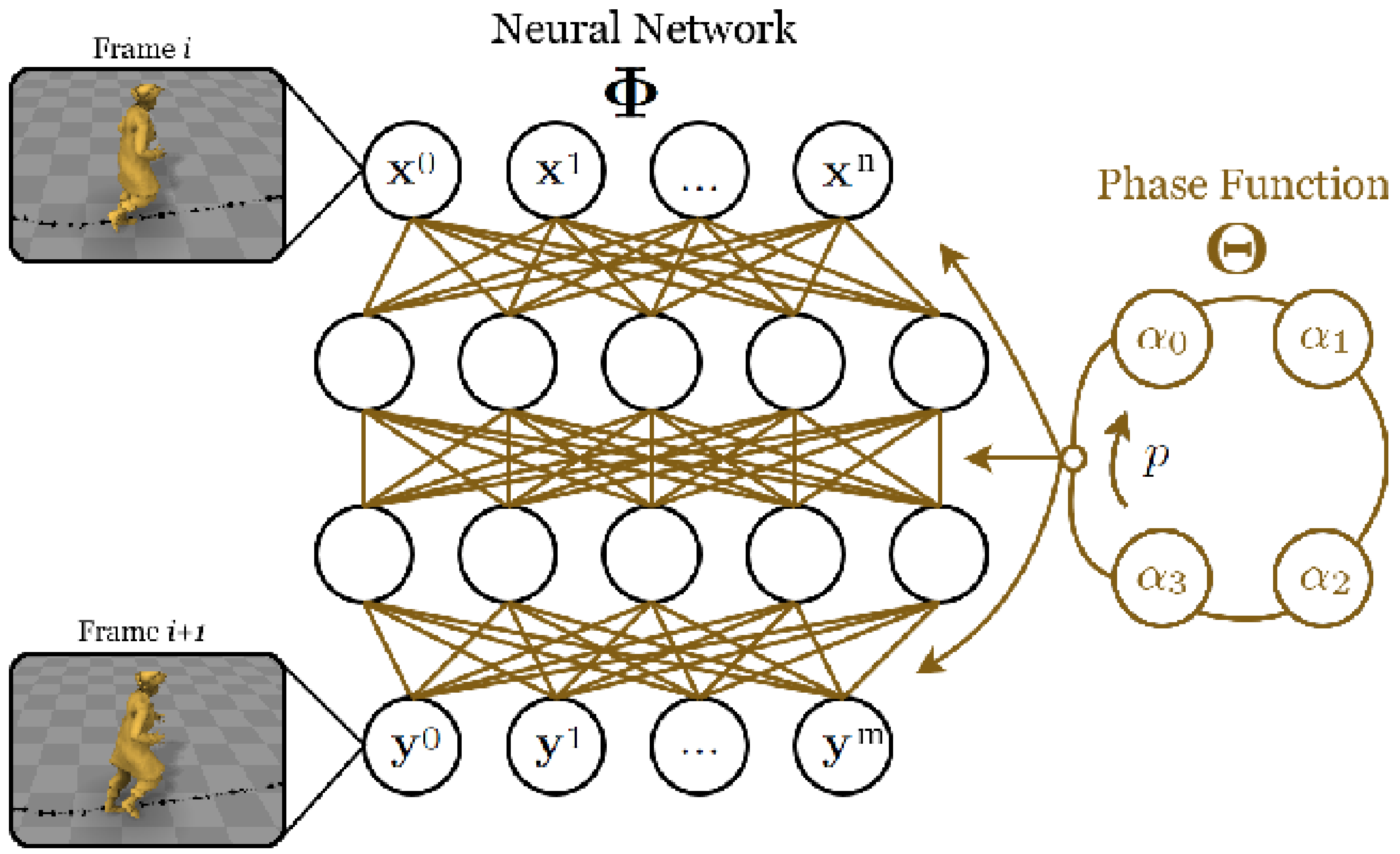


- Euclidean distance of joint positions/locations
- MSE on joint rotations
- **Offset loss** → maintain skeletal structure

Deep character animation networks

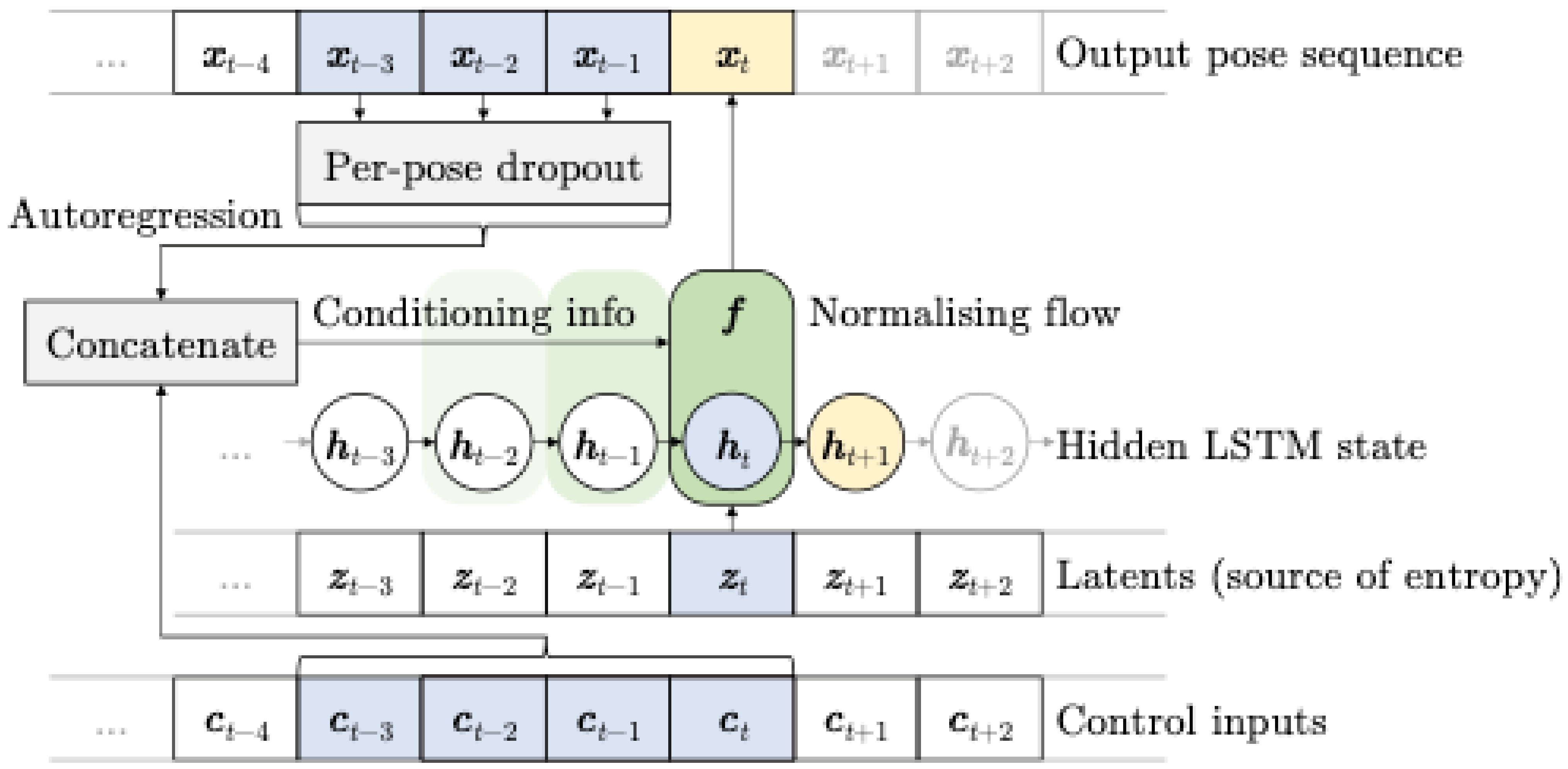


Holden et al., 2016



Holden et al., 2017

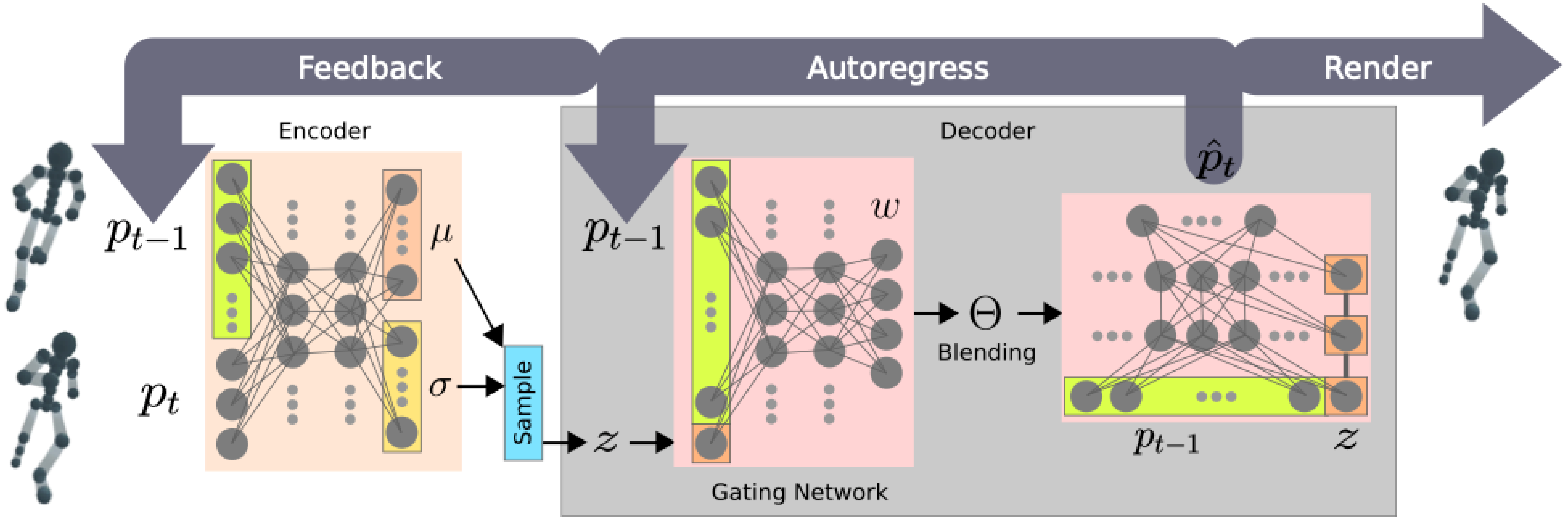
Deep character animation networks



Alexanderson et al., 2020

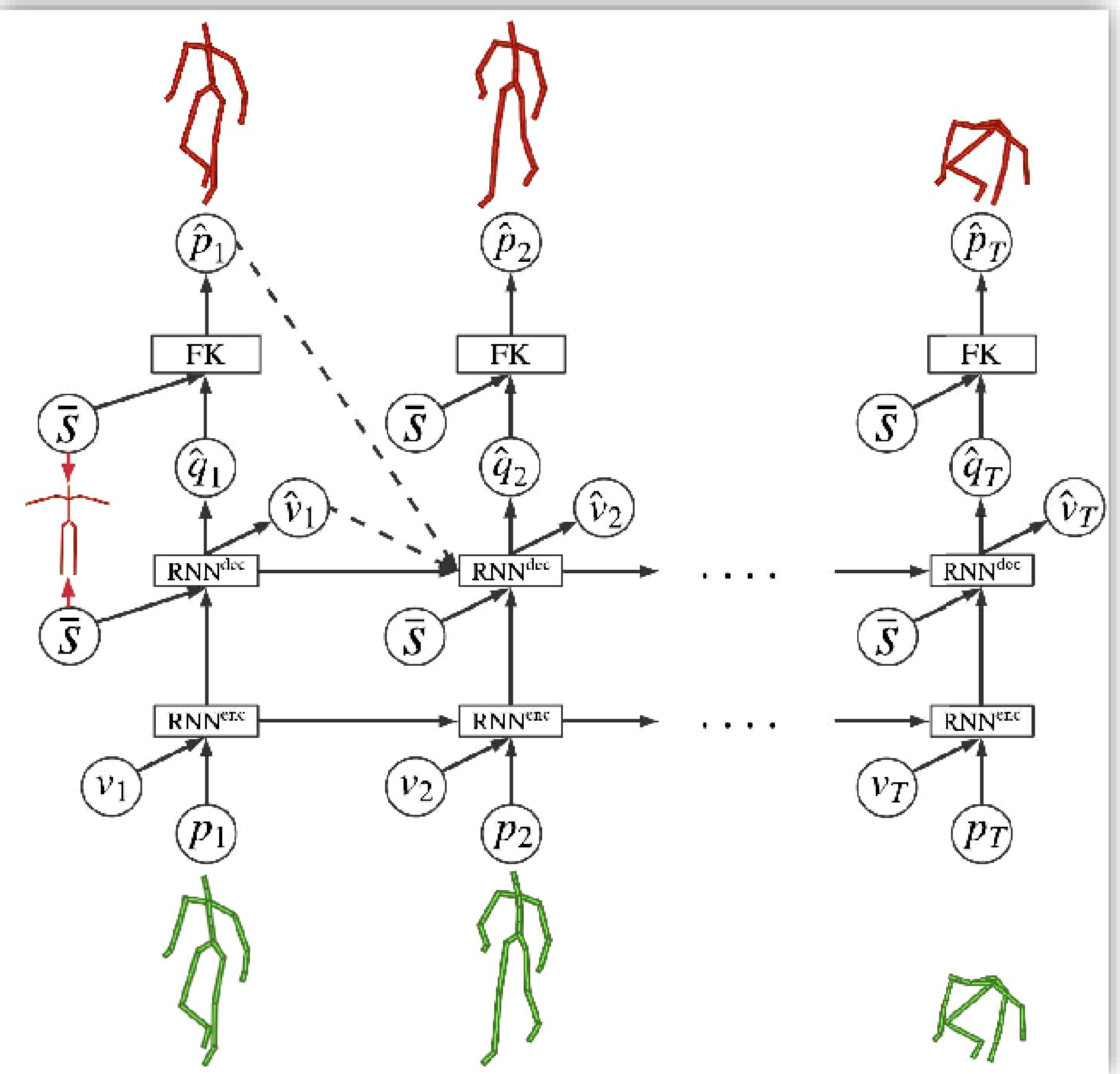
Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep character animation networks

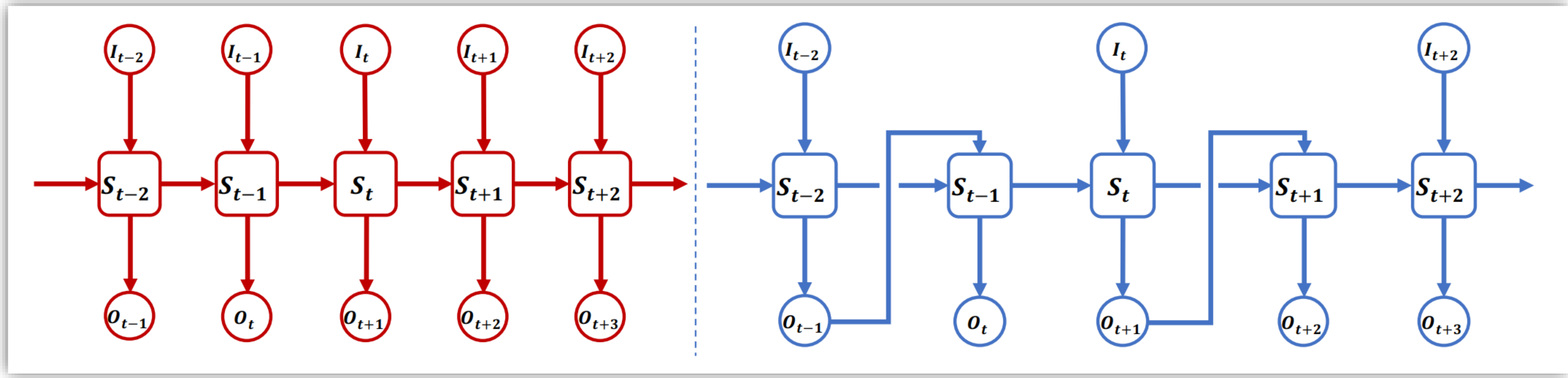


Ling et al., 2021

Deep character animation networks



Frangiadaki et al. 2015



Zhou et al. 2018

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Many challenges in Character Animation have been re-defined

Rigging/Skinning

Motion Synthesis

Motion in-betweening

Motion Control

Motion Retargeting

Style Transfer

Audio/music-driven synthesis

Text-to-animation

etc.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Research in our lab

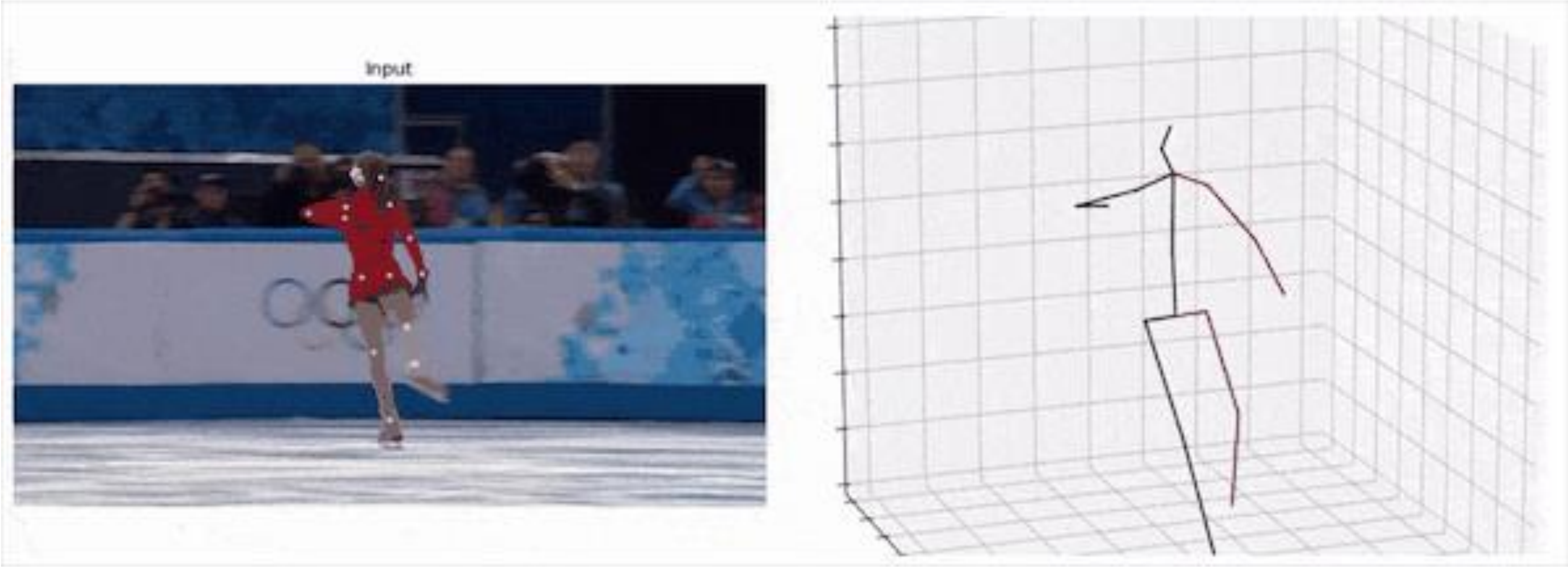
Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

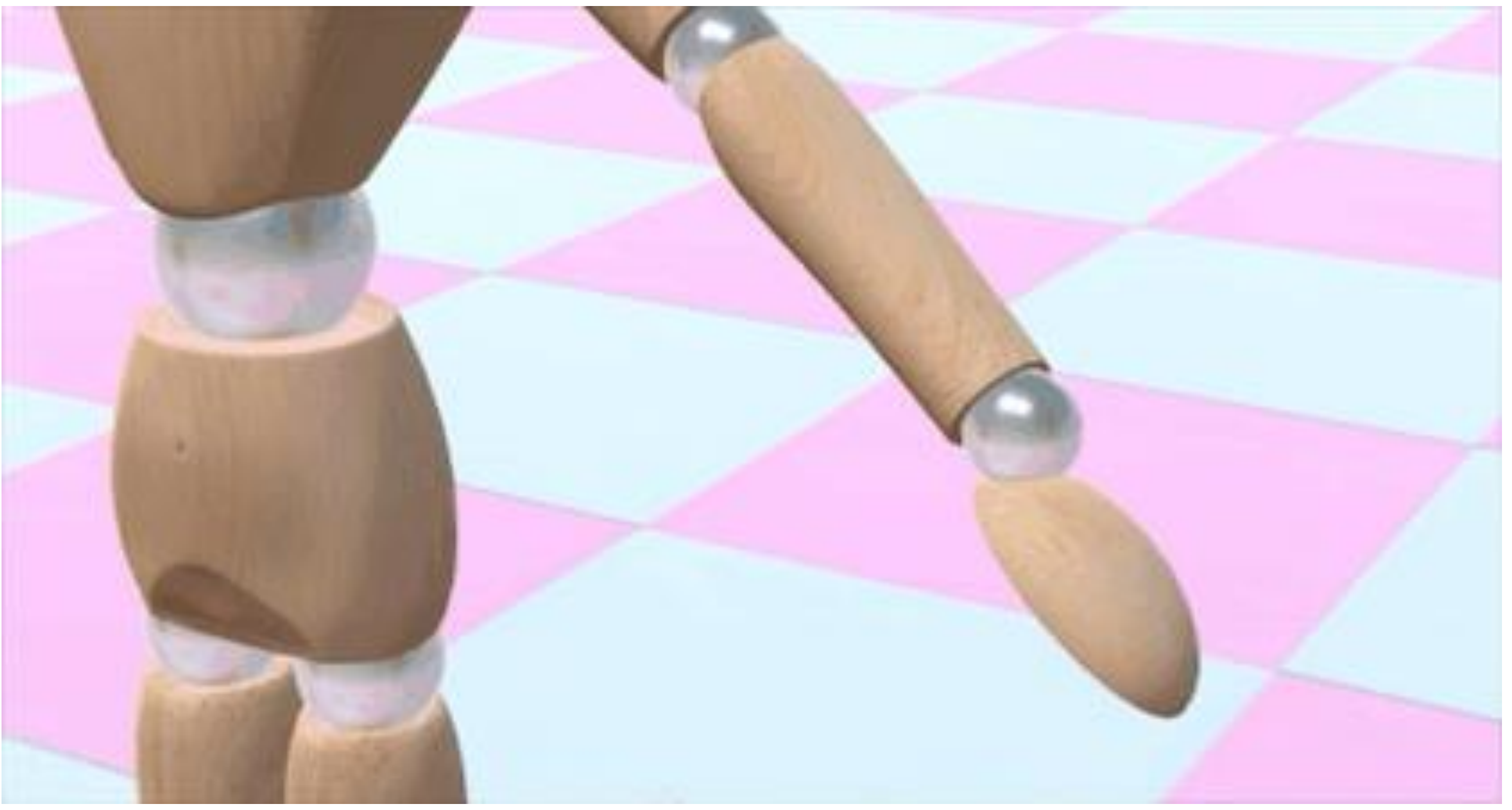
MotionNet

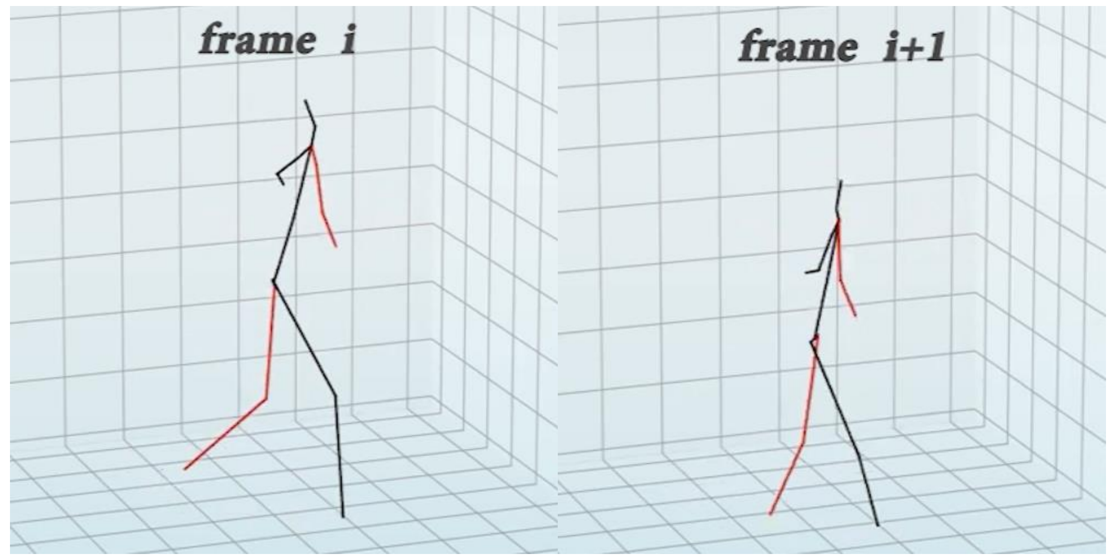
MotioNet: 3D Human Motion Reconstruction from Monocular Video with Skeleton Consistency

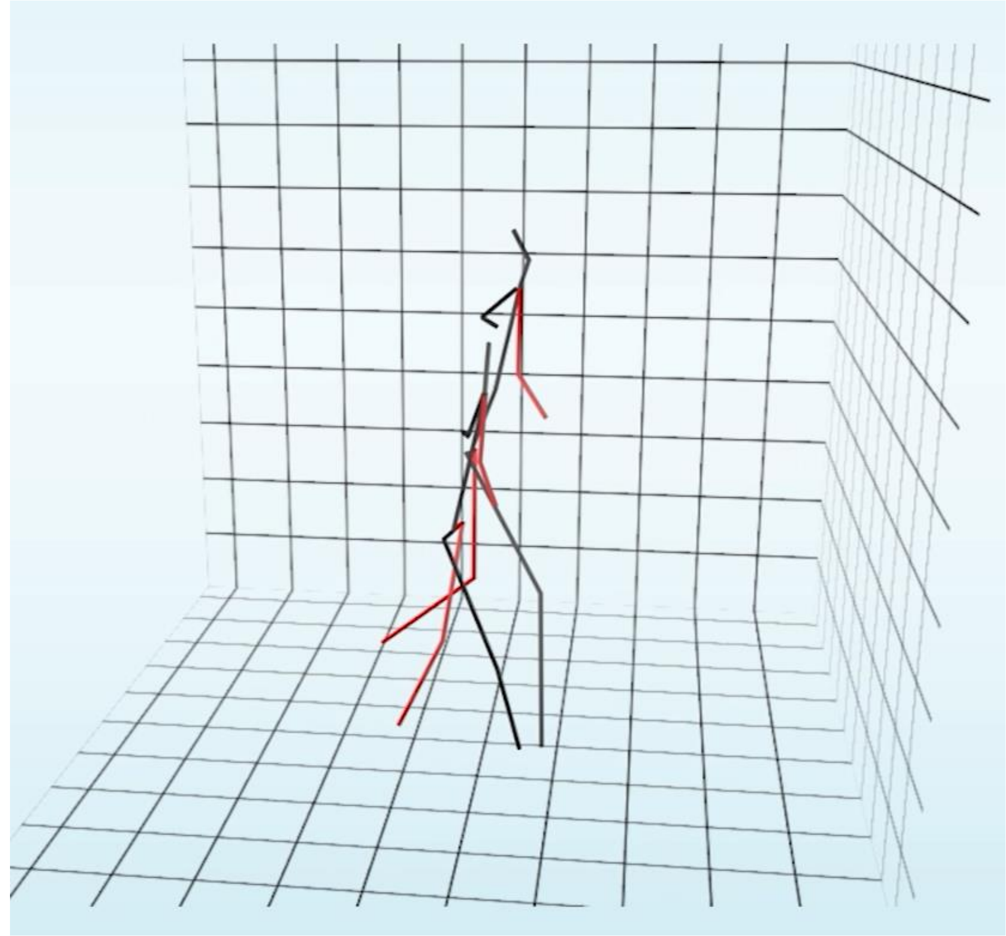
by M. Shi, K. Aberman, A. Aristidou, T. Komura, D. Lischinski, D. Cohen-Or, B. Chen

ACM Transactions on Graphics





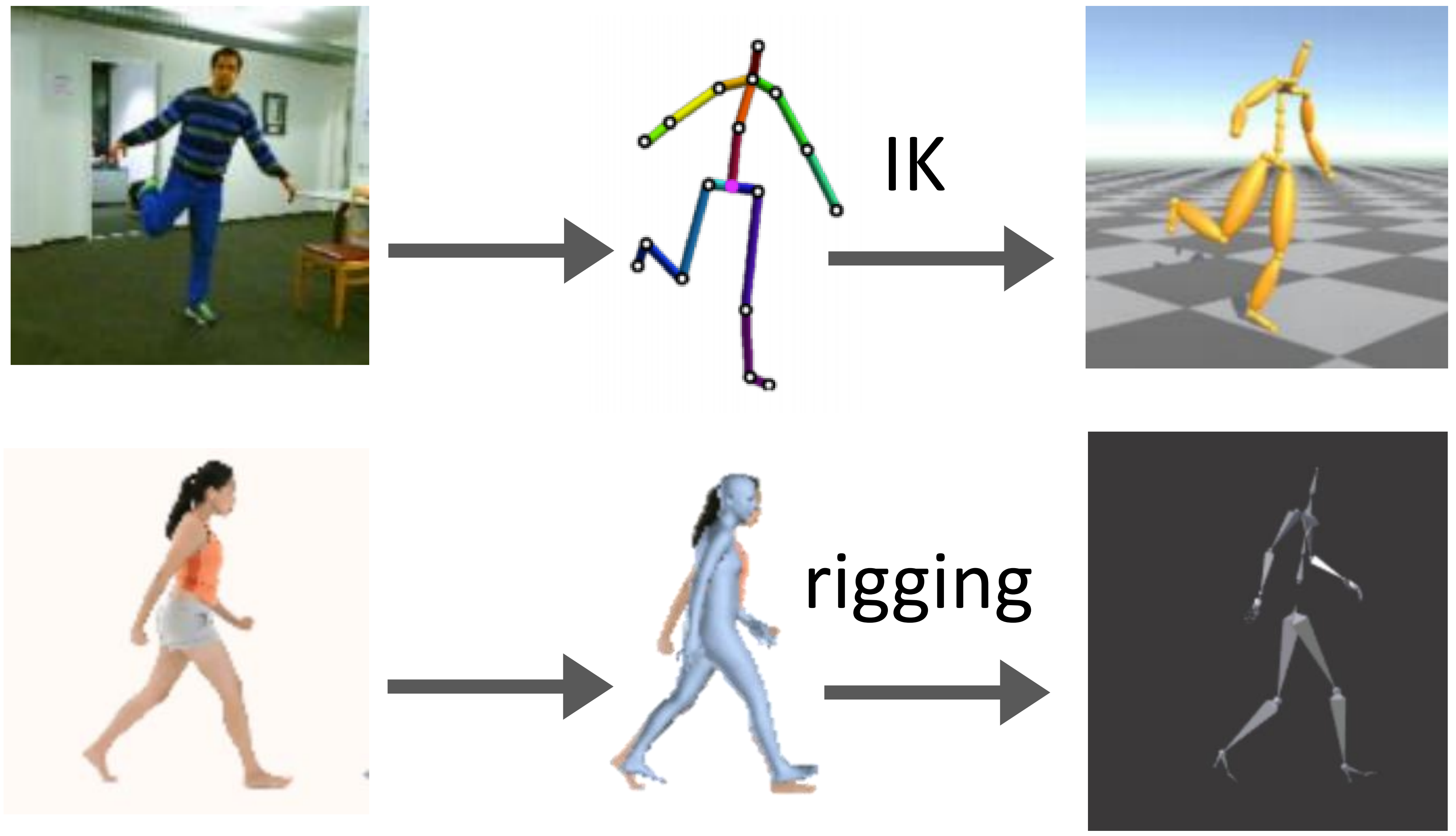




[Pavllo et al., CVPR 2019]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423



Use IK to convert the 3d position to rotation

[VNect, Mehta et al., SIGGRAPH 2017]

Apply rigging to make the rotation to a consistent skeleton

[HMR, Kanazawa et al., CVPR 2018]

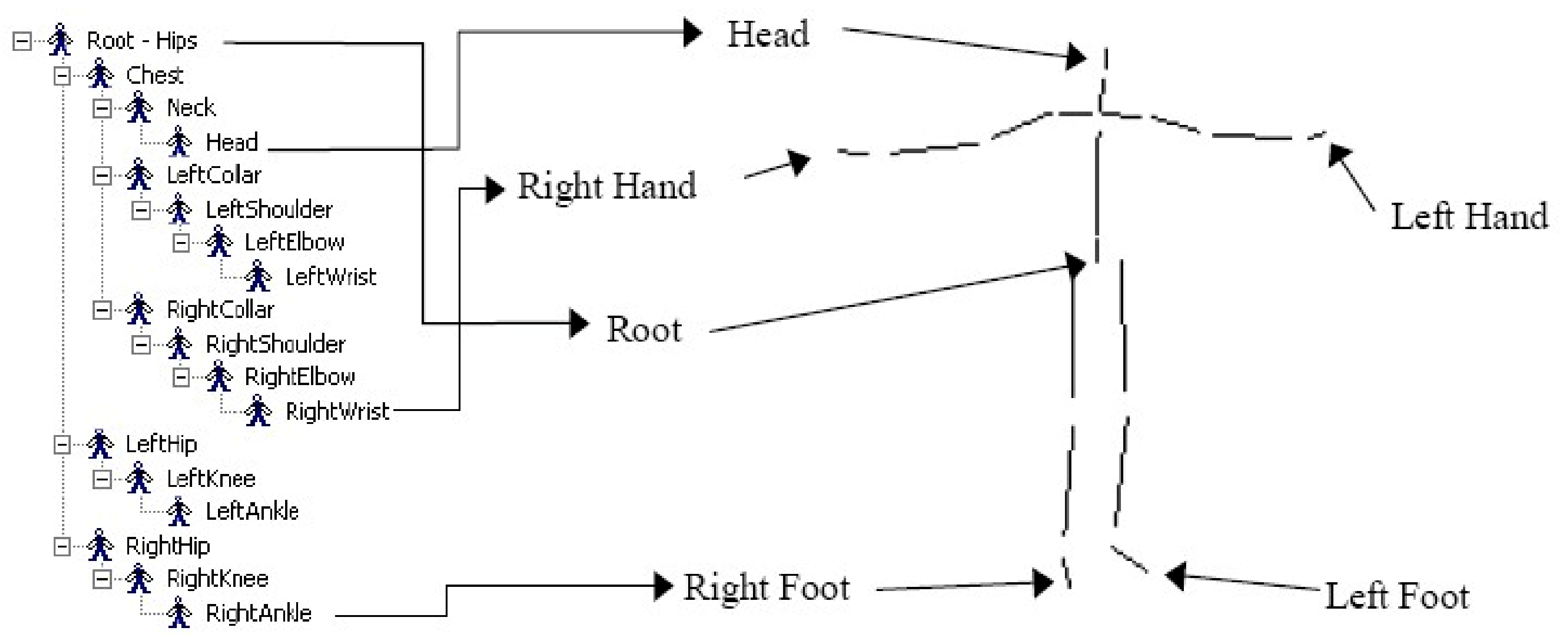
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The Key Idea:

What is the common representation of motion in MoCap datasets?

BVH - the most used output format of MoCap system



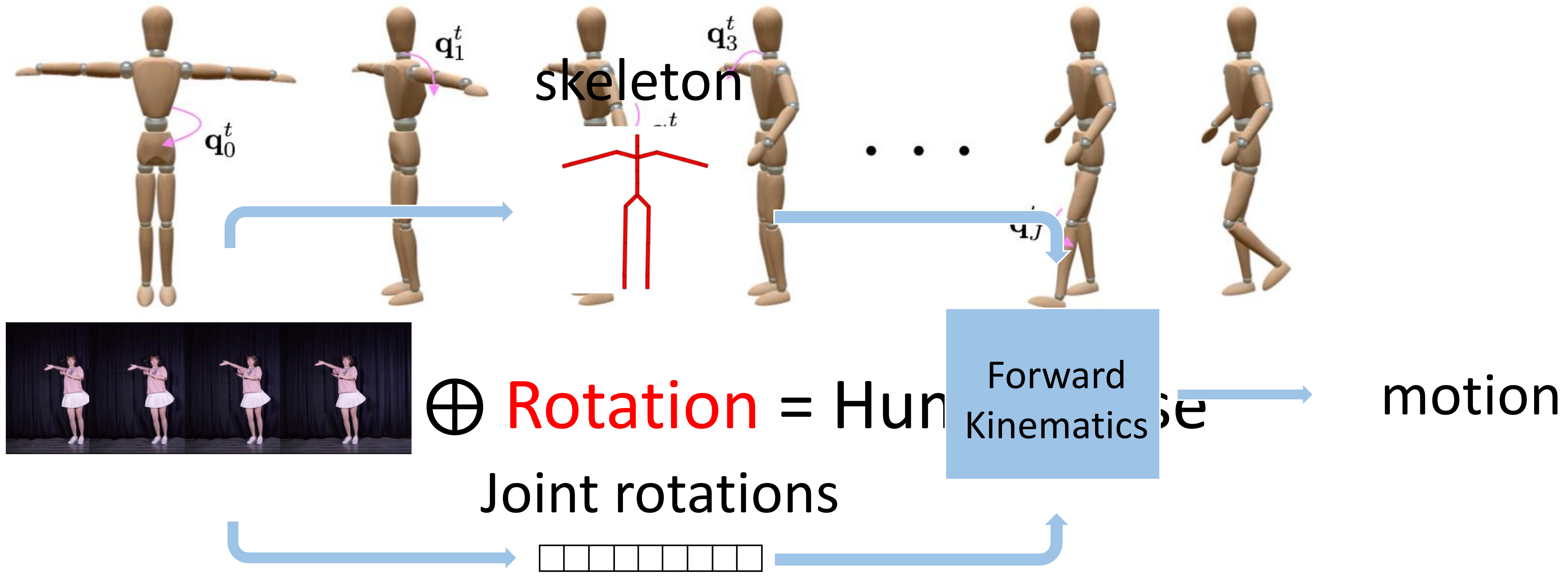
Initial pose with a hierarchical structure

```
Frames: 2
Frame Time:  0.04166667
-9.533684   4.447926    -0.566564   -7.757381   -1.735414   89.207932   9.763572
            6.289016    -1.825344   -6.106647   3.973667    -3.706973   -6.474916
            -14.391472  -3.461282   -16.504230  3.973544    -3.805107   22.204674
            2.533497    -28.283911  -6.862538   6.191492    4.448771    -16.292816
            2.951538    -3.418231   7.634442    11.325822   5.149696    -23.069189
            -18.352753  15.051558   -7.514462   8.397663    2.953842    -7.213992
            2.494318    -1.543435   2.970936    -25.086460  -4.195537   -1.752307
            7.093068    -1.507532   -2.633332   3.858087    0.256802    7.892136
            12.803010   -28.692566  2.151862    -9.164188   8.006427    -5.641034
            -12.596124  4.366460
```

Time-framed joint information(rotation)

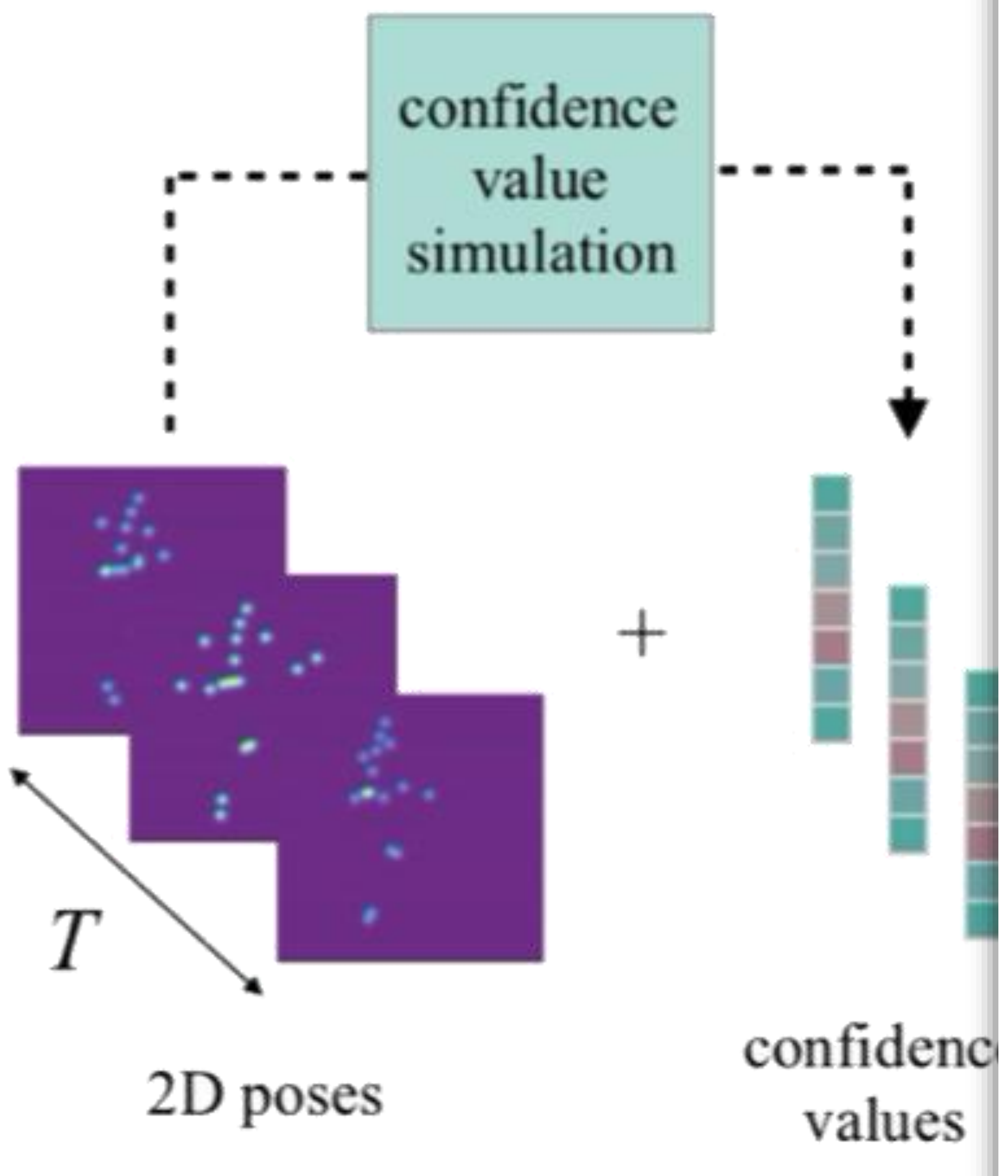
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423



⊕: Forward kinematics

MAI4CARE



MotioNet: 3D Human Motion Reconstruction from Monocular Video with Skeleton Consistency

MINGYI SHI, Shandong University, China, and AICFVE, Beijing Film Academy, China
KFIR ABERMAN, AICFVE, Beijing Film Academy, China, and Tel-Aviv University, Israel
ANDREAS ARISTIDOU, University of Cyprus and RISE Research Centre, Cyprus
TAKU KOMURA, Edinburgh University, Japan
DANI LISCHINSKI, Shandong University, China and The Hebrew University of Jerusalem, Israel and AICFVE, Beijing Film Academy, Israel
DANIEL COHEN-OR, Tel-Aviv University, Israel, and AICFVE, Beijing Film Academy, Israel
BAOQUAN CHEN, CFCS, Peking University, China, and AICFVE, Beijing Film Academy, China

Fig. 1. Given a monocular video of a performer, our approach, MotioNet, reconstructs a complete representation of the motion, consisting of a single symmetric skeleton, and a sequence of global root positions and 3D joint rotations. Thus, inverse kinematics is effectively integrated within the network and is data-driven rather than based on a universal prior. The images on the right were rendered from the output of our system after a simple rigging process.

We introduce *MotioNet*, a deep neural network that directly reconstructs the motion of a 3D human skeleton from a monocular video. While previous methods rely on either rigging or inverse kinematics (IK) to associate a consistent skeleton with temporally coherent joint rotations, our method is the first data-driven approach that directly outputs a kinematic skeleton, which is a complete, commonly used motion representation. At the crux of our approach lies a deep neural network with embedded kinematic priors, which decomposes sequences of 2D joint positions into two separate attributes: a single, symmetric skeleton encoded by bone lengths, and a sequence of 3D joint rotations associated with global root positions and foot contact labels. These attributes are fed into an integrated forward kinematics (FK) layer that outputs 3D positions, which are compared to a ground truth. In addition, an adversarial loss is applied to the velocities of the recovered rotations to ensure that they lie on the manifold of natural joint rotations. The key advantage of our approach is that it learns to infer natural joint rotations directly from the training data rather than assuming an underlying model, or inferring them from joint positions using a data-agnostic IK solver. We show that enforcing a single consistent skeleton along with temporally coherent joint rotations constrains the solution space, leading to a more robust handling of self-occlusions and depth ambiguities.

CCS Concepts: • **Computing methodologies** → **Motion processing**; **Neural networks**;

Additional Key Words and Phrases: Pose estimation, motion capturing, motion analysis

ACM Reference format:
Mingyi Shi, Kfir Aberman, Andreas Aristidou, Taku Komura, Dani Lischinski, Daniel Cohen-Or, and Baoquan Chen. 2020. MotioNet: 3D Human Motion Reconstruction from Monocular Video with Skeleton Consistency. *ACM Trans. Graph.* 40, 1, Article 1 (September 2020), 15 pages.
https://doi.org/10.1145/3407659

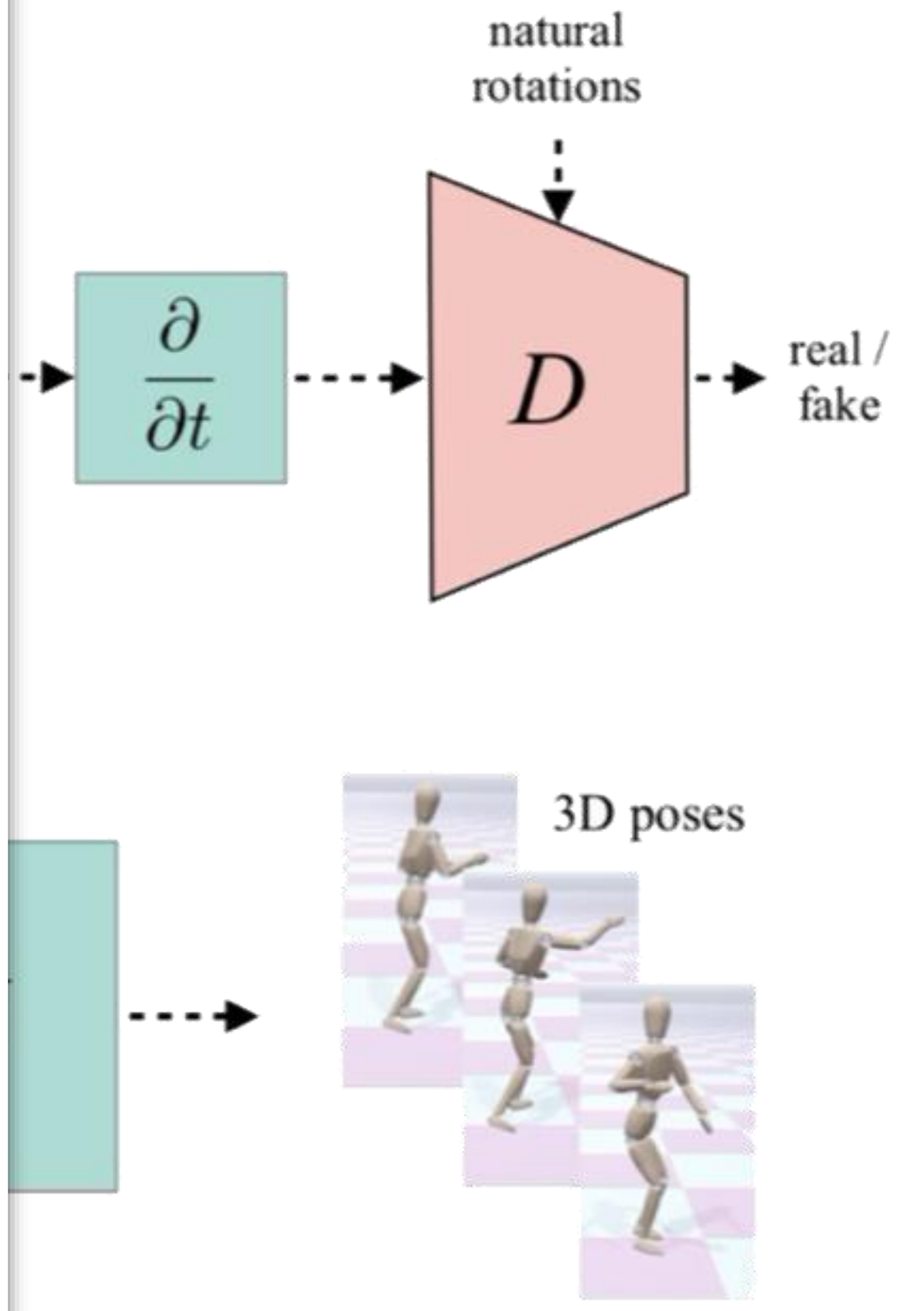
1 INTRODUCTION

Capturing the motion of humans has long been a fundamental task with a wide spectrum of applications in data-driven computer animation, special effects, gaming, activity recognition, and behavioral analysis. Motion is most accurately captured in a controlled setting using specialized hardware, such as magnetic

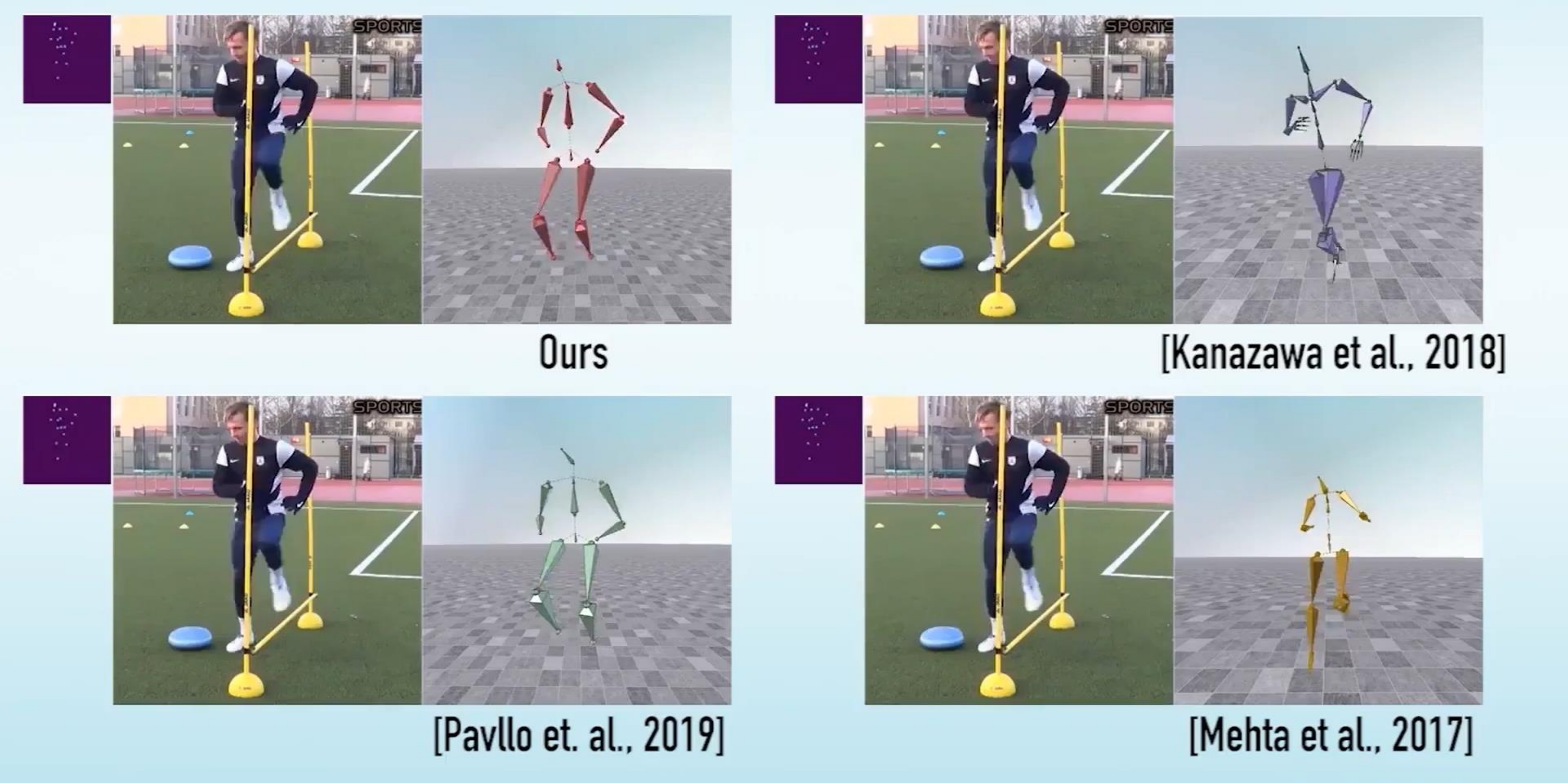
This work was supported in part by the National Key R&D Program of China (2018YFB1403900, 2019YFF0302902), the Israel Science Foundation (grant no. 2366/16), and by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739578 and the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

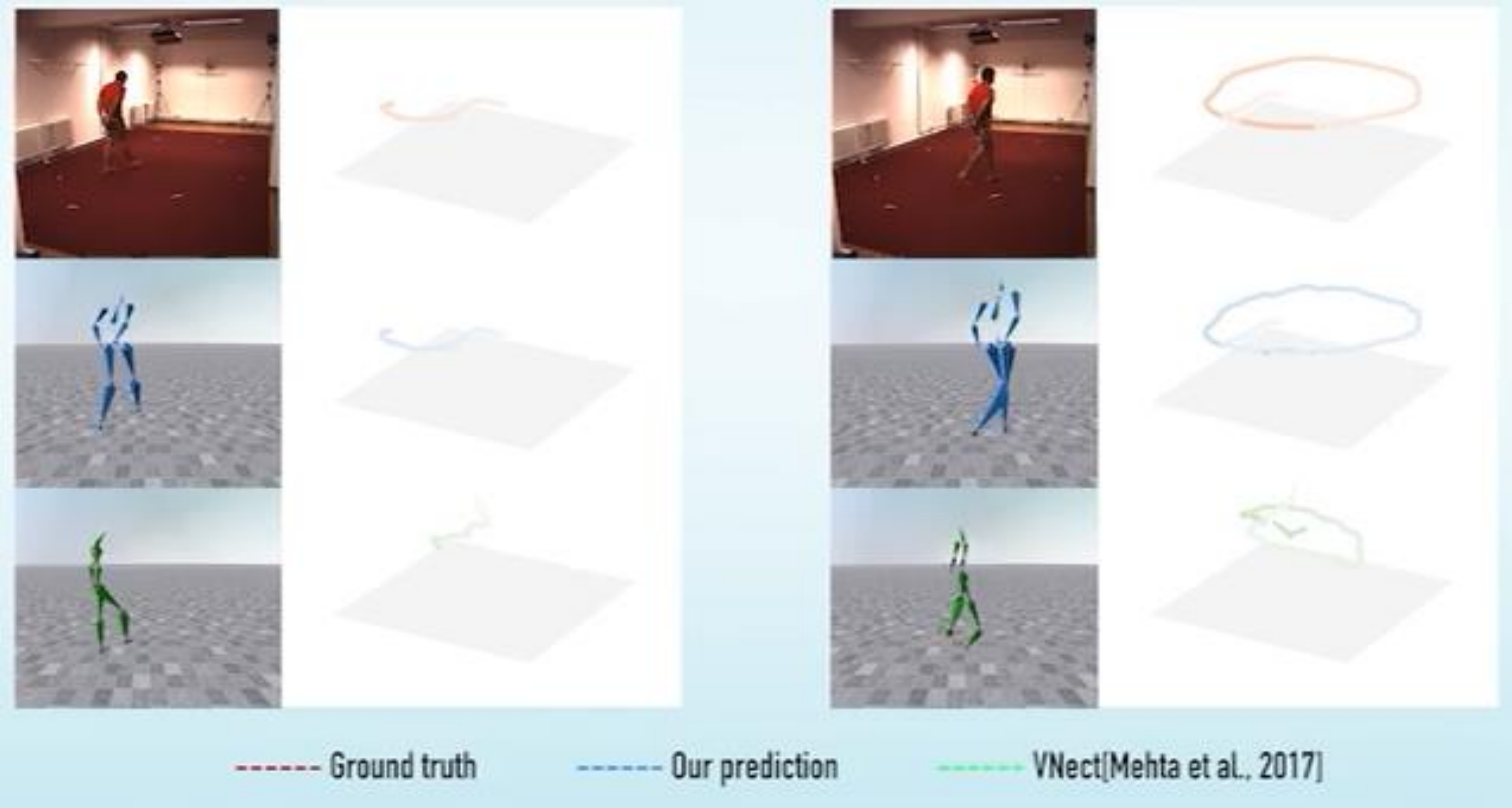
Authors' addresses: M. Shi, Shandong University, AICFVE, Beijing Film Academy; email: irubbly@gmail.com; K. Aberman, AICFVE, Beijing Film Academy, Tel-Aviv University; A. Aristidou, University of Cyprus, 75, Kallipoleos, Nicosia, Cyprus, 1678, RISE Research Centre; email: a.aristidou@ieee.org; T. Komura, Edinburgh University; D. Lischinski, Shandong University, The Hebrew University of Jerusalem, AICFVE, Beijing Film Academy; D. Cohen-Or, Tel-Aviv University, AICFVE, Beijing Film Academy; B. Chen (corresponding author), CFCS, Peking University, AICFVE, Beijing Film Academy; email: baoquan@pku.edu.cn.

0730-0301/2020/09-ART1 $15.00
https://doi.org/10.1145/3407659

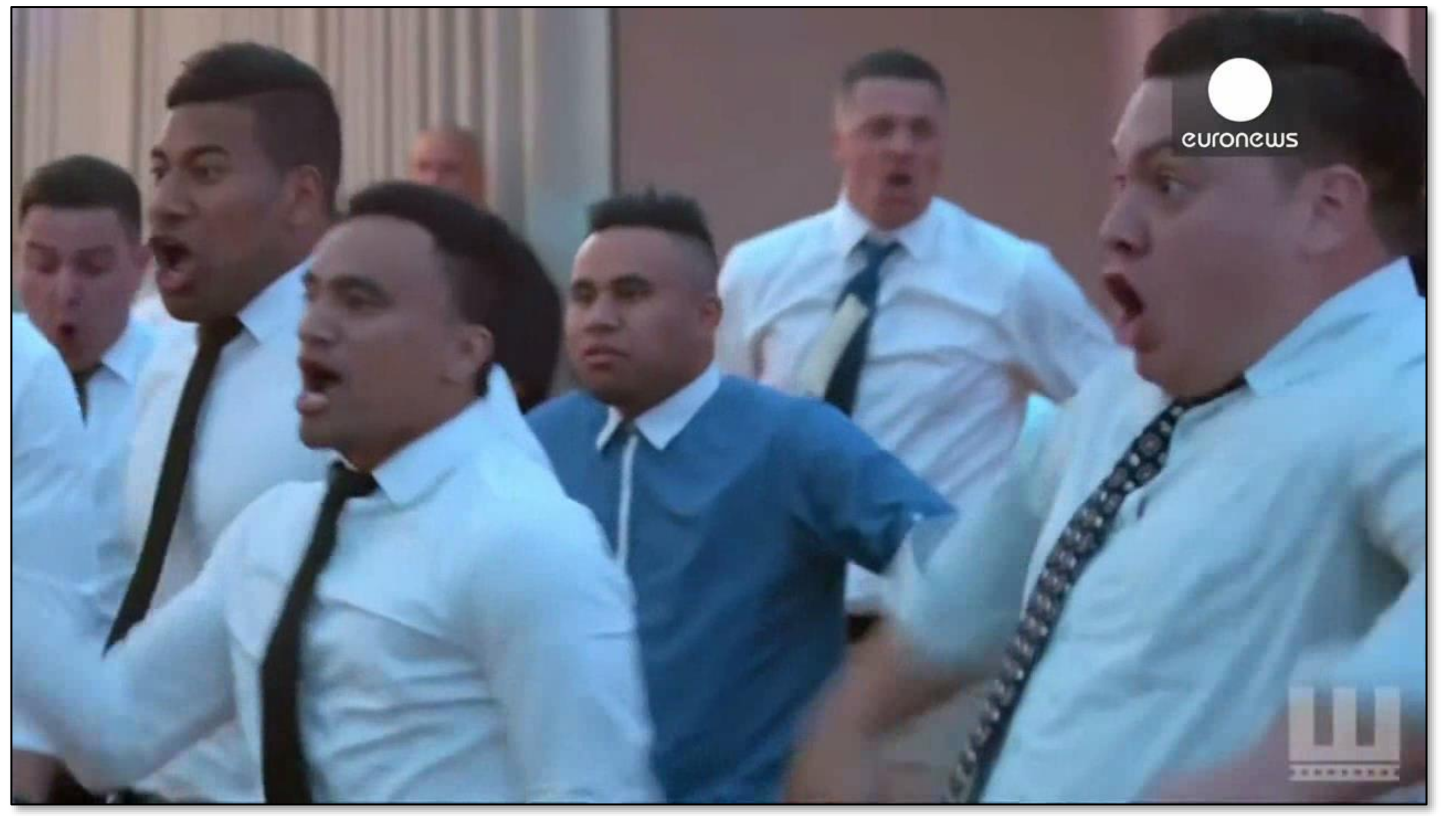


Results


Ours
[Kanazawa et al., 2018]
[Pavllo et. al., 2019]
[Mehta et al., 2017]


Ground truth
Our prediction
VNect[Mehta et al., 2017]

Emotion and Style



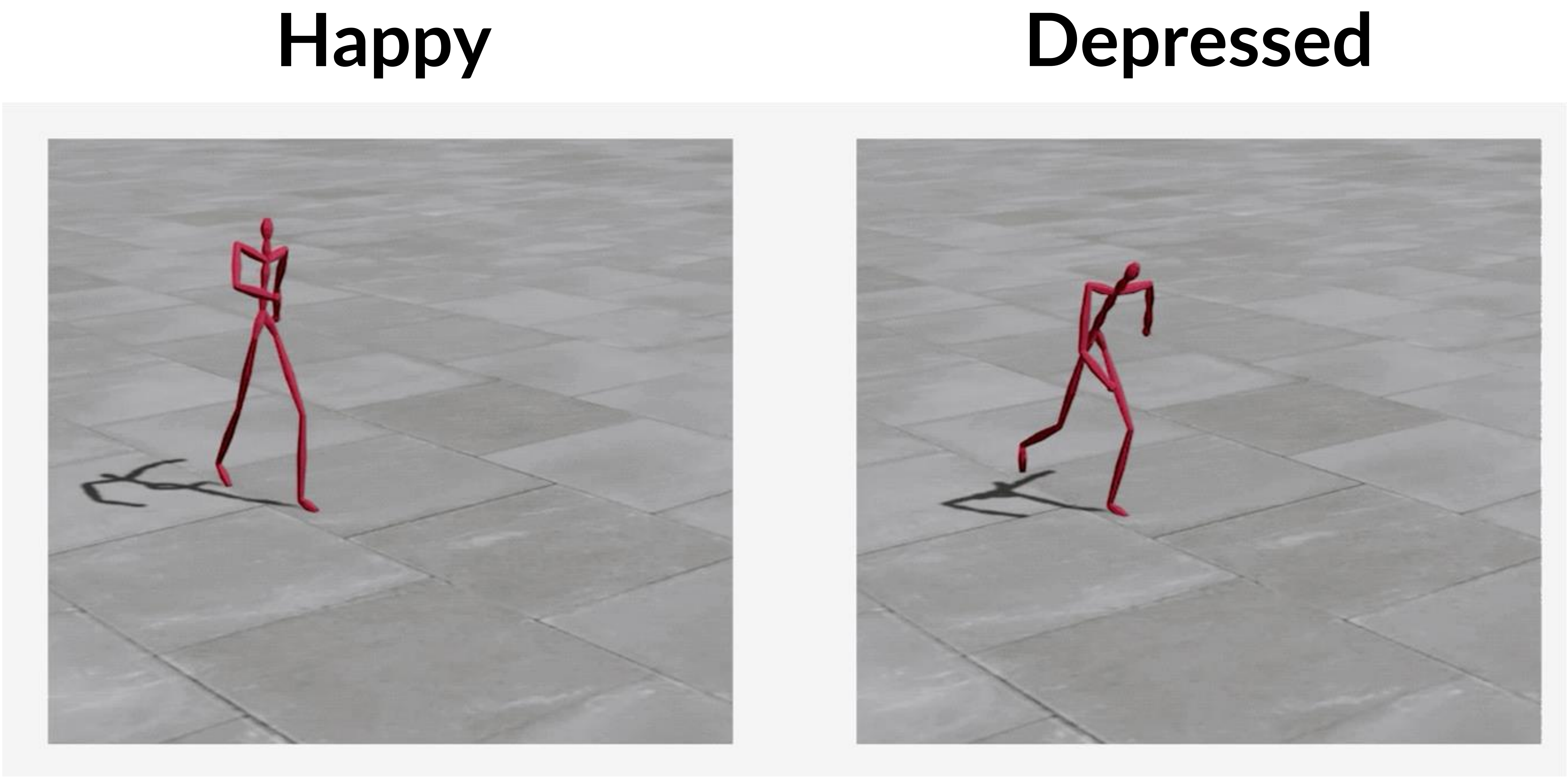
Maori wedding (Haka)
https://youtu.be/QUbx-AcDgXo

Emotions through dance
https://youtu.be/m0R-ftFBm38

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Human Motion Style



Style is an abstract attribute

Related Work

Model **style**, which is **not well-defined**, as some **hand-crafted representations**, such as,

- Difference in spectral domain.
- Physical parameters of human body.
- Low-level features based on the LMA theories on human analysis

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Related Work

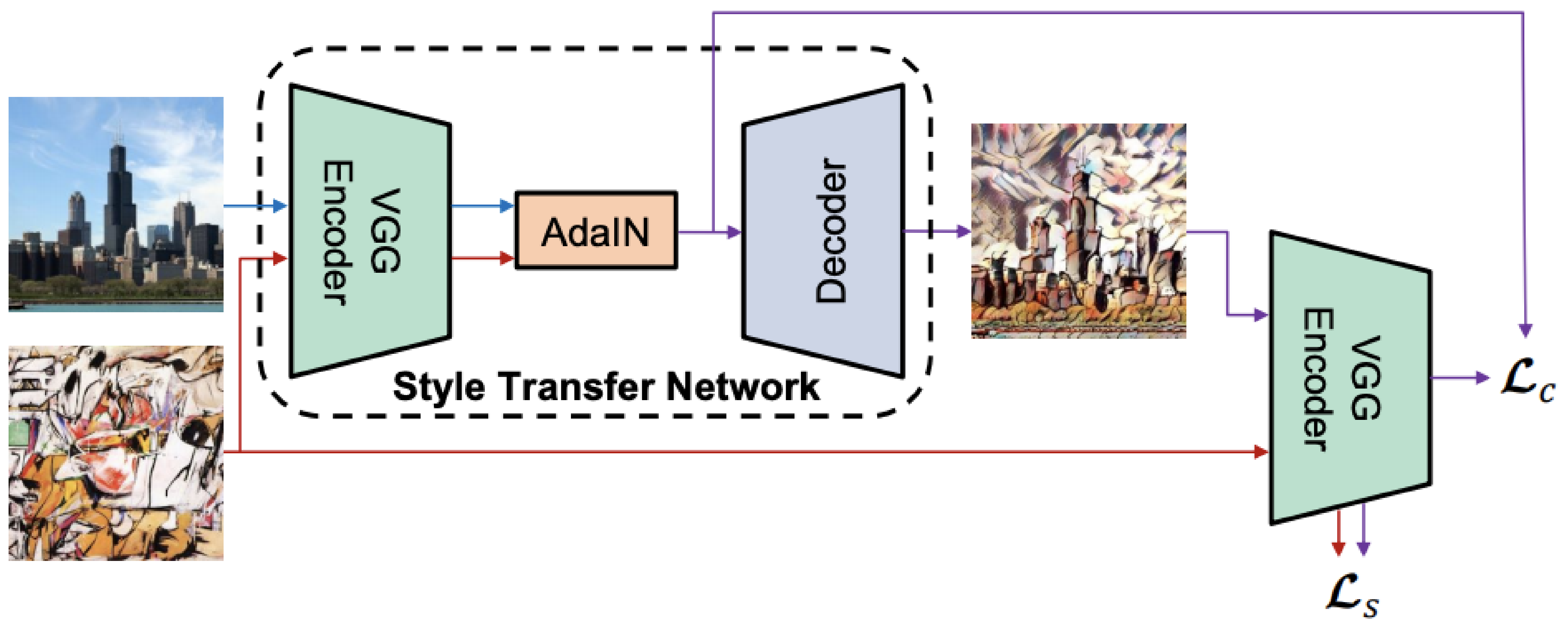
Data Driven

Learn the mapping based on **labeled & paired motion data**.

- limited to **structurally similar** motions in the dataset
- limited to a **pre-defined set** of styles in mocap data
- limited to **style recorded by MoCap** systems

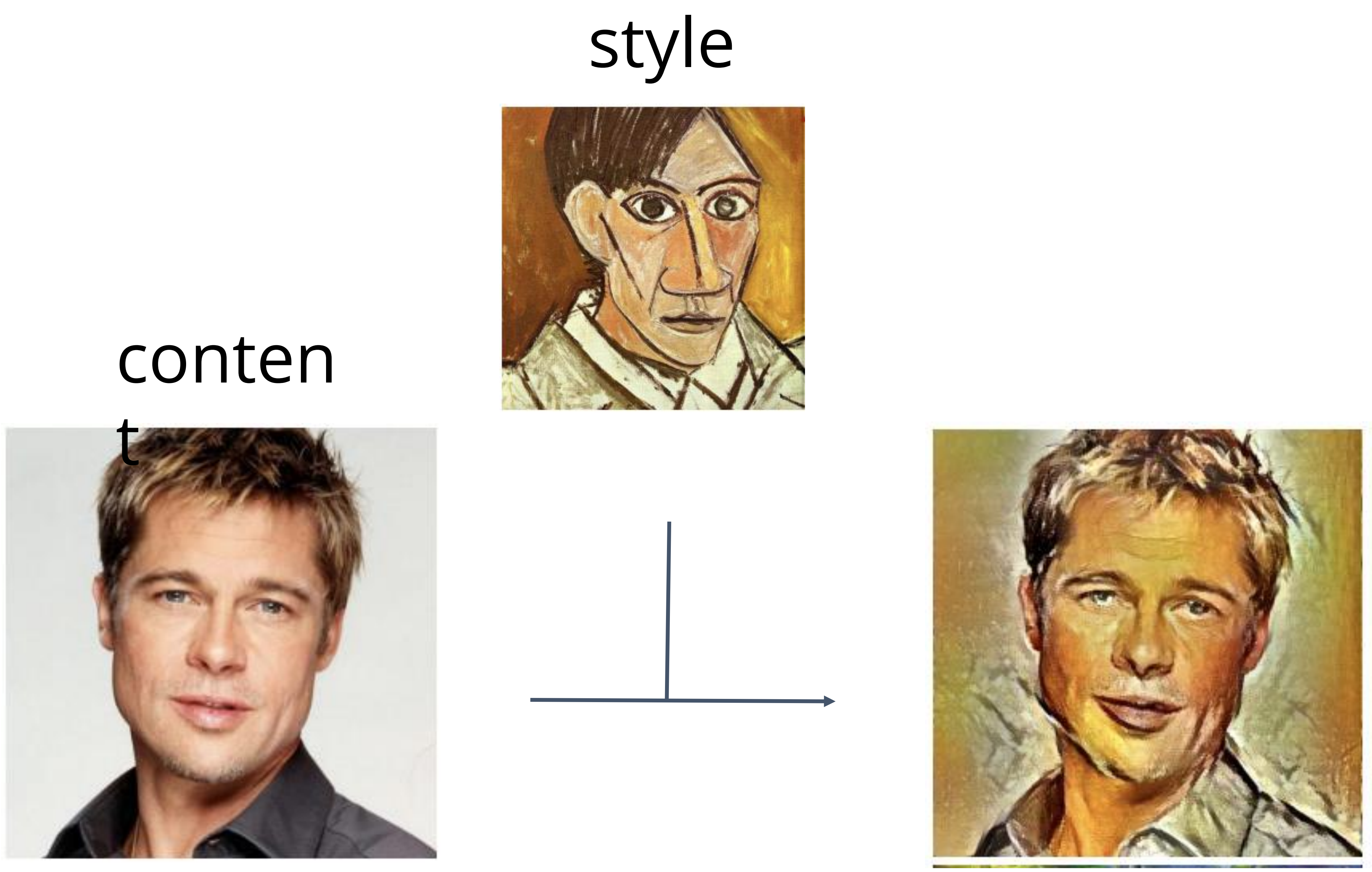
Inspiration from Image Style Transfer

Adaptive Instance Normalization **(AdaIN) layer -** spatially invariant, maintains geometry, manipulates style.



$$\mathrm{AdaIN}(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y)$$

[Huang and Belongie, 2017]

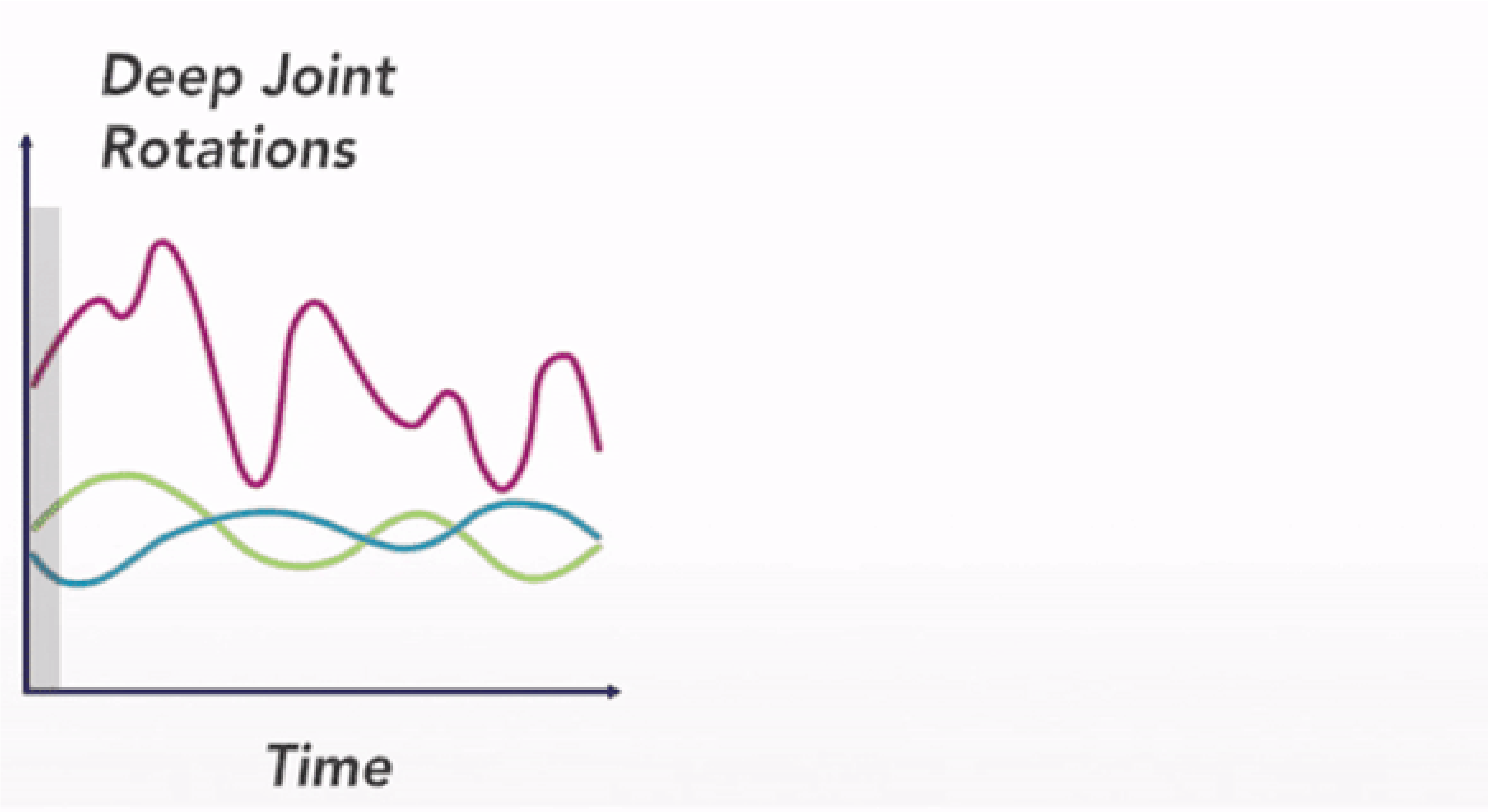


Geometry and shapes are preserved

Adapting to motion?

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

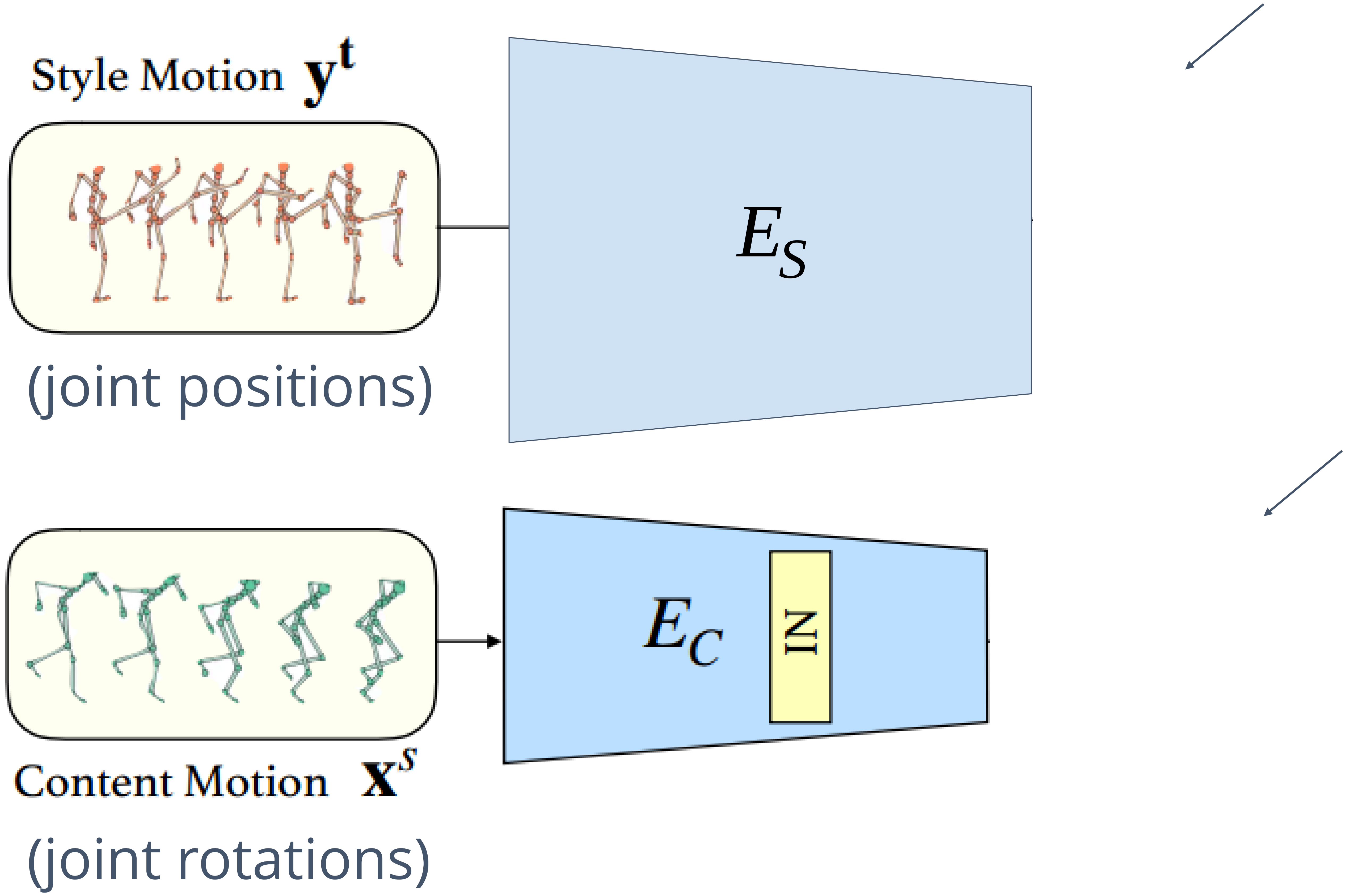
Adaptive Instance Normalization (AdaIN)



Shape of signal is preserved

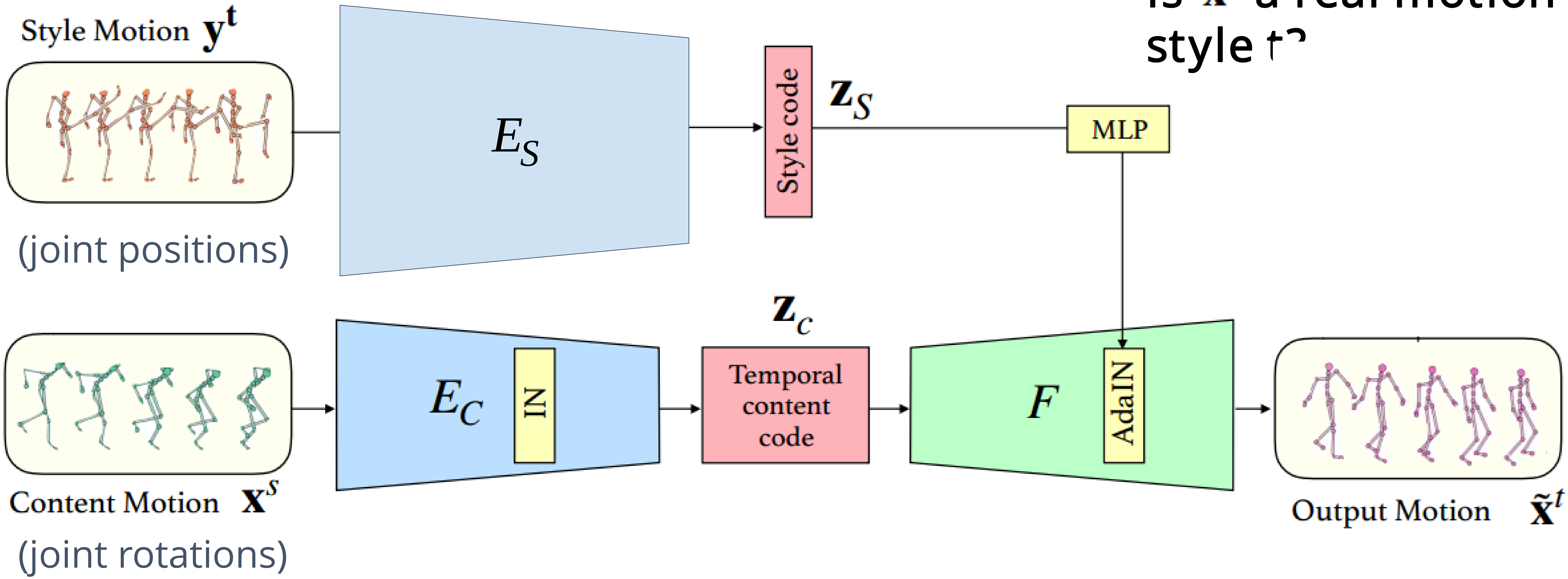
Architecture

Temporally-invariant

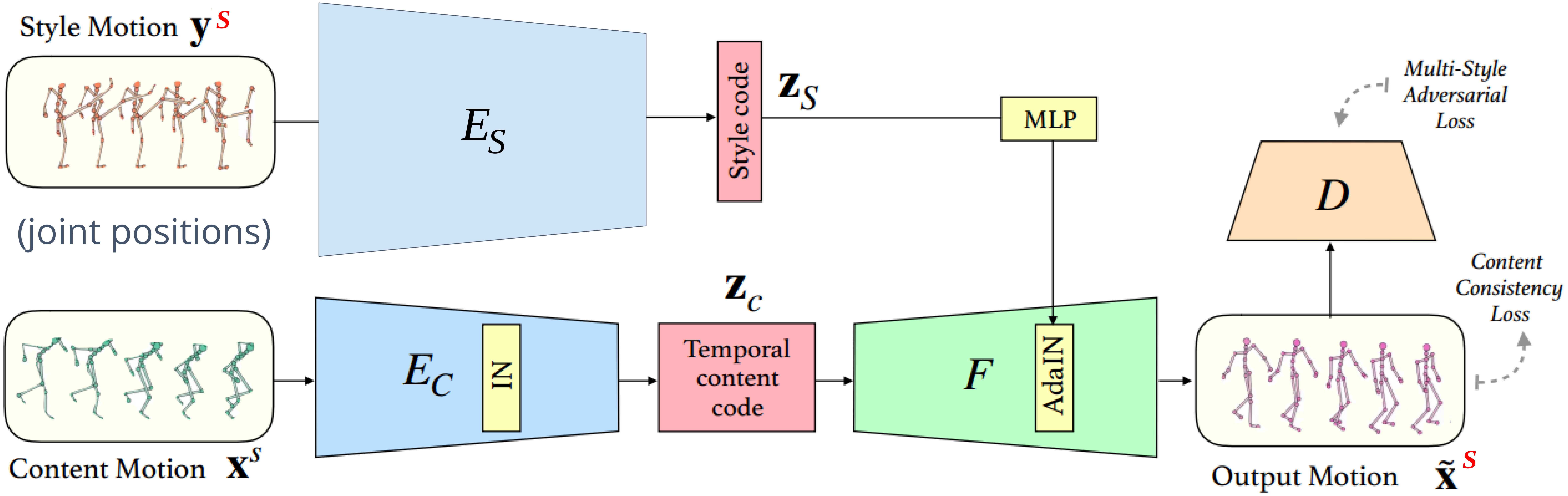


Loss Terms

Is $\tilde{\mathbf{x}}^t$ a real motion of style t?



Loss Terms



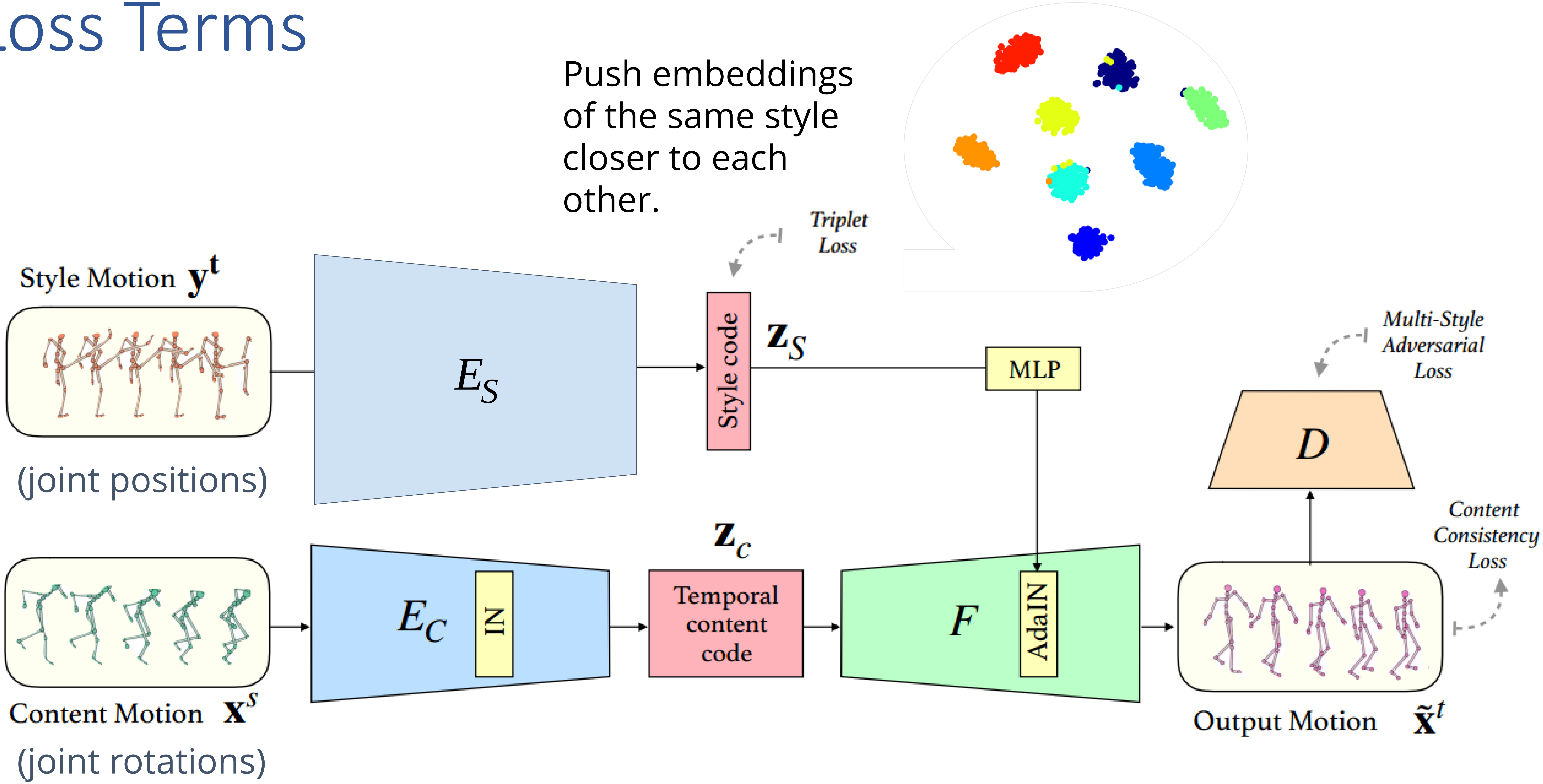
(joint positions)

(joint rotations)

$\tilde{\mathbf{x}}^S$ reconstructs $\mathbf{x}^S$?

Loss Terms

Push embeddings of the same style closer to each other.



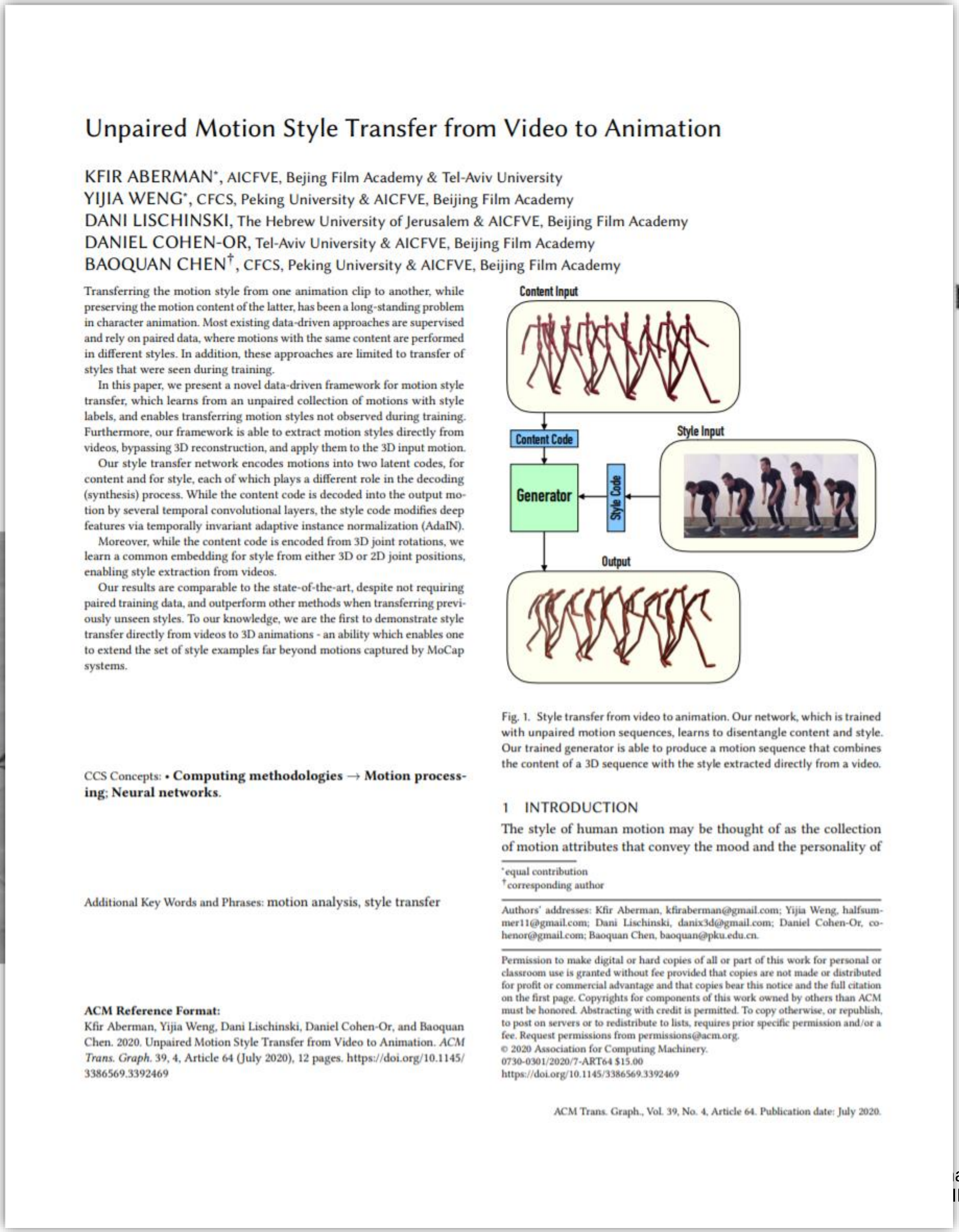
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results

Style Input (proud)

Content Input



Unpaired Motion Style Transfer from Video to Animation

KFIR ABERMAN*, AICFVE, Bejing Film Academy & Tel-Aviv University
YIJIA WENG*, CFCS, Peking University & AICFVE, Beijing Film Academy
DANI LISCHINSKI, The Hebrew University of Jerusalem & AICFVE, Beijing Film Academy
DANIEL COHEN-OR, Tel-Aviv University & AICFVE, Beijing Film Academy
BAOQUAN CHEN†, CFCS, Peking University & AICFVE, Beijing Film Academy

Transferring the motion style from one animation clip to another, while preserving the motion content of the latter, has been a long-standing problem in character animation. Most existing data-driven approaches are supervised and rely on paired data, where motions with the same content are performed in different styles. In addition, these approaches are limited to transfer of styles that were seen during training.

In this paper, we present a novel data-driven framework for motion style transfer, which learns from an unpaired collection of motions with style labels, and enables transferring motion styles not observed during training. Furthermore, our framework is able to extract motion styles directly from videos, bypassing 3D reconstruction, and apply them to the 3D input motion.

Our style transfer network encodes motions into two latent codes, for content and for style, each of which plays a different role in the decoding (synthesis) process. While the content code is decoded into the output motion by several temporal convolutional layers, the style code modifies deep features via temporally invariant adaptive instance normalization (AdaIN).

Moreover, while the content code is encoded from 3D joint rotations, we learn a common embedding for style from either 3D or 2D joint positions, enabling style extraction from videos.

Our results are comparable to the state-of-the-art, despite not requiring paired training data, and outperform other methods when transferring previously unseen styles. To our knowledge, we are the first to demonstrate style transfer directly from videos to 3D animations - an ability which enables one to extend the set of style examples far beyond motions captured by MoCap systems.

Fig. 1. Style transfer from video to animation. Our network, which is trained with unpaired motion sequences, learns to disentangle content and style. Our trained generator is able to produce a motion sequence that combines the content of a 3D sequence with the style extracted directly from a video.

CCS Concepts: • **Computing methodologies → Motion processing; Neural networks.**

Additional Key Words and Phrases: motion analysis, style transfer

ACM Reference Format:
Kfir Aberman, Yijia Weng, Dani Lischinski, Daniel Cohen-Or, and Baoquan Chen. 2020. Unpaired Motion Style Transfer from Video to Animation. *ACM Trans. Graph.* 39, 4, Article 64 (July 2020), 12 pages. https://doi.org/10.1145/3386569.3392469

1 INTRODUCTION

The style of human motion may be thought of as the collection of motion attributes that convey the mood and the personality of

*equal contribution
†corresponding author

Authors' addresses: Kfir Aberman, kfiraberman@gmail.com; Yijia Weng, halfsummer11@gmail.com; Dani Lischinski, danix3d@gmail.com; Daniel Cohen-Or, cohenor@gmail.com; Baoquan Chen, baoquan@pku.edu.cn.

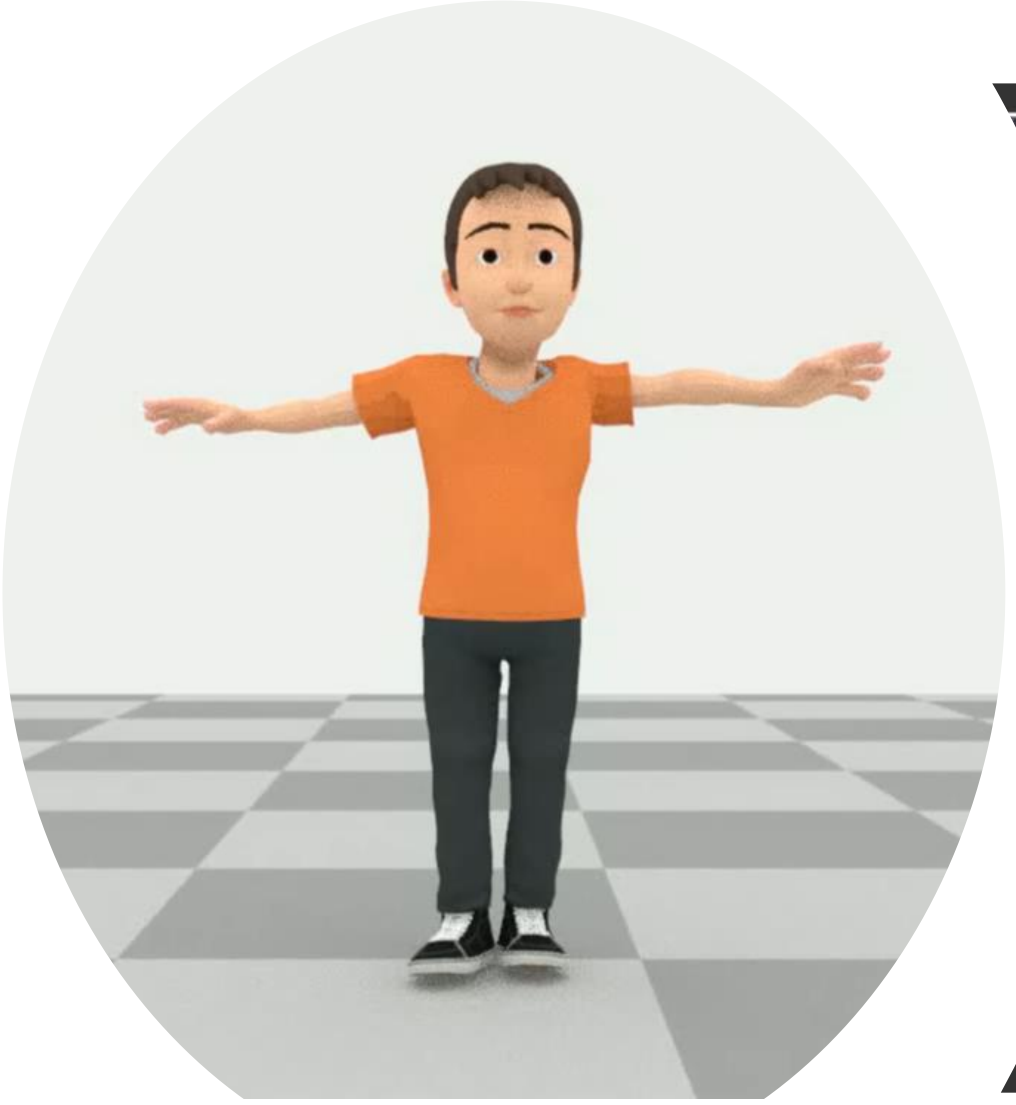
0730-0301/2020/7-ART64 $15.00
https://doi.org/10.1145/3386569.3392469

ry)

Output

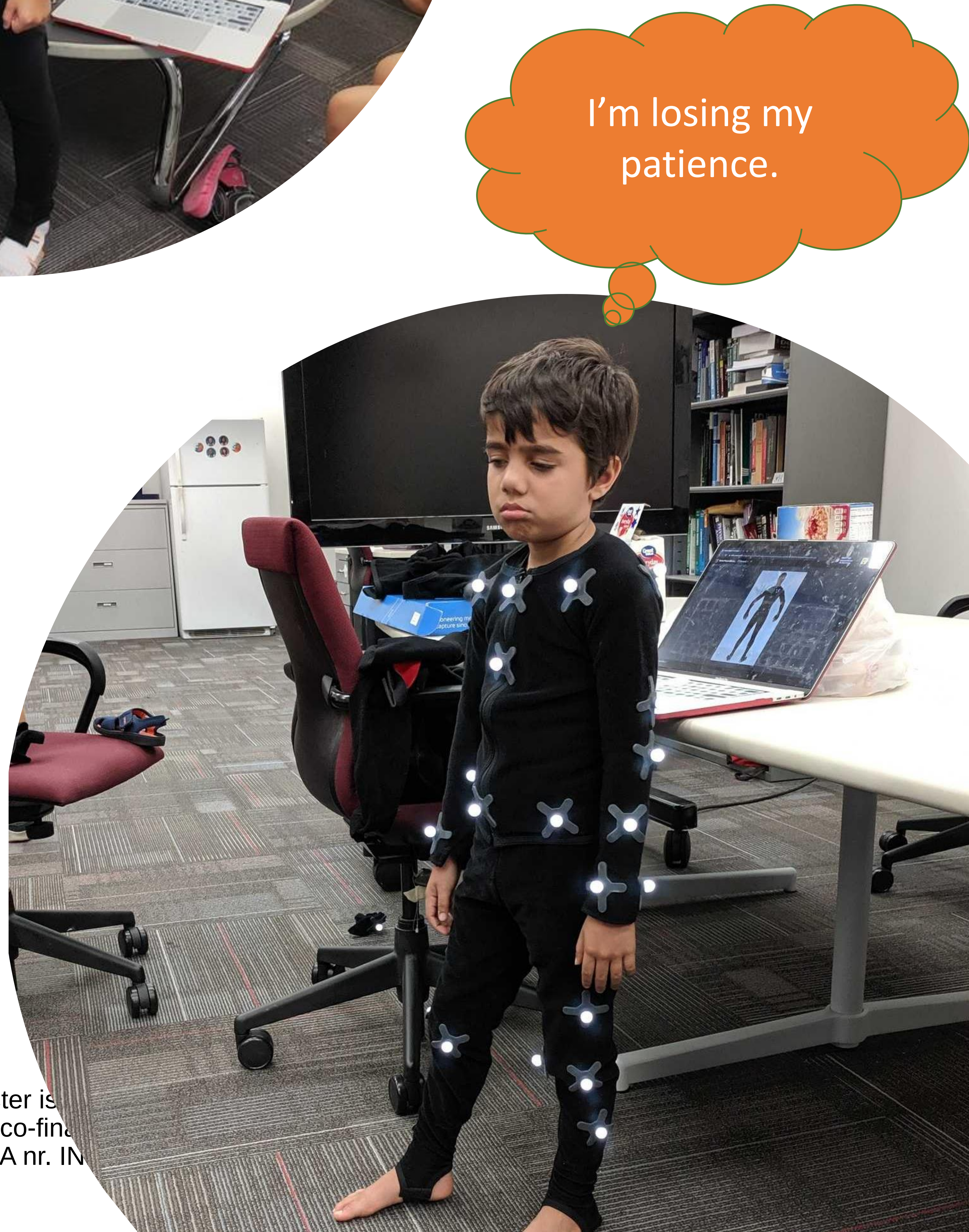
Challenges

- Data are not always available…



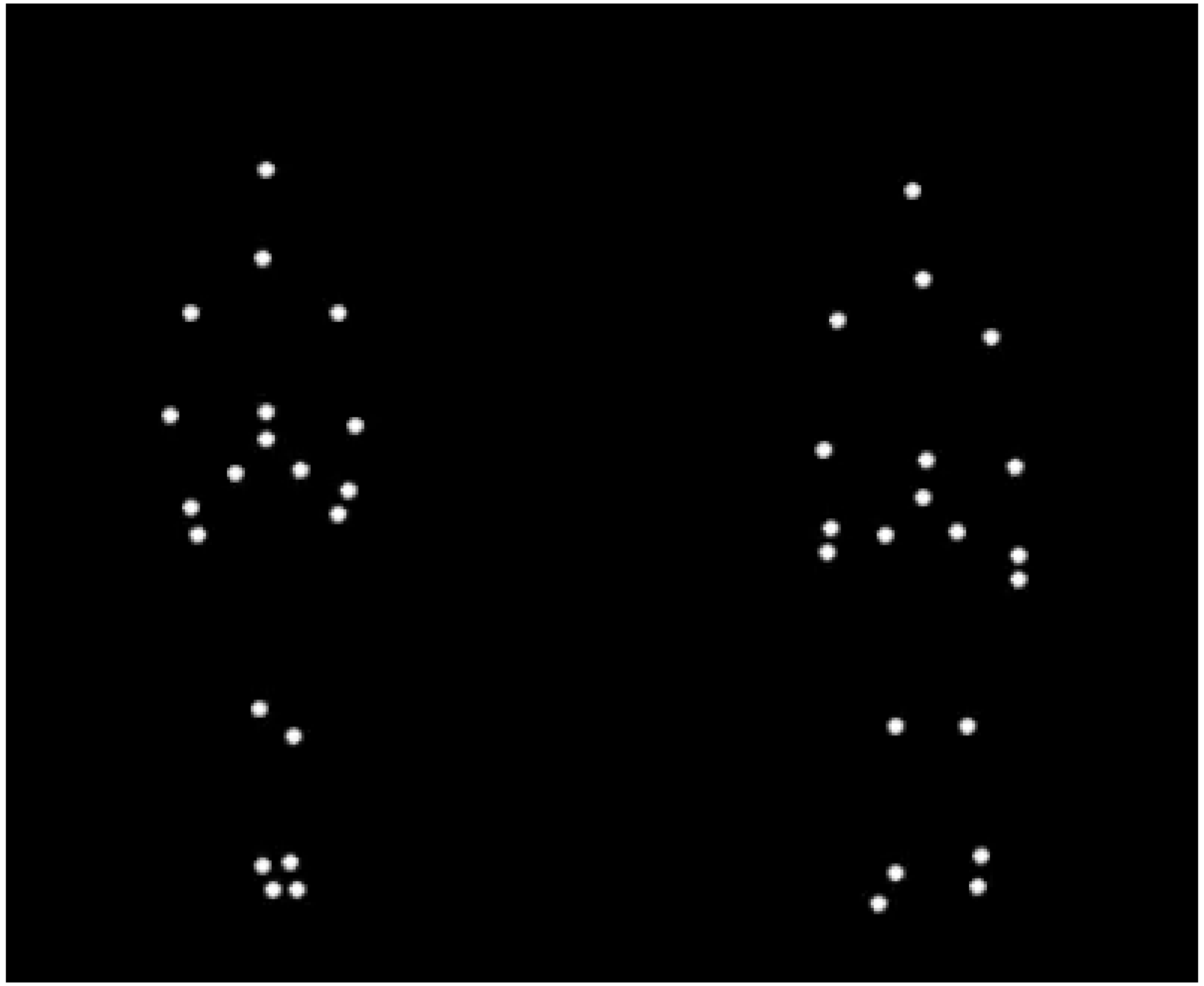
Russell from movie "Up"

Bonnie from "Toy Story 4"



Challenges

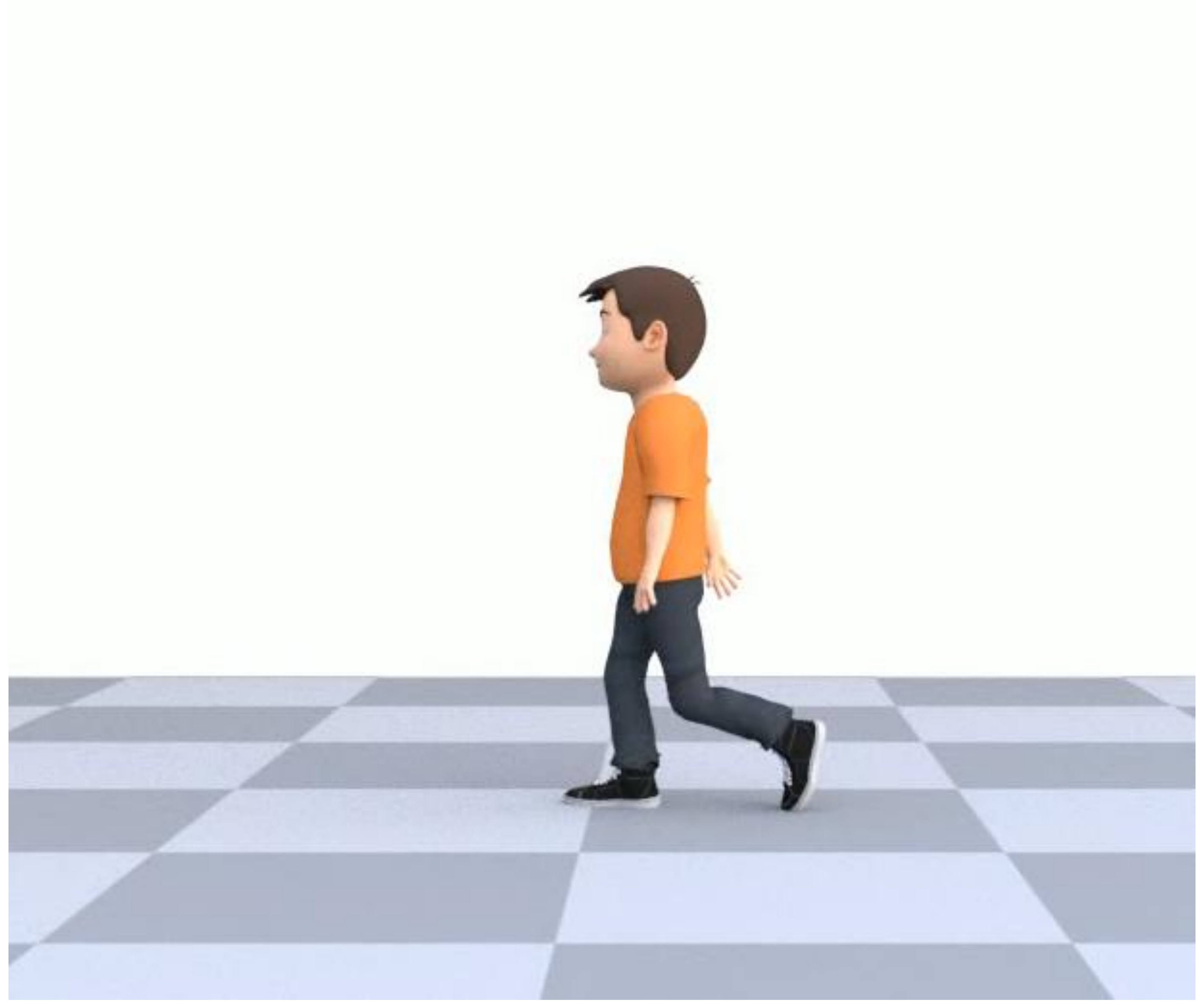
- Given the difficulties of motion capturing children, can we just use adult mocap data on child characters?
- Can we convince the viewers that the motions are from children?
- Jain et al[2016] found that viewers can differentiate child motion from adult motion by viewing point light display videos.



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Key Ideas

- Adapt adult motions to child motions that captures both the postures and the timing of child motions.
- Achieve this goal without temporally aligned data given that adult motions and child motions can be drastically different.



Overall Architecture

Adversarial loss

$$\mathcal{L}_{G_{c2a}} = 0.5 * \mathbb{E}_{c \sim p(c)}[D_a(G_{c2a}(c)) - 1]$$
$$\mathcal{L}_{G_{a2c}} = 0.5 * \mathbb{E}_{a \sim p(a)}[D_c(G_{a2c}(a)) - 1]$$

Cycle loss

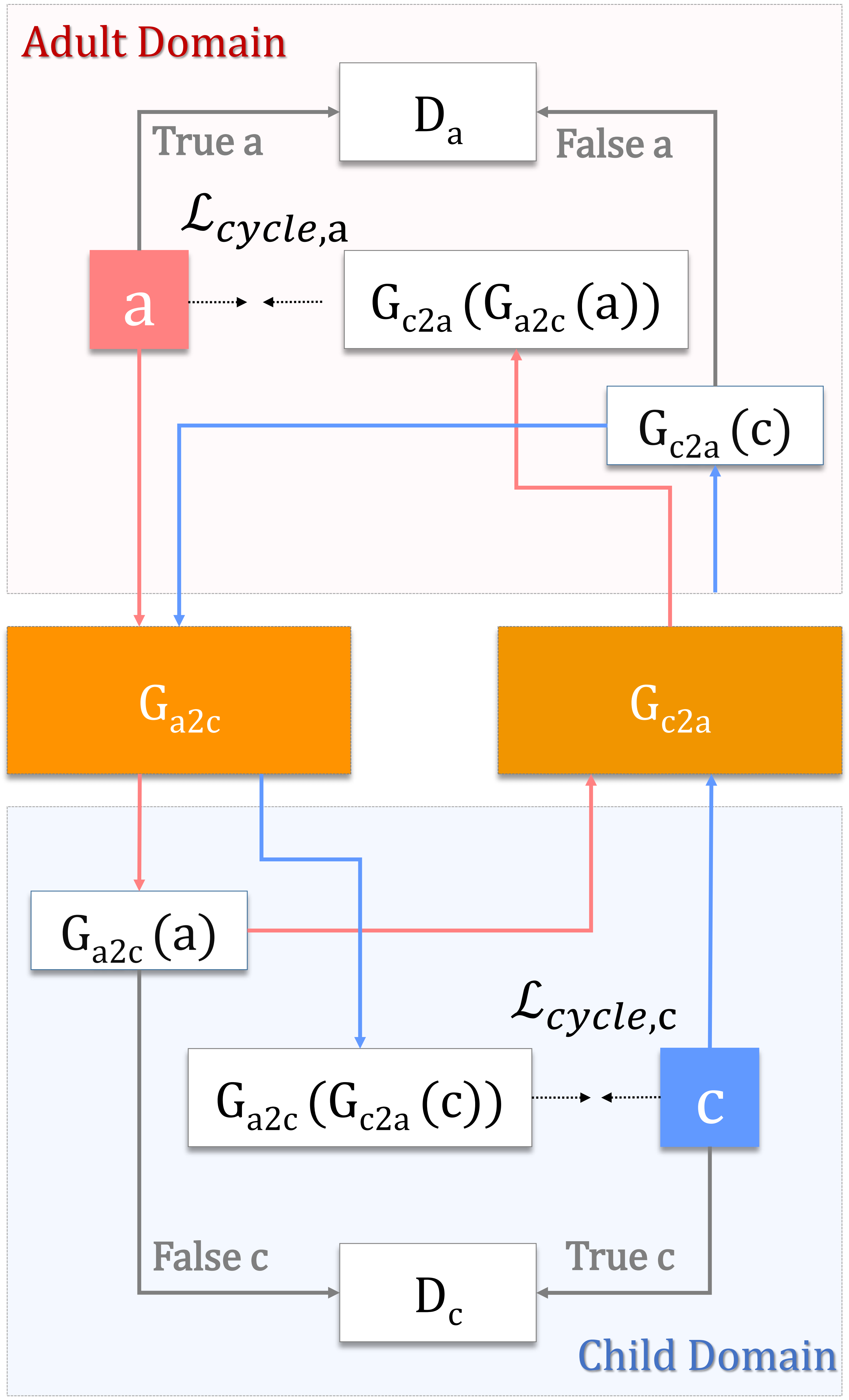
$$\mathcal{L}_{cycle,c} = G_{a2c}(G_{c2a}(c)) - c$$
$$\mathcal{L}_{cycle,a} = G_{c2a}(G_{a2c}(a)) - a$$

Coherence loss

$$\mathcal{L}_{coherence,a} = \sum_t \sum_{DOF} ||G_{a2c}(a)(t) - G_{a2c}(a)(t-1)||$$
$$\mathcal{L}_{coherence,c} = \sum_t \sum_{DOF} ||G_{c2a}(c)(t) - G_{c2a}(c)(t-1)||$$

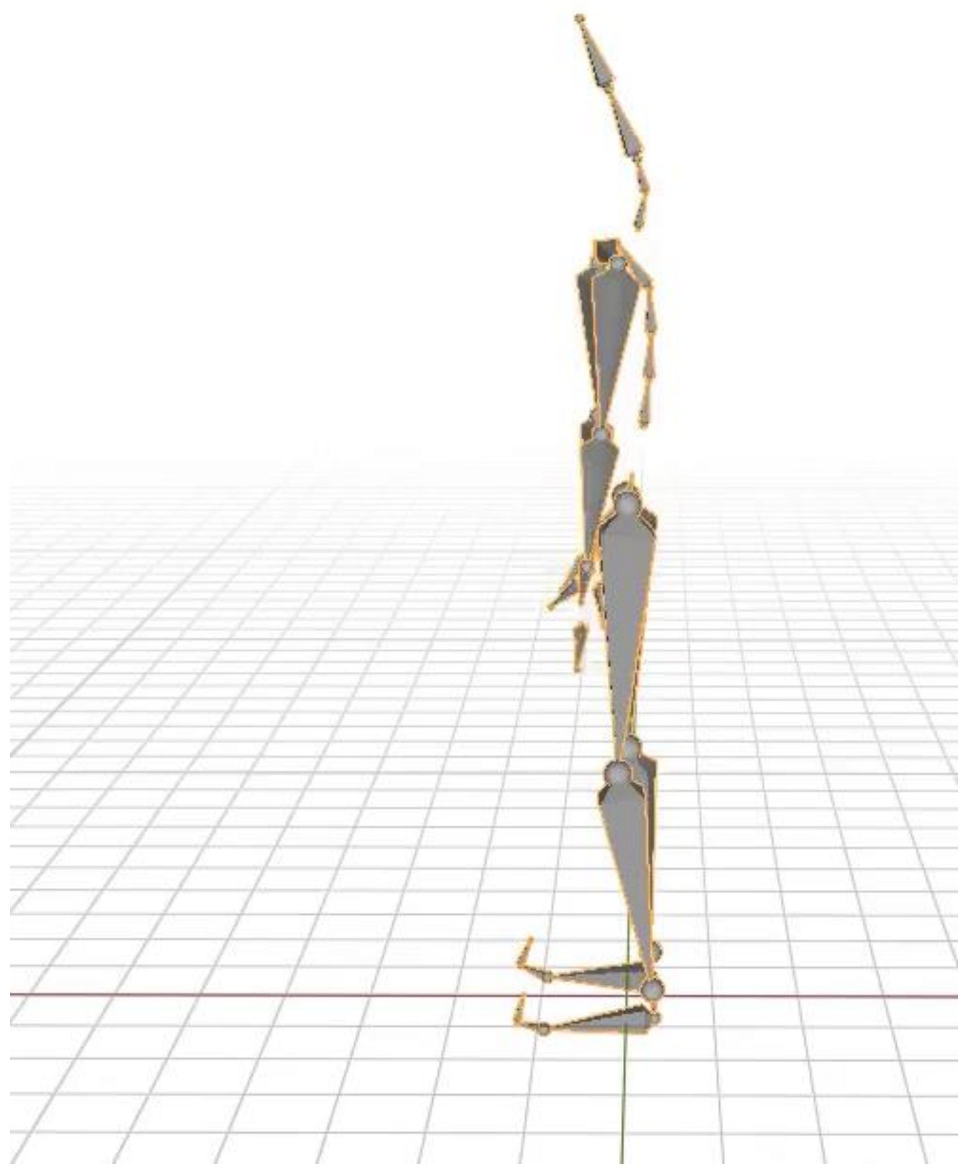
Transition loss

$$y = G_{c2a}(c)$$
$$\mathcal{L}_{transition,c} = \sum_t \sum_{DOF} ||y_i(t_{overlap:end}) - y_{i+1}(0 : t_{overlap})||$$

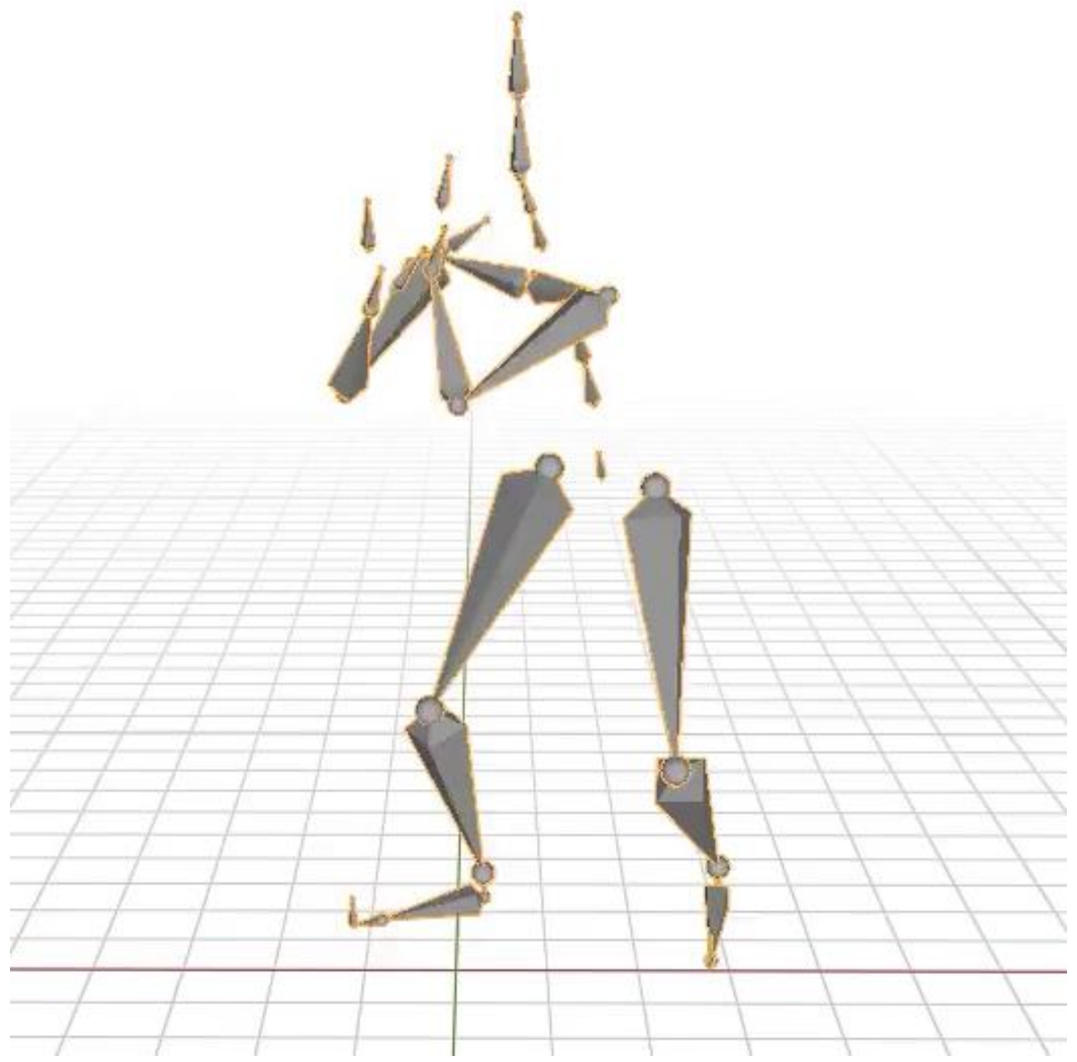


Results: *Punch*

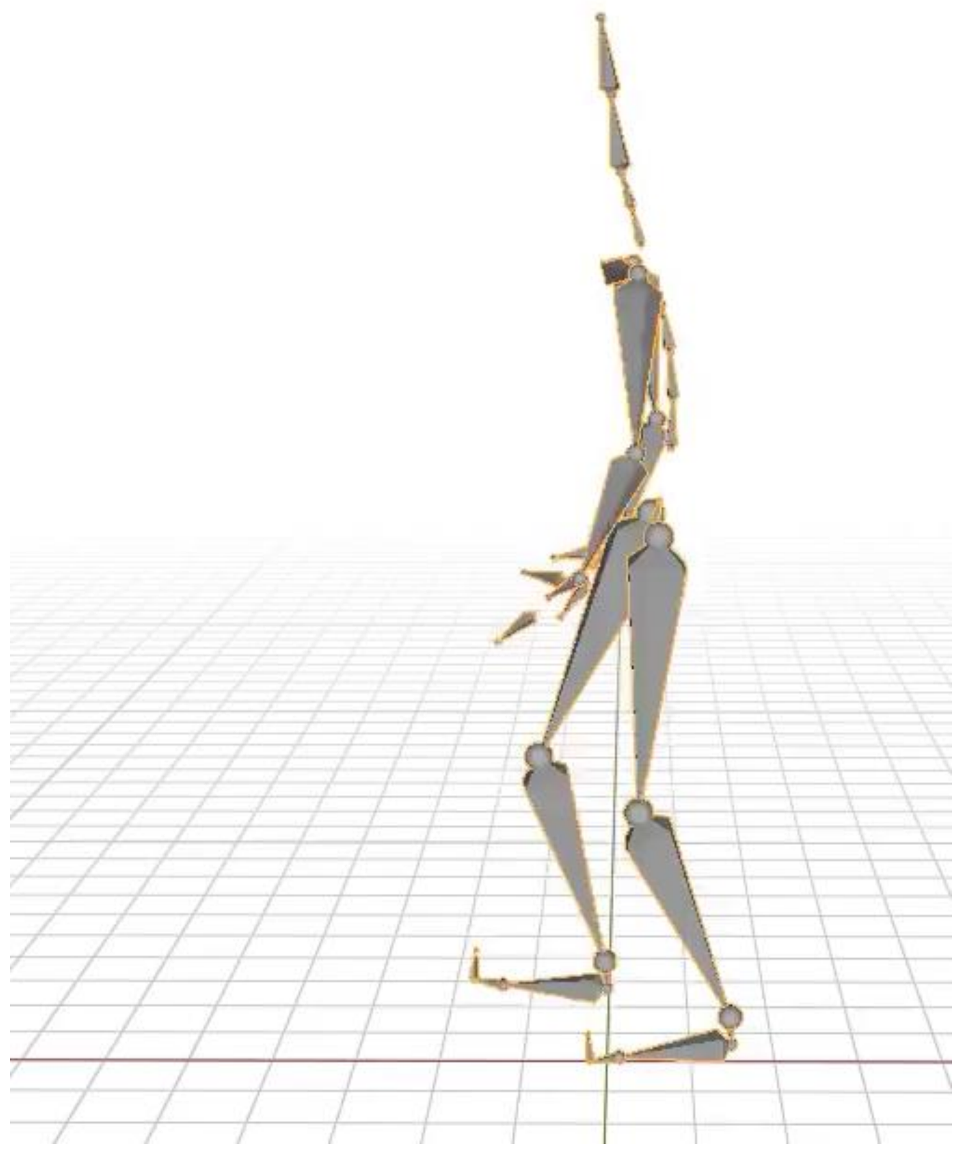
Input adult



Ours



Reference child

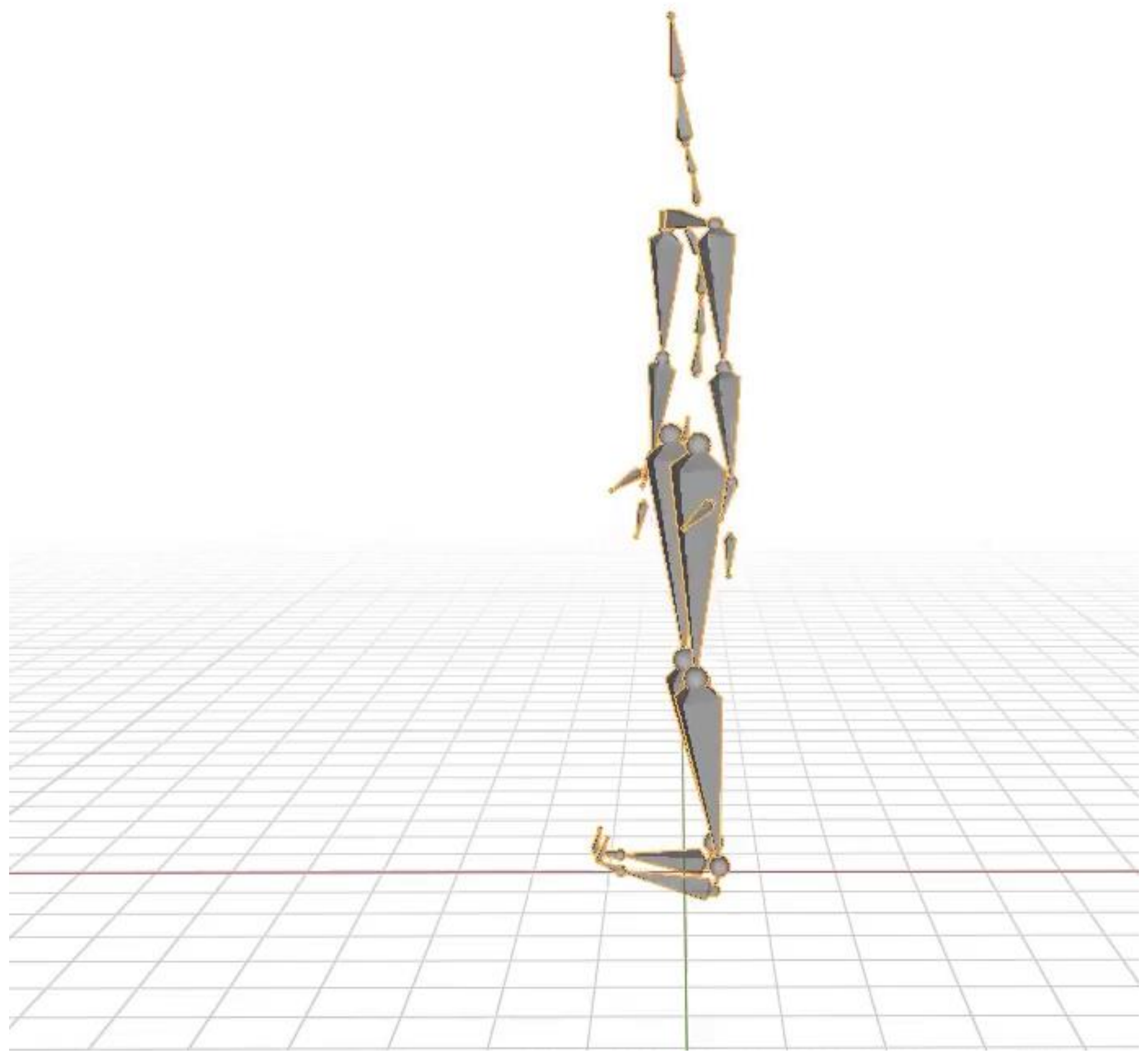


Co-financed by the European Union
Connecting Europe Facility

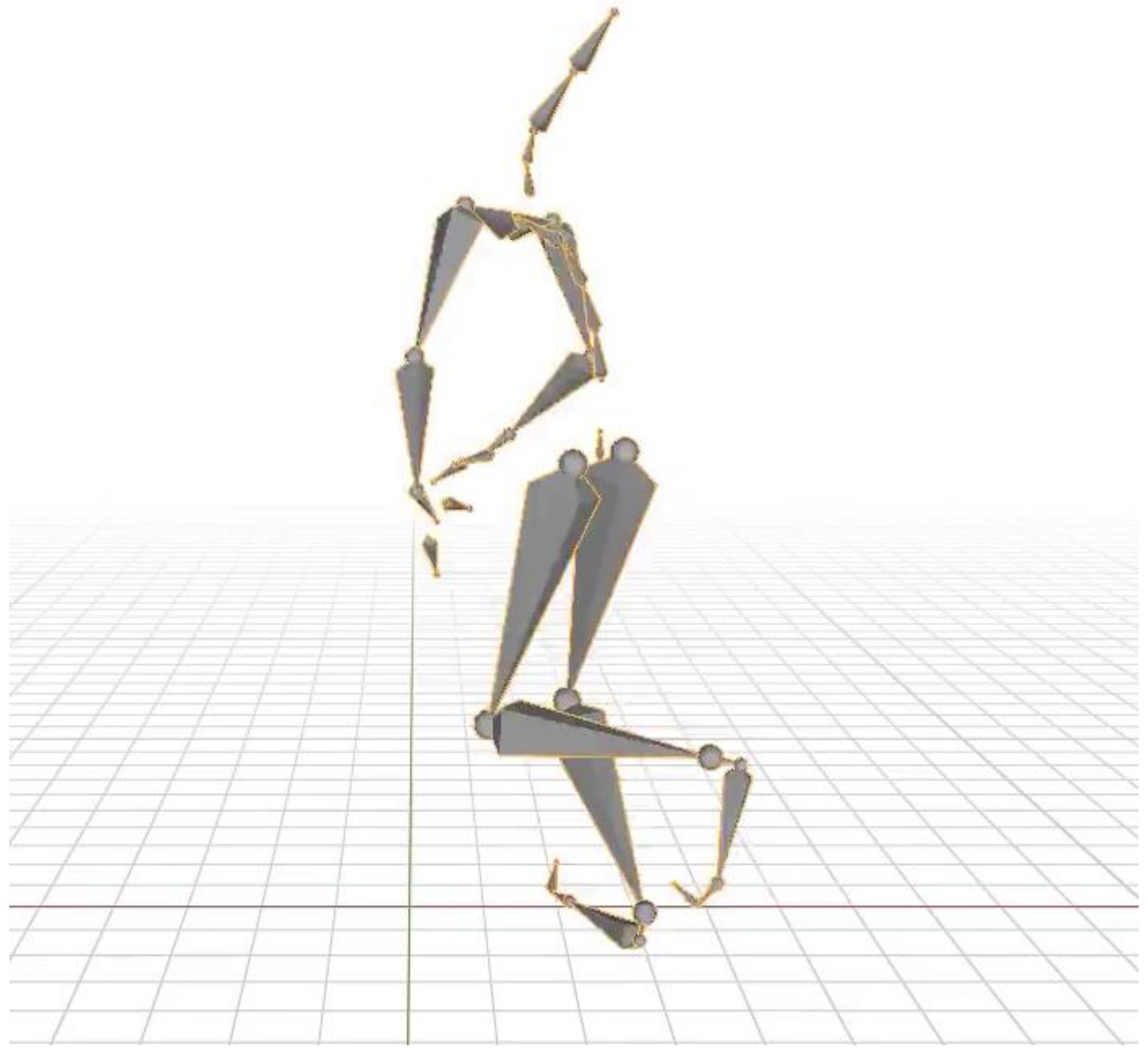
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results: *Run as fast as you can*

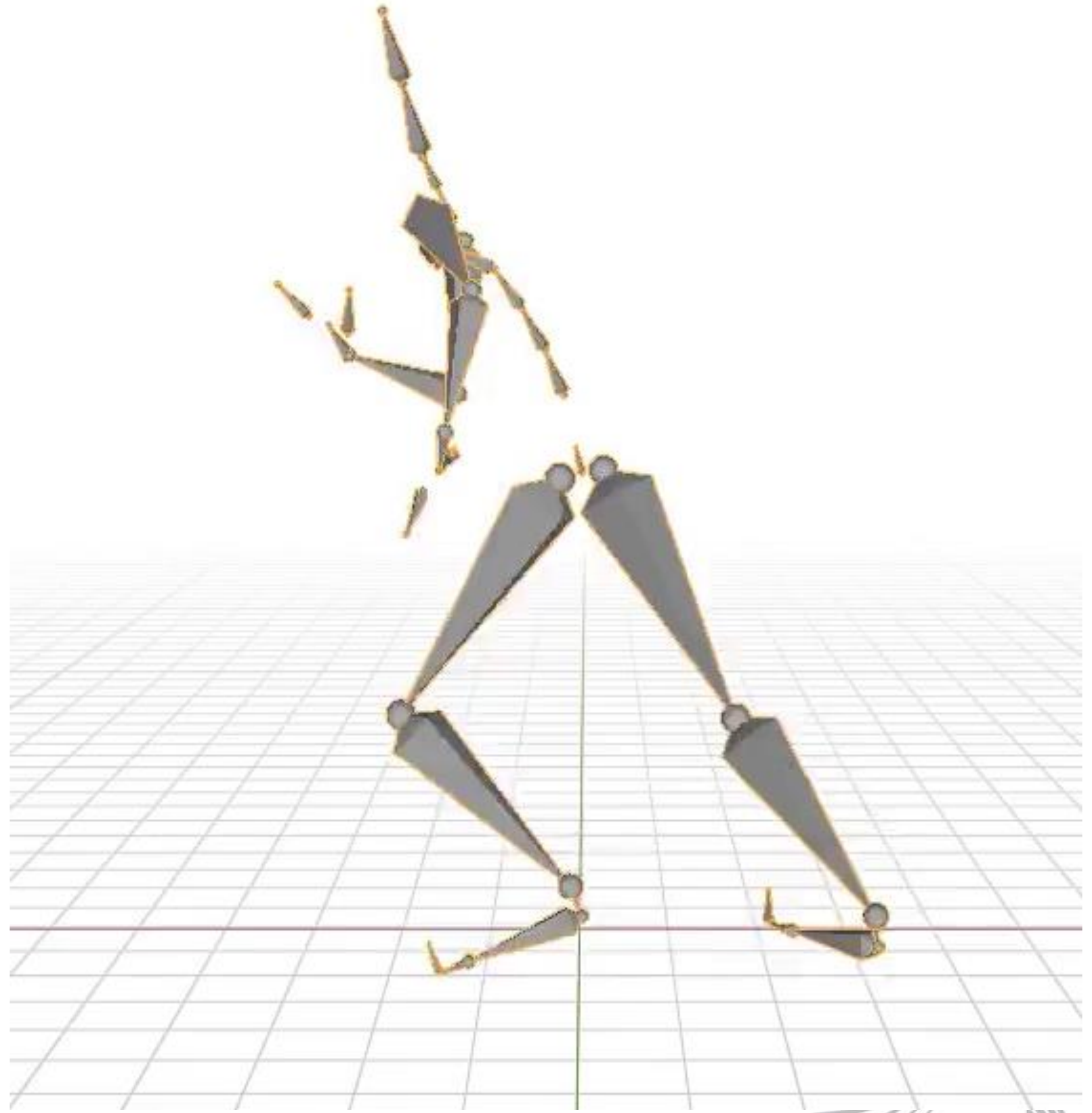
Input adult



Ours



Reference child



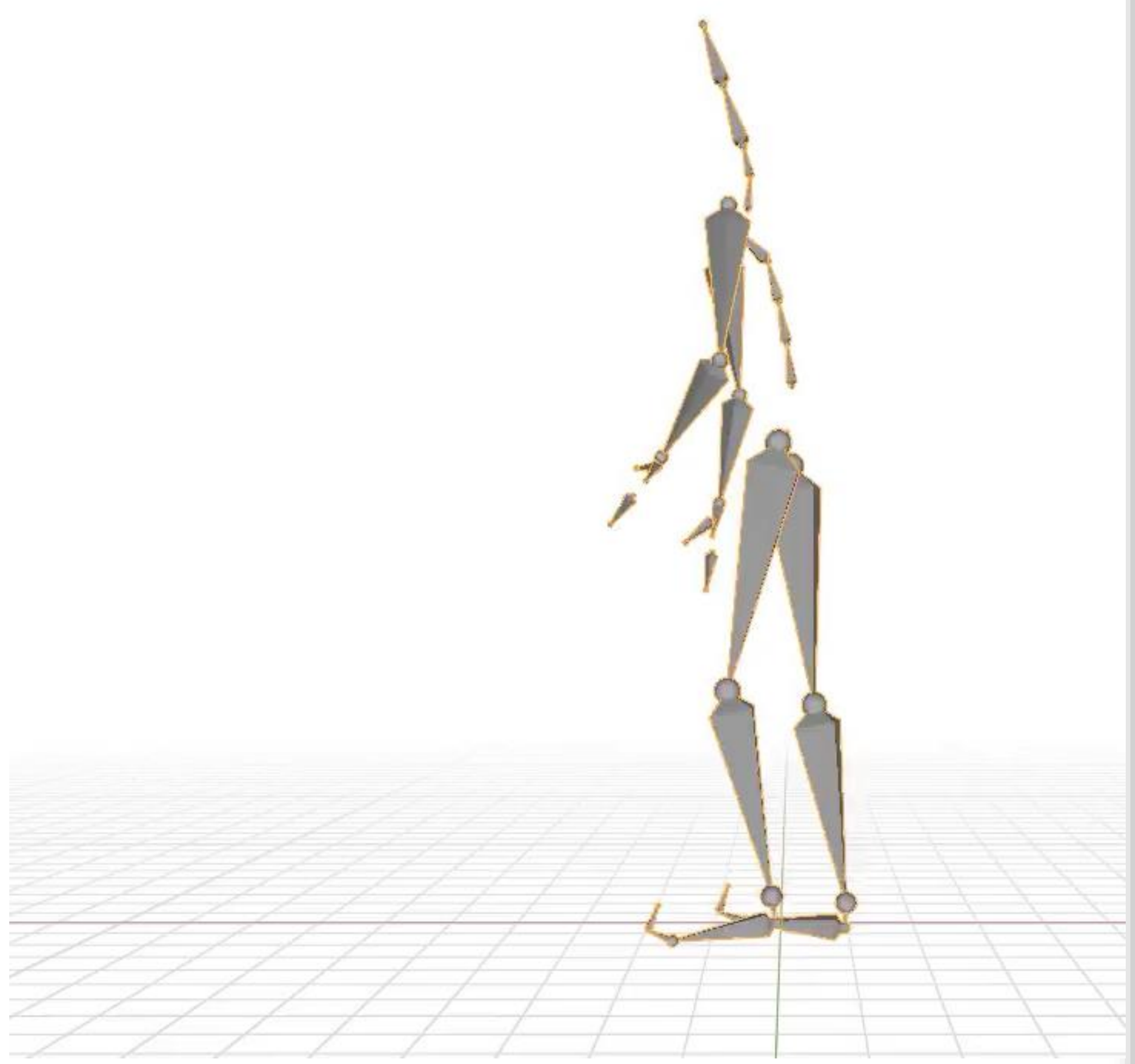
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results: *Walk as*

Input adul

Reference child



Adult2child: Motion Style Transfer using CycleGANs

Yuzhu Dong
University of Florida
yuzhudo1g@ufl.edu

Andreas Aristidou
University of Cyprus
RISE Research Centre
a.aristidou@ieee.org

Ariel Shamir
The Interdisciplinary Center Herzliya
arik@idc.ac.il

Moshe Mahler
Carnegie Mellon University
mmahler@andrew.cmu.edu

Eakta Jain
University of Florida
ejain@cise.ufl.edu

ABSTRACT

Child characters are commonly seen in leading roles in top-selling video games. Previous studies have shown that child motions are perceptually and stylistically different from those of adults. Creating motion for these characters by motion capturing children is uniquely challenging because of confusion, lack of patience and regulations. Retargeting adult motion, which is much easier to record, onto child skeletons, does not capture the stylistic differences. In this paper, we propose that style translation is an effective way to transform adult motion capture data to the style of child motion. Our method is based on CycleGAN, which allows training on a relatively small number of sequences of child and adult motions that do not even need to be temporally aligned. Our *adult2child* network converts short sequences of motions called motion words from one domain to the other. The network was trained using a motion capture database collected by our team containing 23 locomotion and exercise motions. We conducted a perception study to evaluate the success of style translation algorithms, including our algorithm and recently presented style translation neural networks. Results show that the translated adult motions are recognized as child motions significantly more often than adult motions.

CCS CONCEPTS

• **Computing methodologies → Motion capture; Motion processing; Machine learning; Animation.**

KEYWORDS

Style transfer, CycleGAN, Unpaired data, Motion Analysis

ACM Reference Format:
Yuzhu Dong, Andreas Aristidou, Ariel Shamir, Moshe Mahler, and Eakta Jain. 2020. Adult2child: Motion Style Transfer using CycleGANs. In *Motion, Interaction and Games (MIG '20), October 16–18, 2020, Virtual Event, SC, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3424636.3426909

ACM ISBN 978-1-4503-8171-0/20/10…$15.00
https://doi.org/10.1145/3424636.3426909

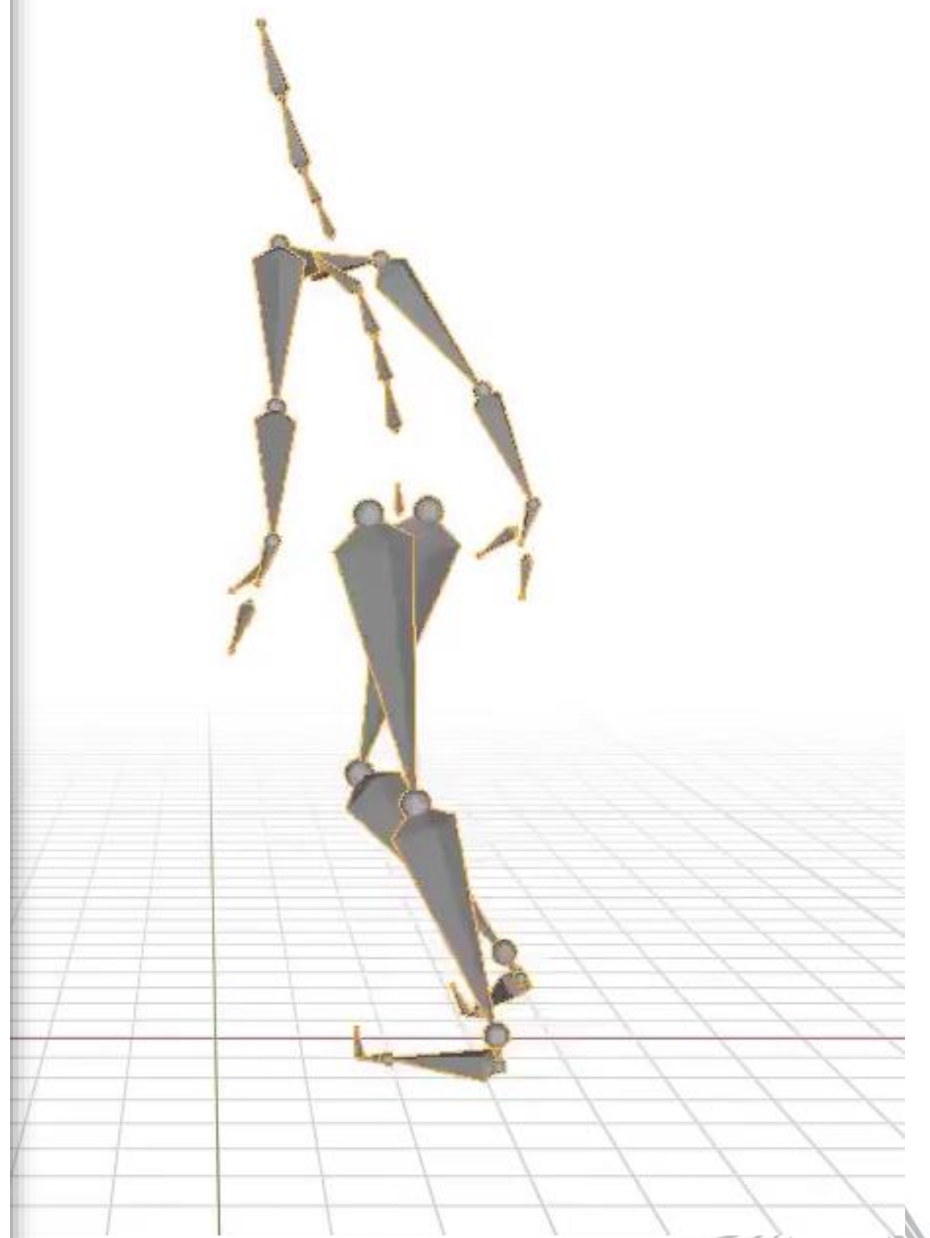
1 INTRODUCTION

Children in the age group 8 to 11 years old have been found to spend as much as 8 hours weekly on video games [Johnson 2018]. Such trends make children important markets for the video game and electronic entertainment industry. Games such as Just Dance, by Ubisoft, and Ring Fit Adventure, by Nintendo, are designed to motivate children to exercise. As a result, there is a need to identify methods for synthesizing child motions.

Keyframing requires hours of manual effort from trained animators to create realistic and compelling motion. Motion capture (mocap), the leading technology for creating animated characters from actual human motion data, has the advantage of maintaining realism, capturing subtle secondary movements, and following real world physics [Menache 2000]. However, motion capturing children is full of difficulties. Children get confused with the instructions, lack patience, and are hard to collaborate with [Piaget 2015], especially at very young ages. These difficulties are the reason there are few online motion repositories. The most well-known mocap repositories, such as the CMU [2020] and OSU [2020] databases, consist only of adult motions. Currently, the Kinder-gator [Aloba et al. 2018] and the Human Motion Database [Guerra-Filho and Biswas 2012] are the only publicly accessible repositories that contain child motion. For games in particular, an abundance of action types, repetitions and variations allows for realism in real time play.

One way to overcome this scarcity of child motion data is to retarget easily available motion from adults to a child sized skeleton. However, retargeting mostly involves changes in the dimensions of limbs, so mapping adult motion directly on child characters fails to transfer the *style* and nuances of the children motion such as speed and variability. Style translation, that is, learning a mapping between two labeled motion capture sequences, has been extensively studied, starting with approaches by Brand and Hertzmann [2000] and Gleicher [1998] to recent advances made by deep neural networks [Aberman et al. 2020; Du et al. 2019a; Holden et al. 2017, 2016, 2015; Mason et al. 2018; Smith et al. 2019].

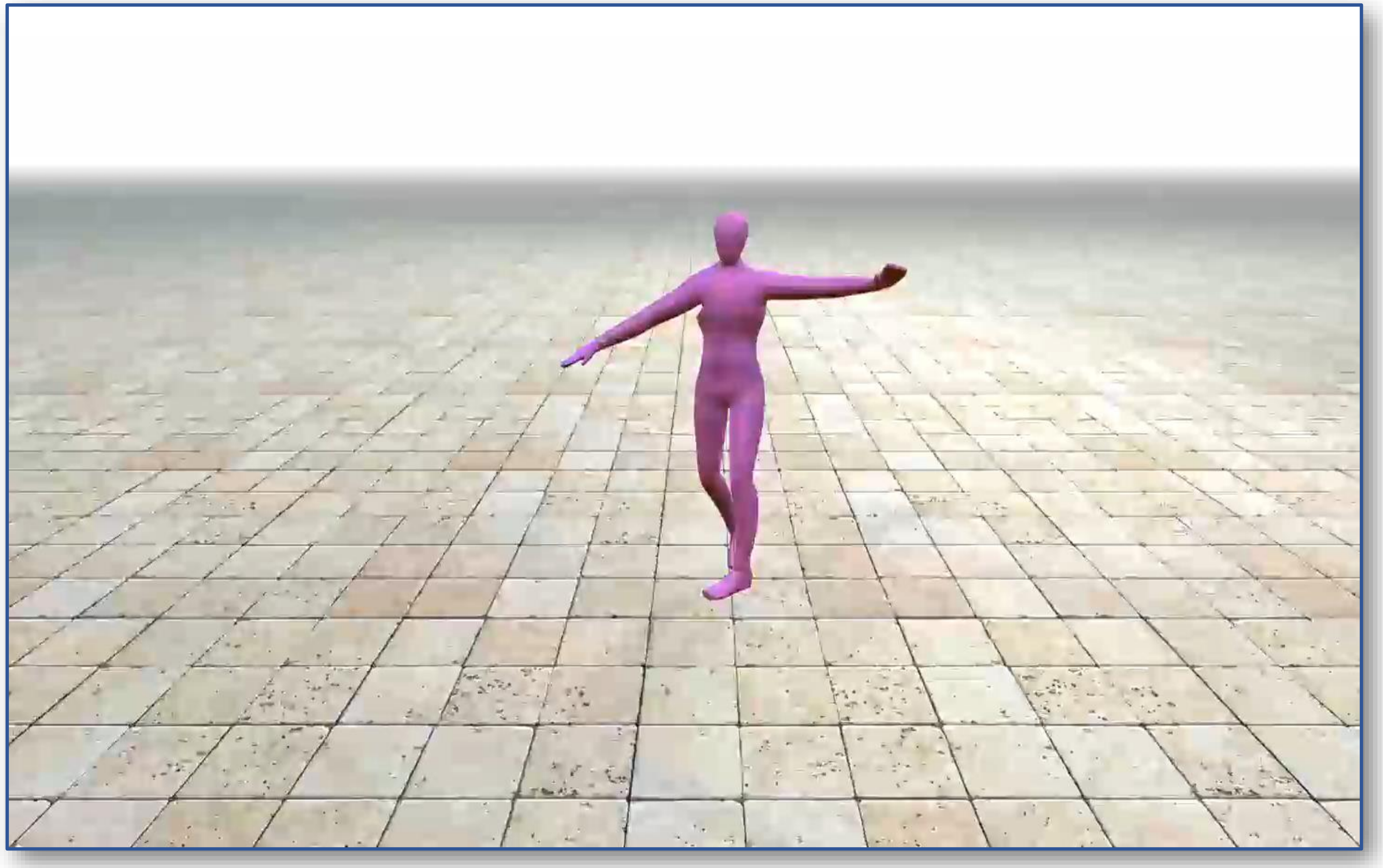
In this paper we devise an adult-to-child motion translation algorithm based on the CycleGAN [Zhu et al. 2017] architecture. CycleGAN has been successfully used in the past for transforming image styles *without* paired training data. This characteristic is critical for adult-to-child translation due to the very limited availability of child data. Generative Adversarial Networks (GANs) have rarely been used in character animation because of the difficulty to train a mapping that exhibits temporal dynamic behavior and generates temporally coherent and realistic movements. We show



Co-financed by the European Union
Connecting Europe Facility

der the context of Action
by the EU CEF Telecom
EF/ICT/A2020/2267423

Contextual Analysis

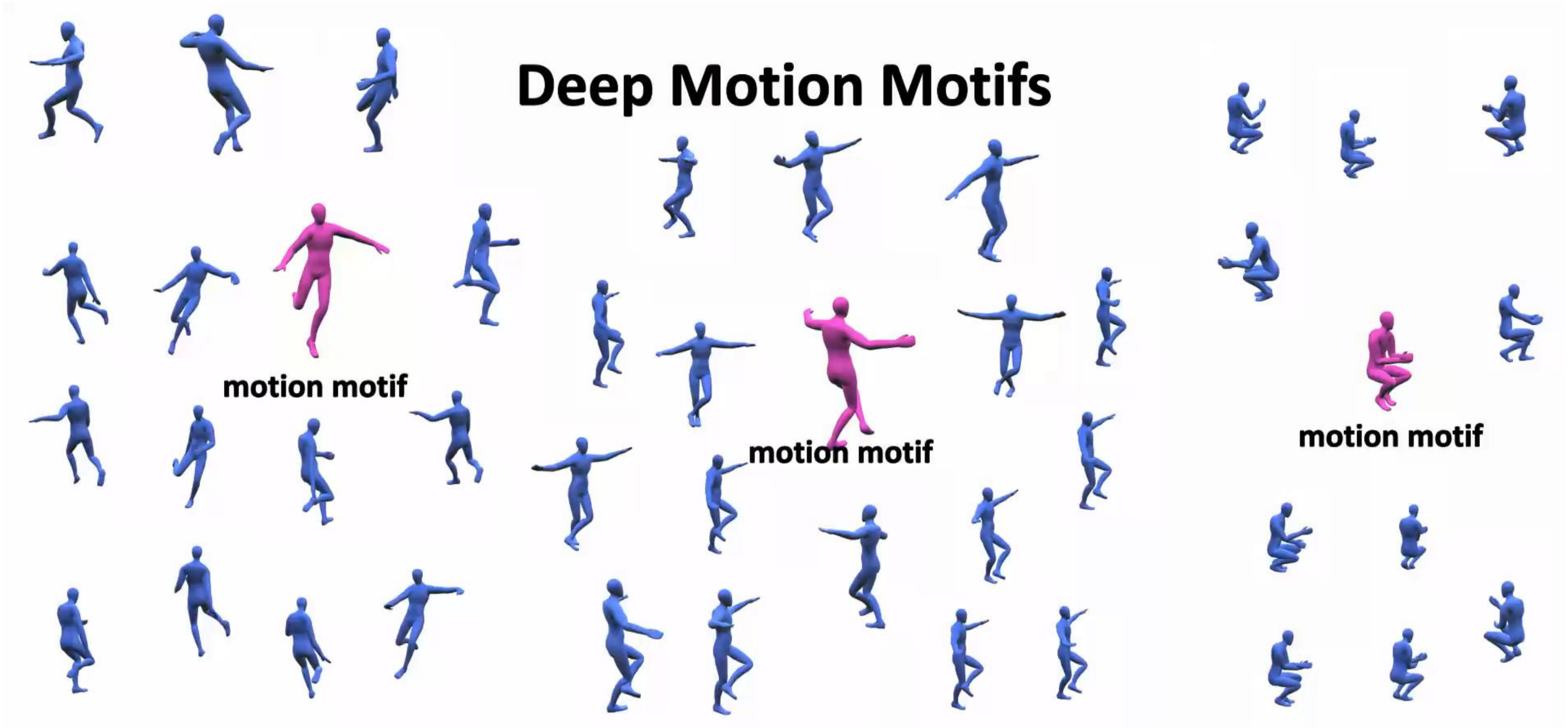


This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

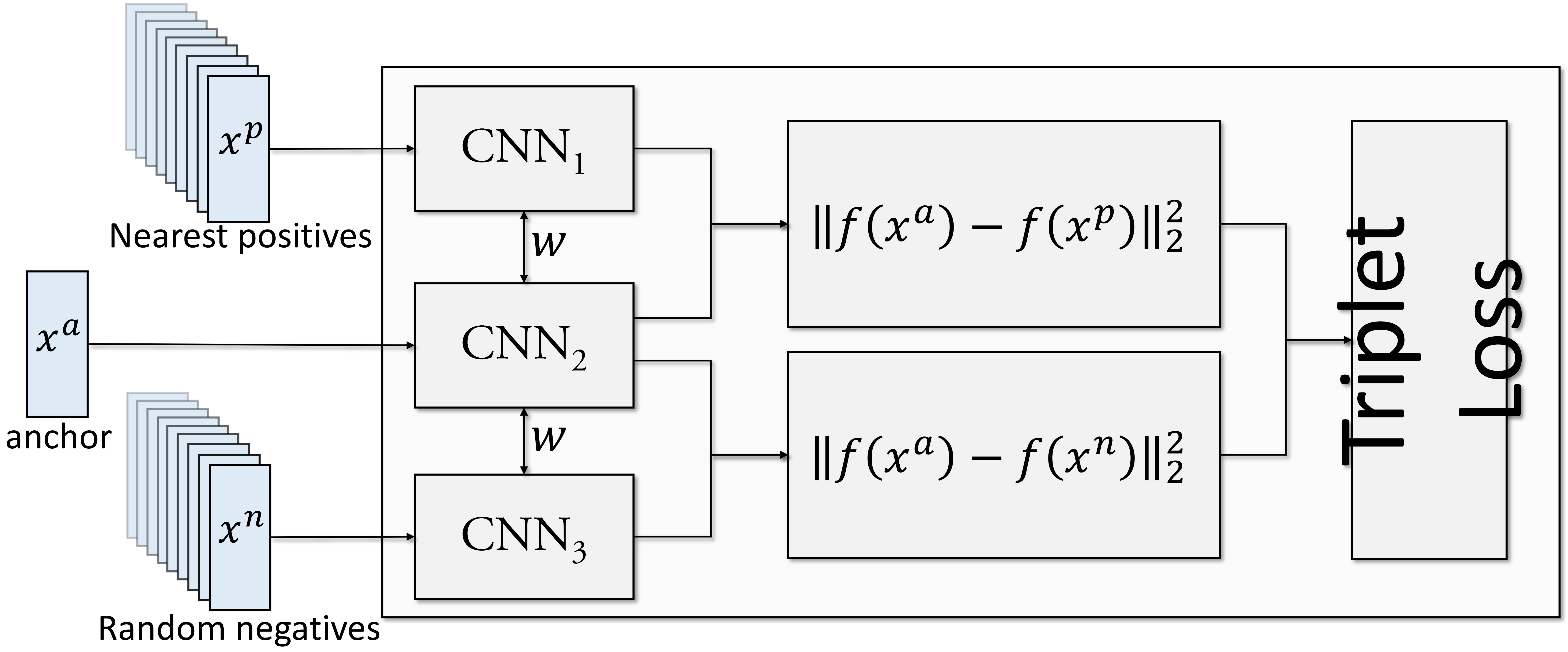
Deep motifs and motion signatures

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Motion Words and Motifs



Triplet Loss Network



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Motion Analysis

Triplet Loss Network

The loss for a single tripl efined as:

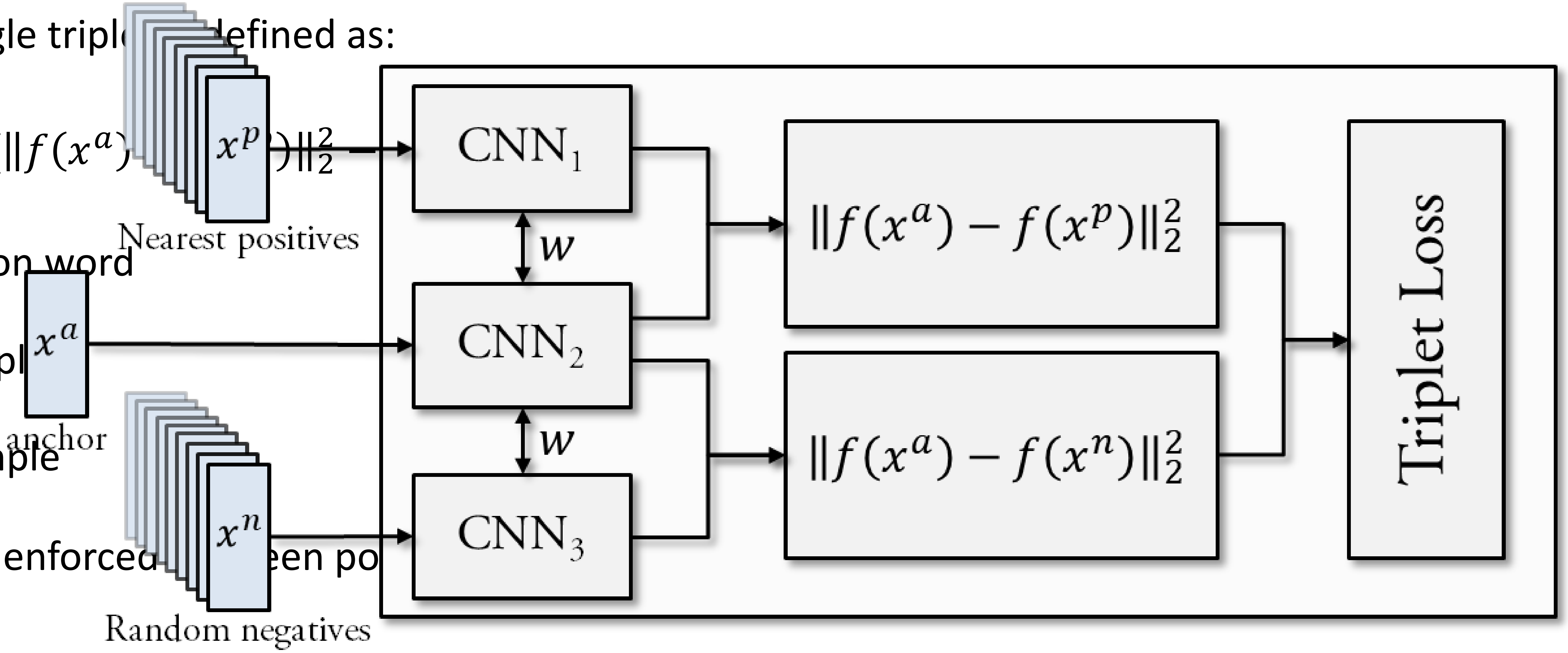
$L(x^a, x^p, x^n) = [\|f(x^a)$ $)\|_2^2$

x^a - anchor motion word

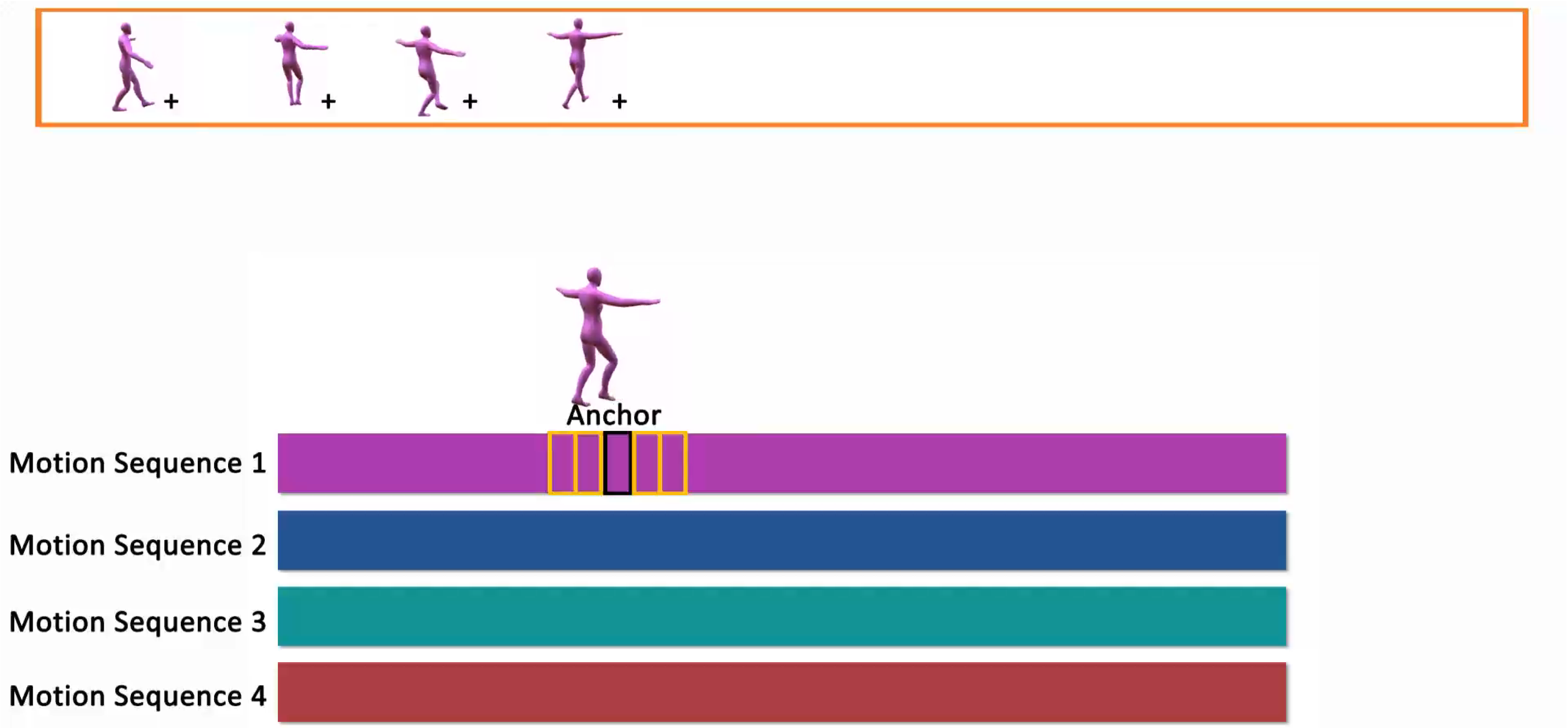
x^p - positive sampl

x^n - negative sample

α - margin that is enforced een po



Triplet Loss Network

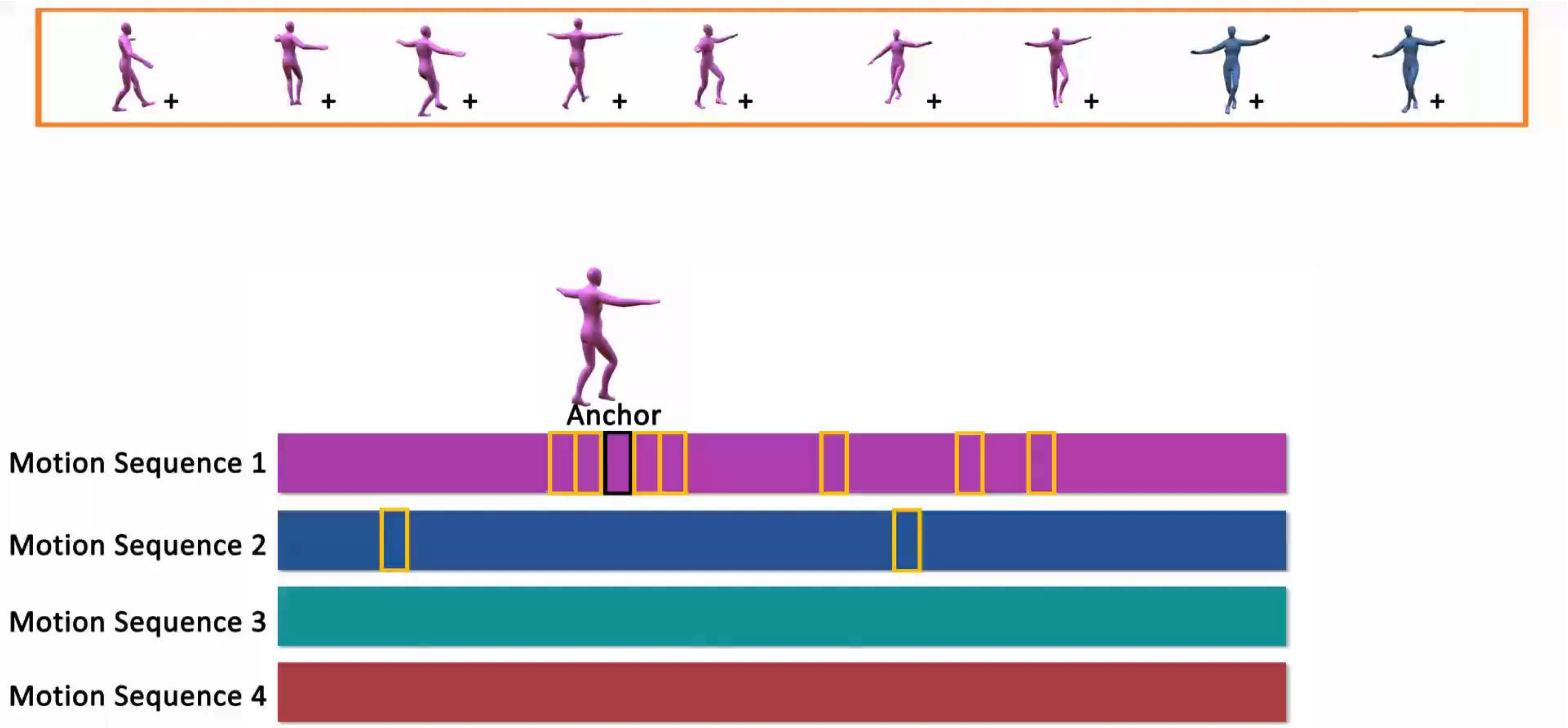


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Motion Analysis

Triplet Loss Network

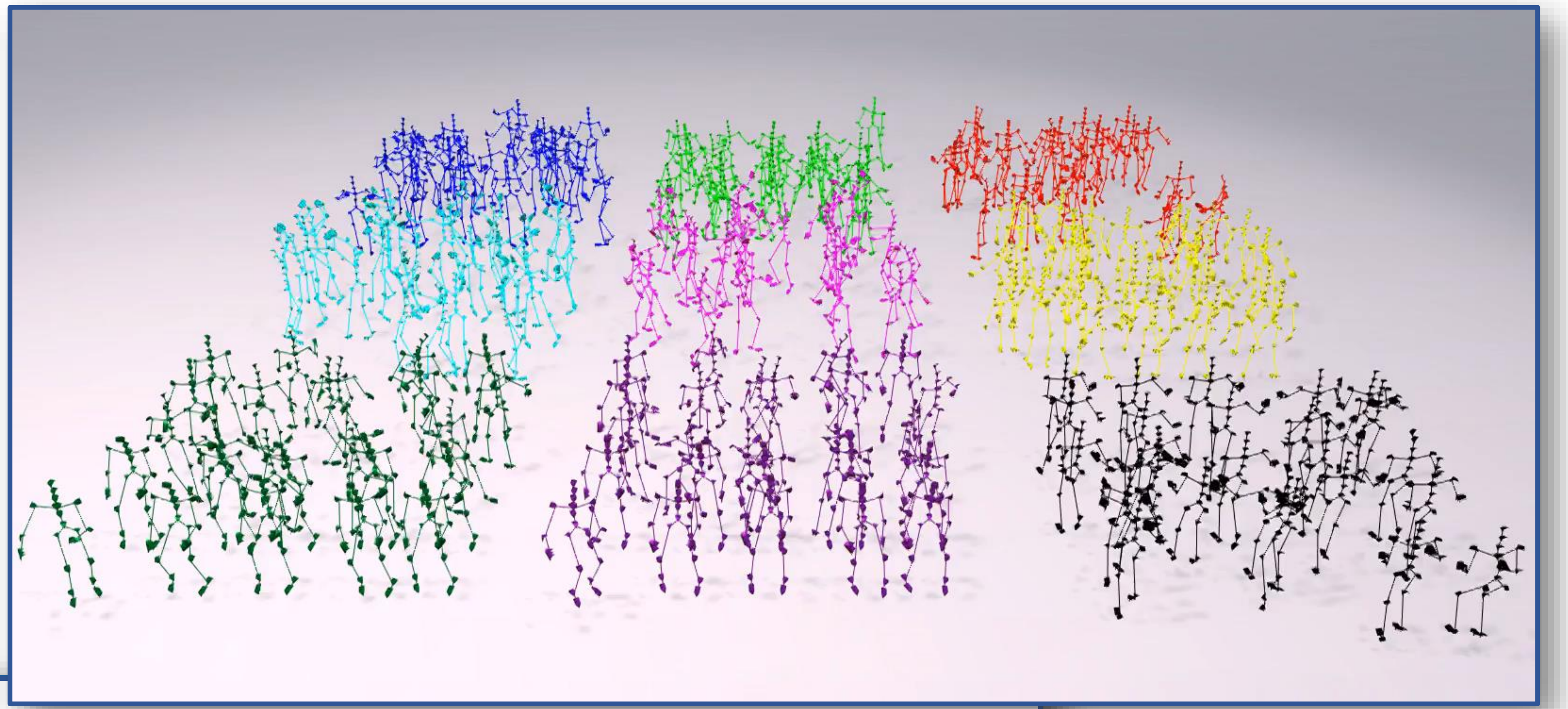


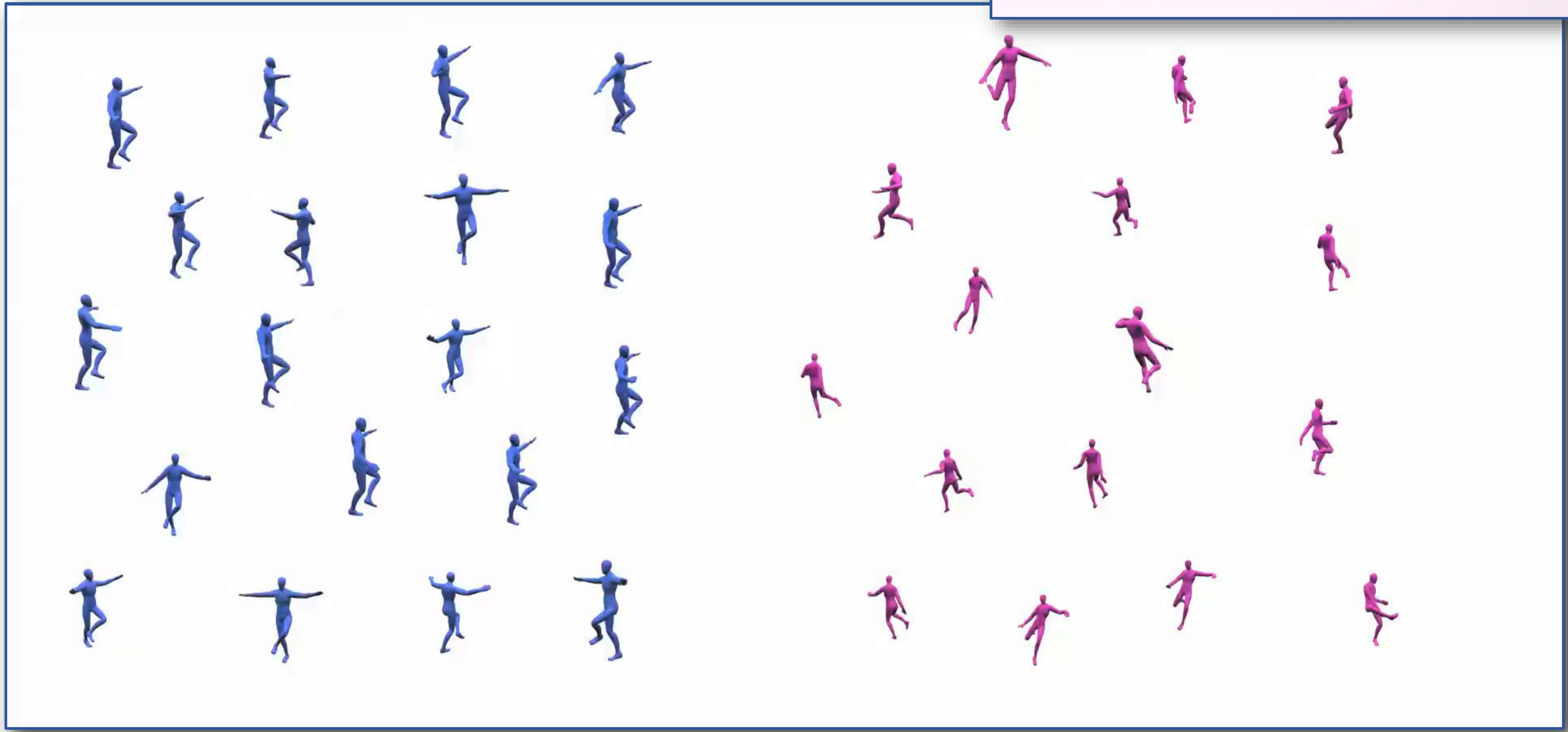
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Feature Space

Motion Words



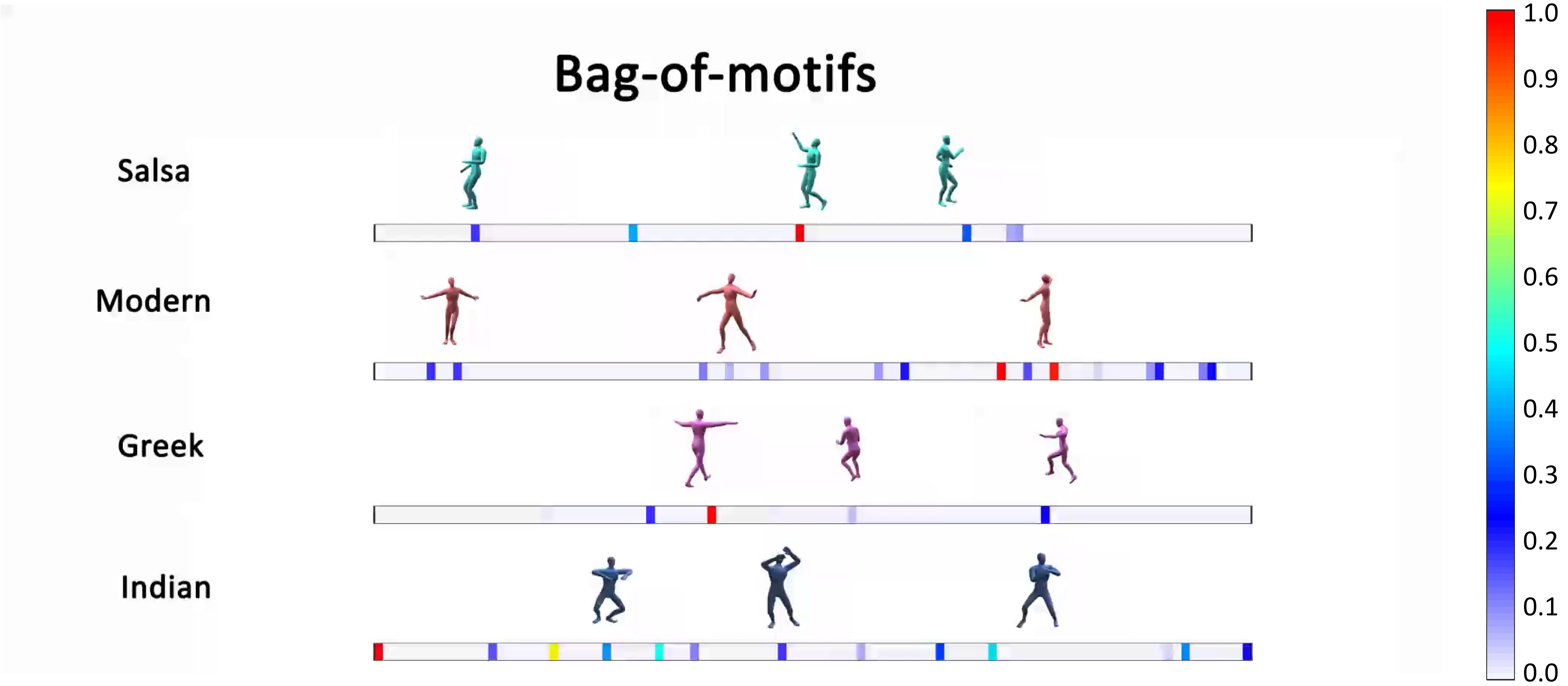


Co-financed by the European Union
Connecting Europe Facility

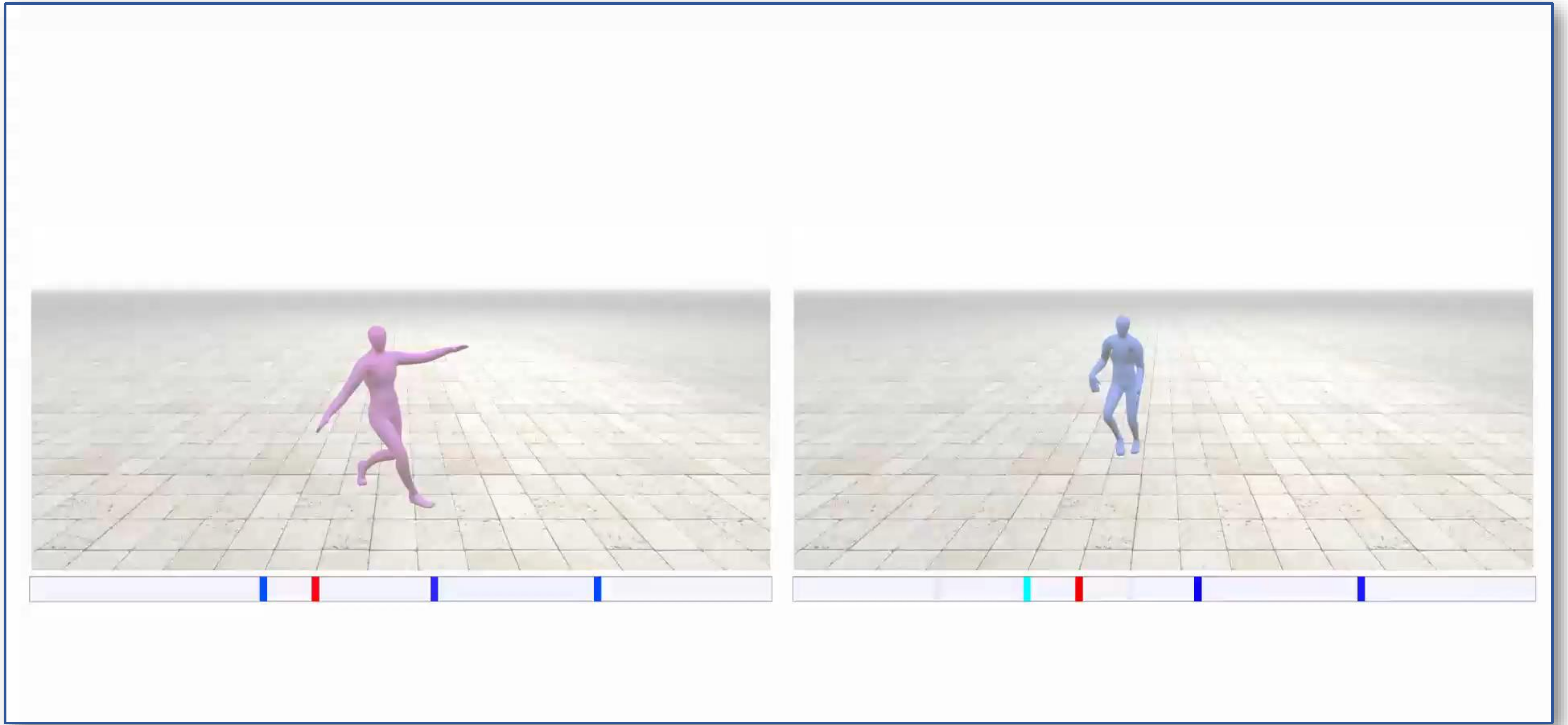
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Bag-of-motifs

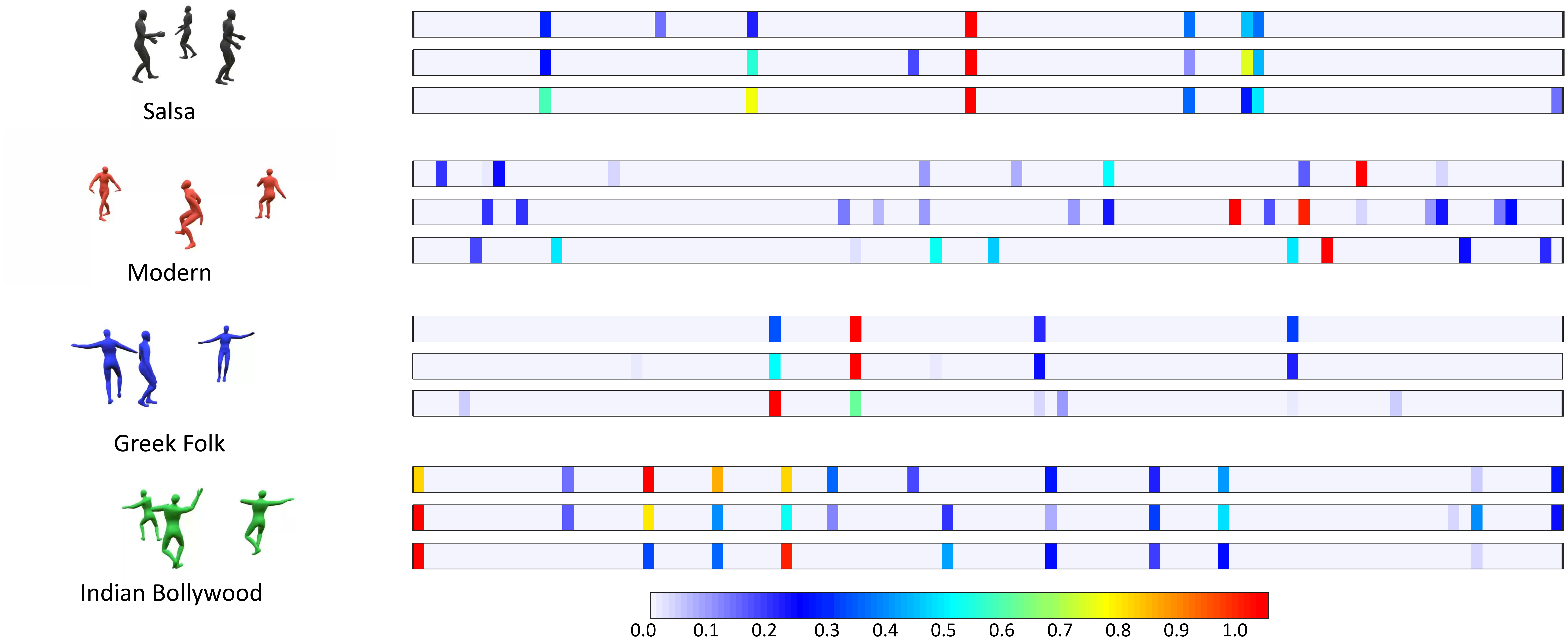
Motion Signatures



Motion Signatures



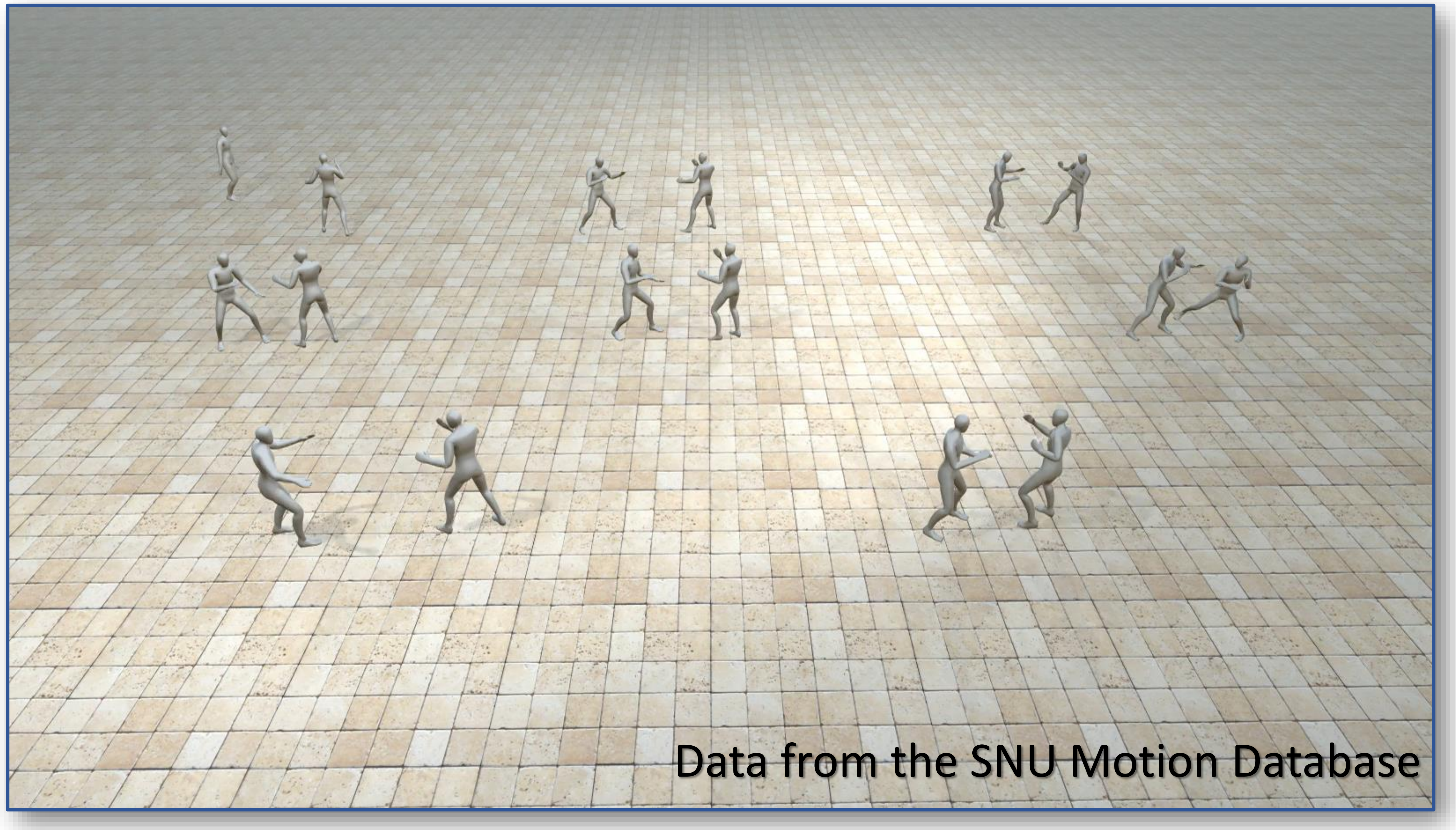
Motion Signatures

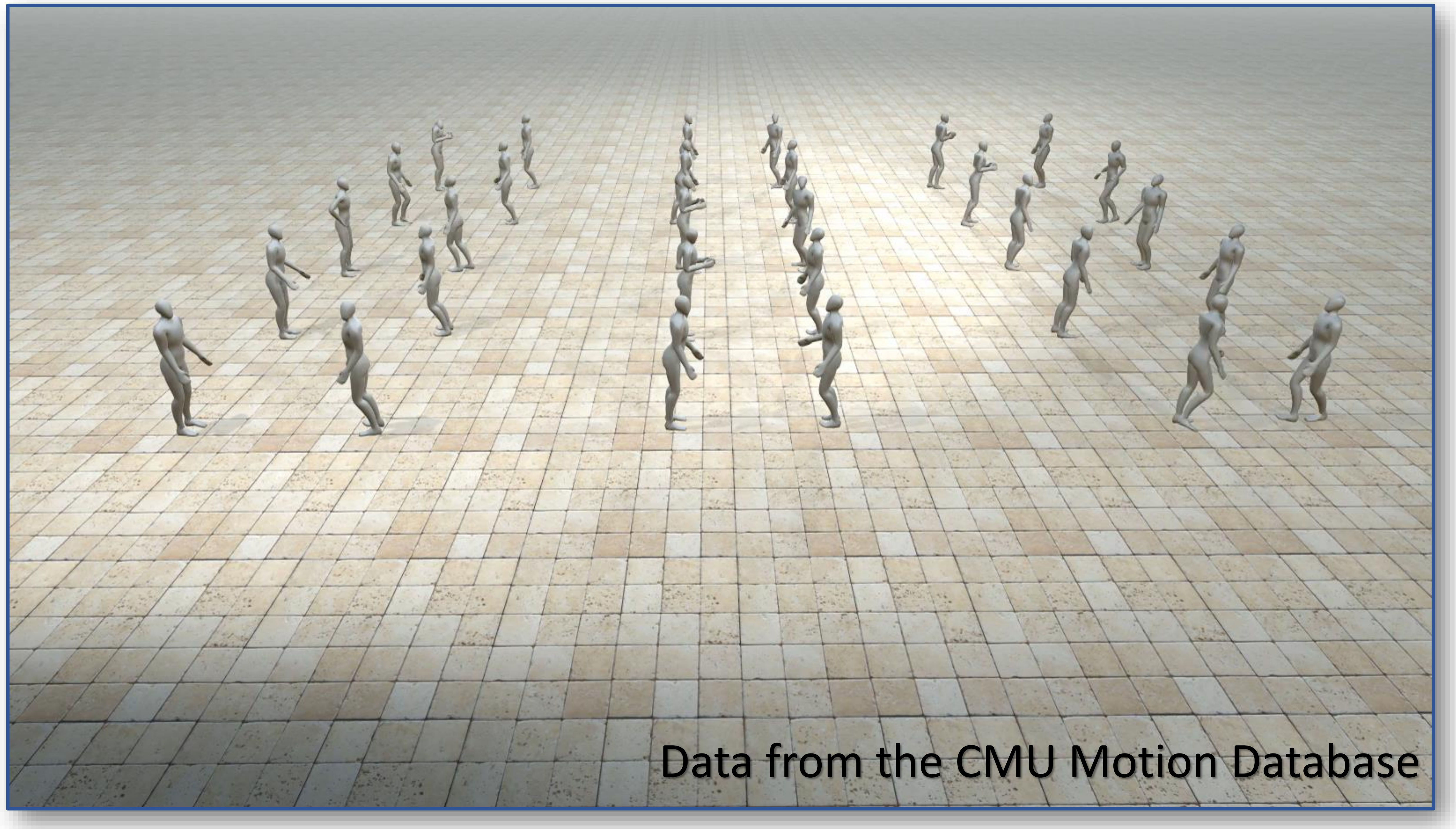


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Fine-grained details



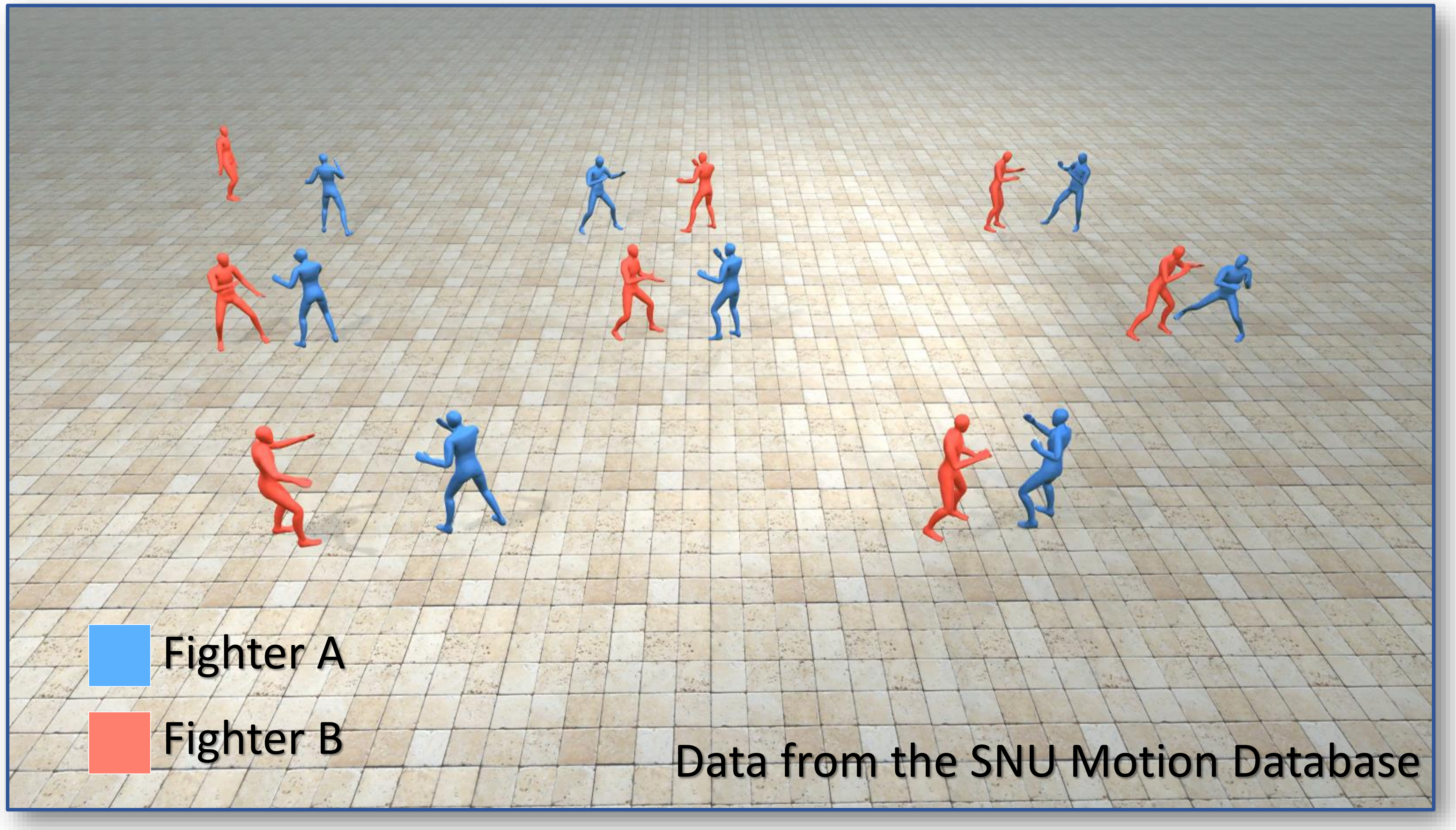


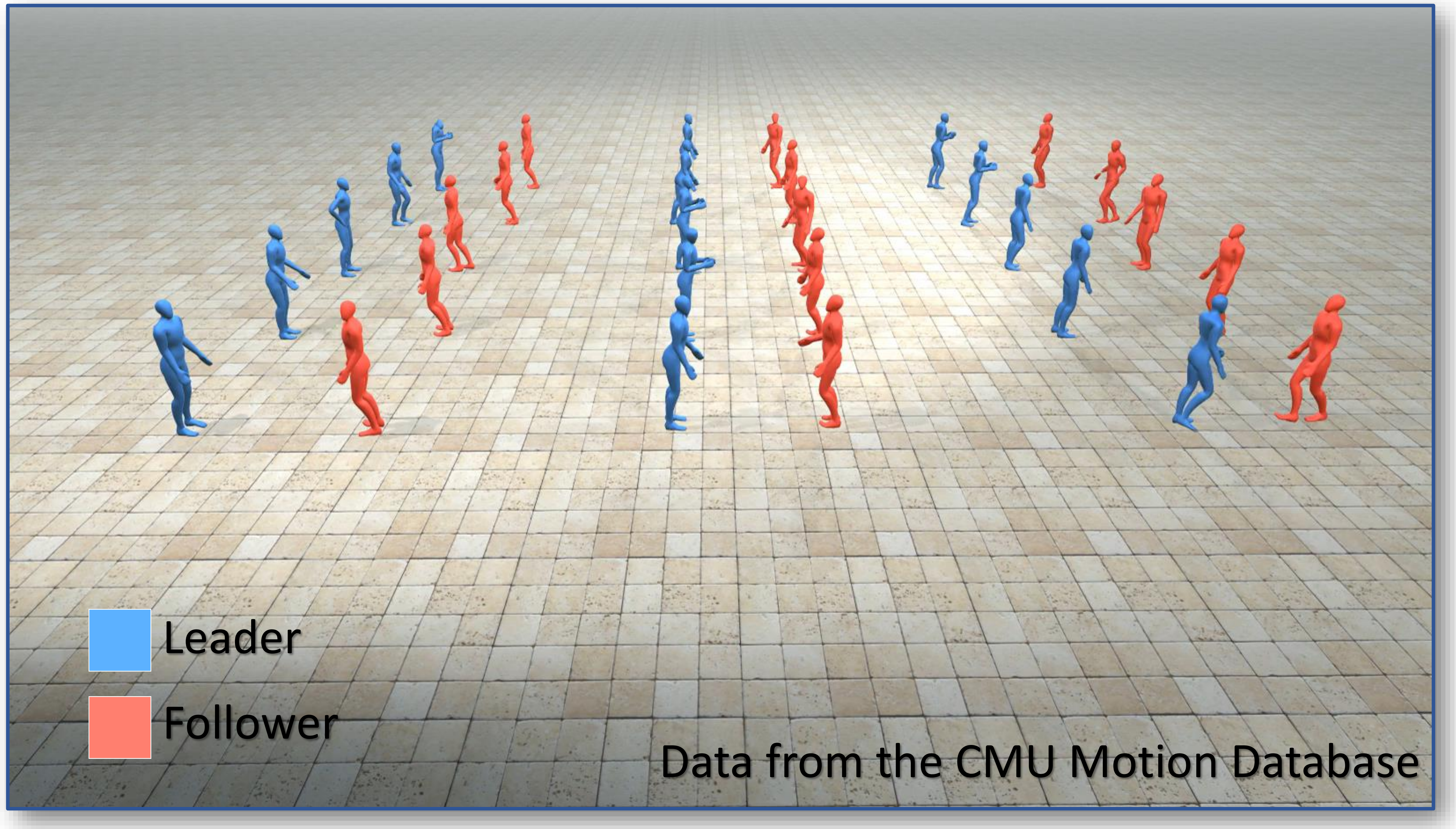
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Deep motifs and motion signatures

Fine-grained details

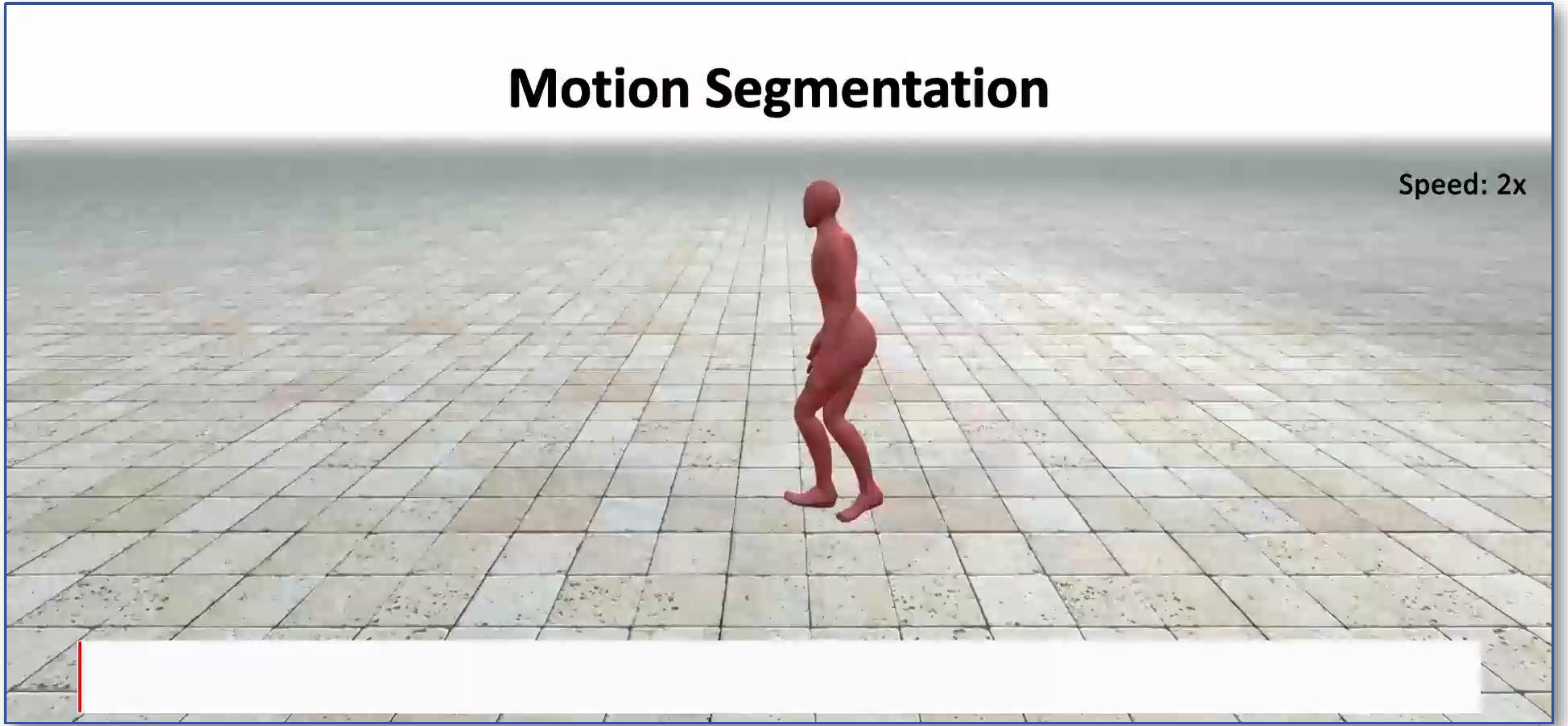




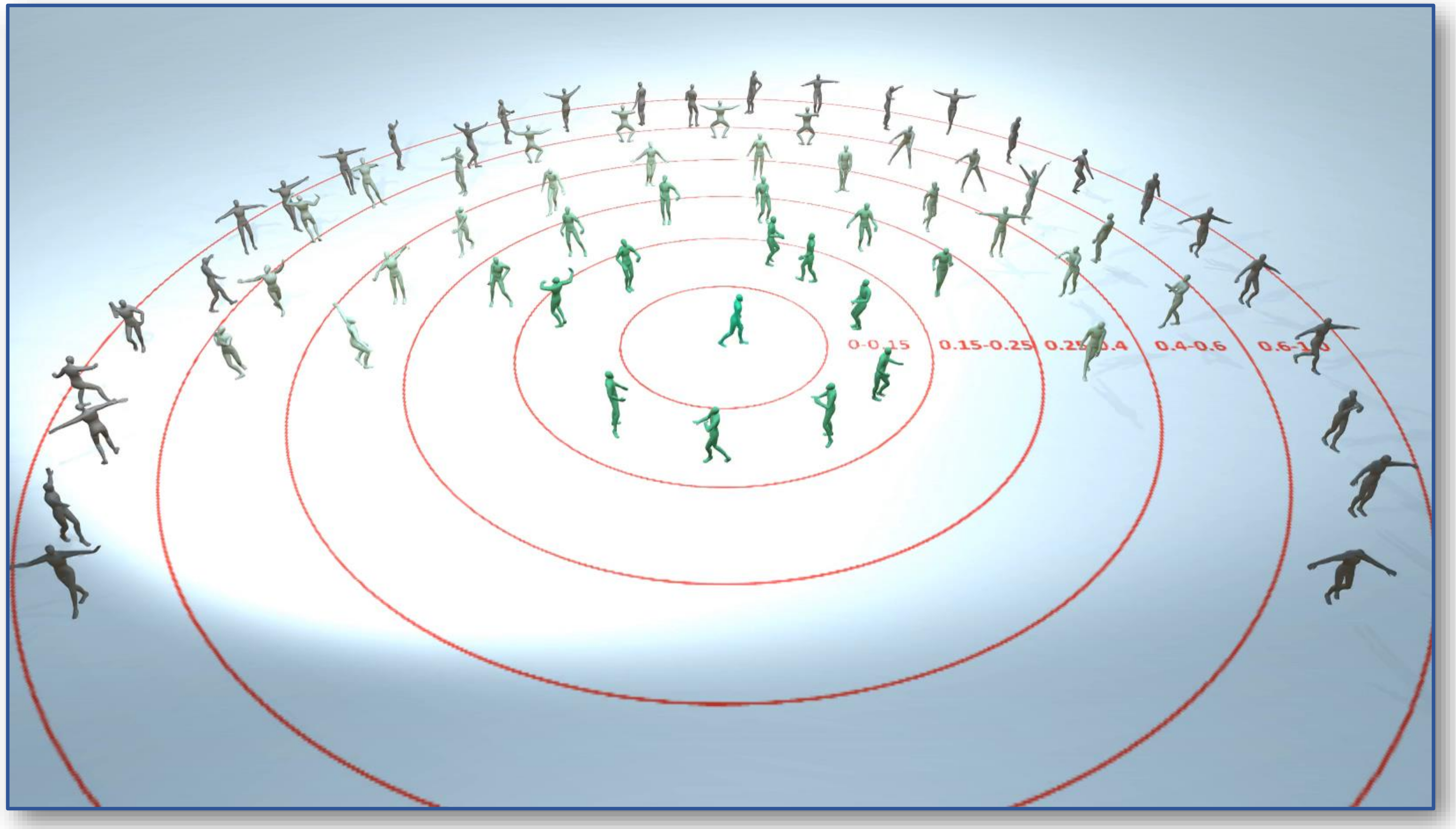
Co-financed by the European Union
Connecting Europe Facility

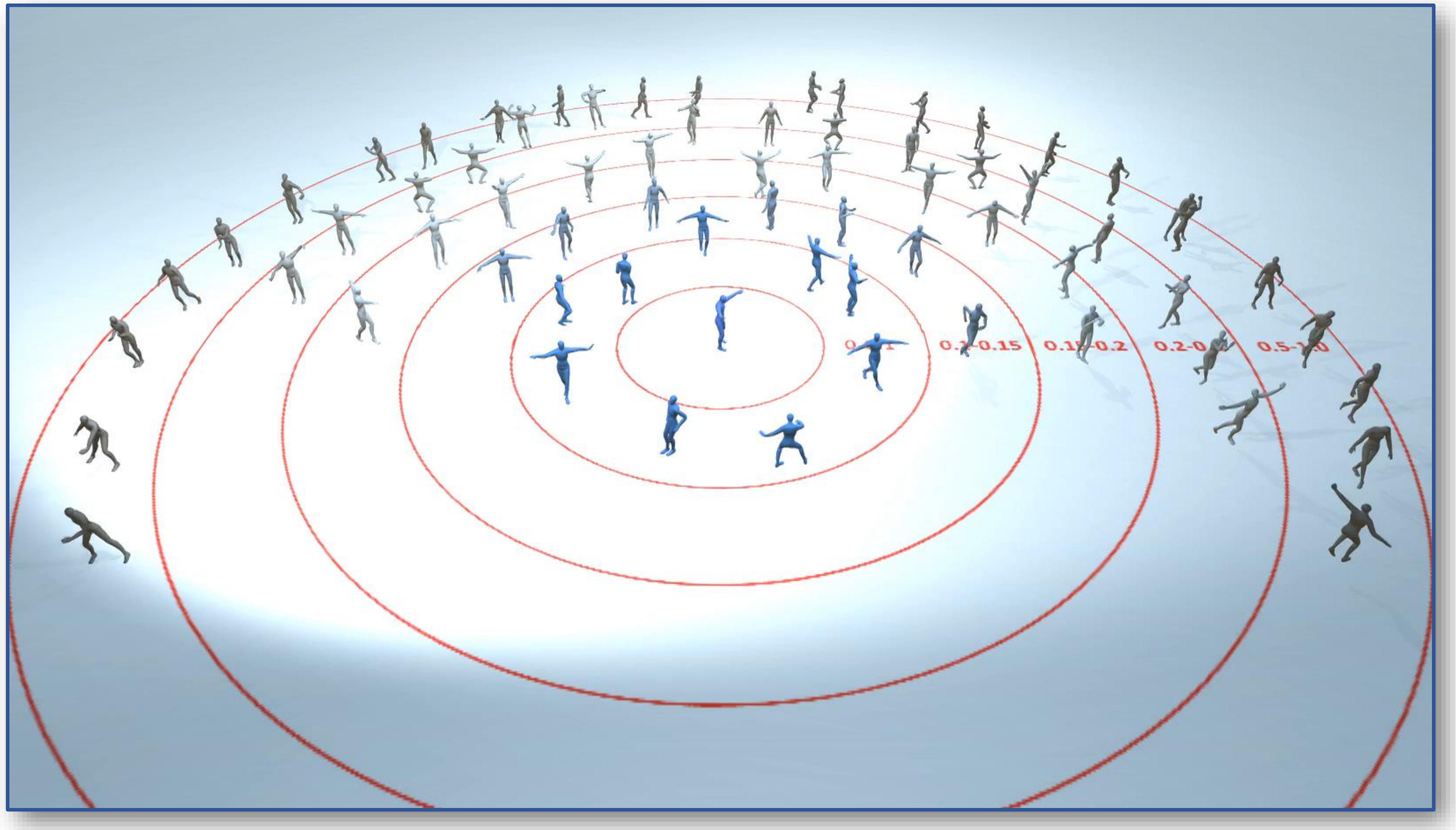
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Motion Segmentation



Organizing large collections: *Dance ethnography*



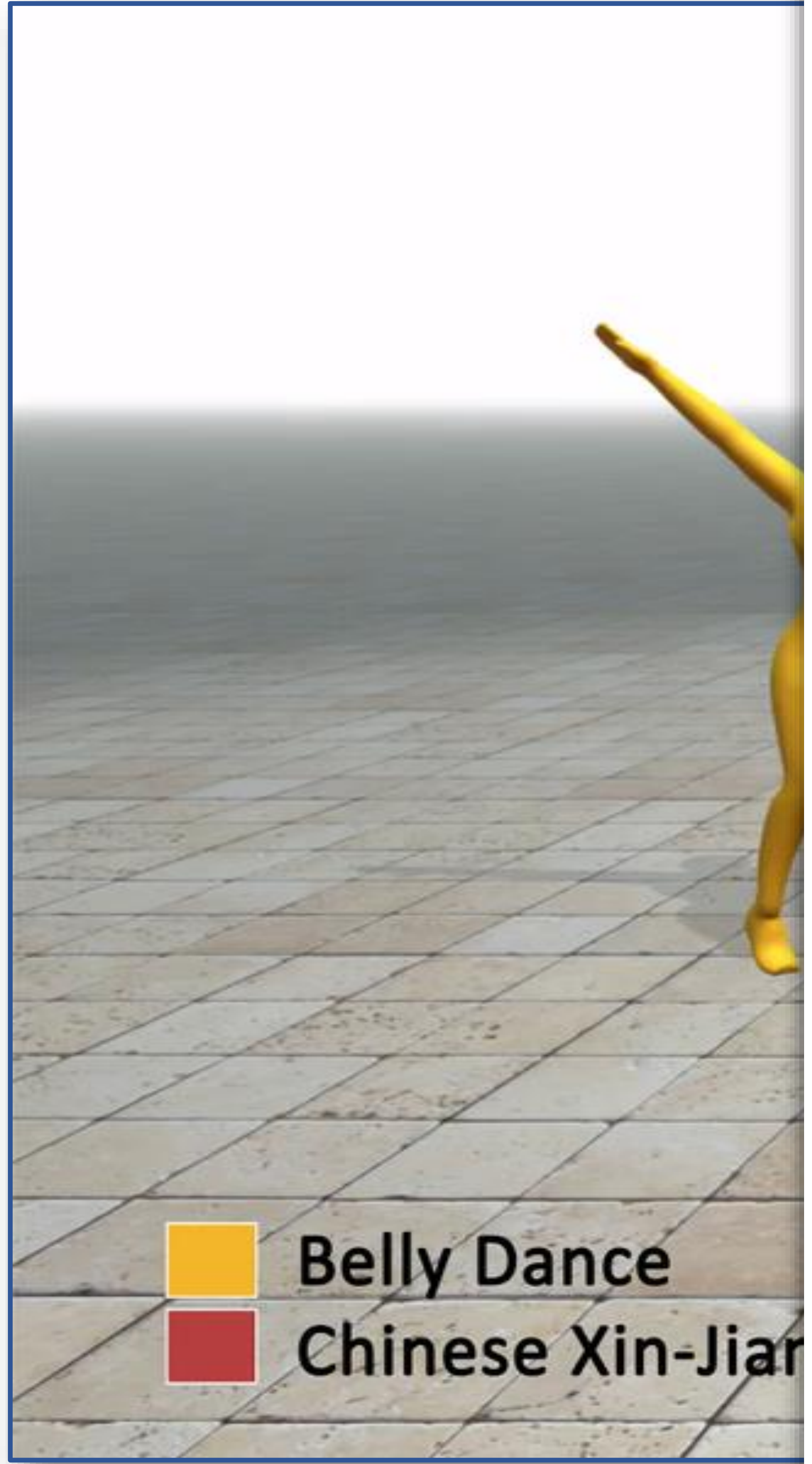


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Deep motifs and moti

Unexpected asso



Deep Motifs and Motion Signatures

ANDREAS ARISTIDOU, The Interdisciplinary Center
DANIEL COHEN-OR, Tel-Aviv University
JESSICA K. HODGINS, Carnegie Mellon University
YIORGOS CHRYSANTHOU, University of Cyprus & RISE Research Center
ARIEL SHAMIR, The Interdisciplinary Center

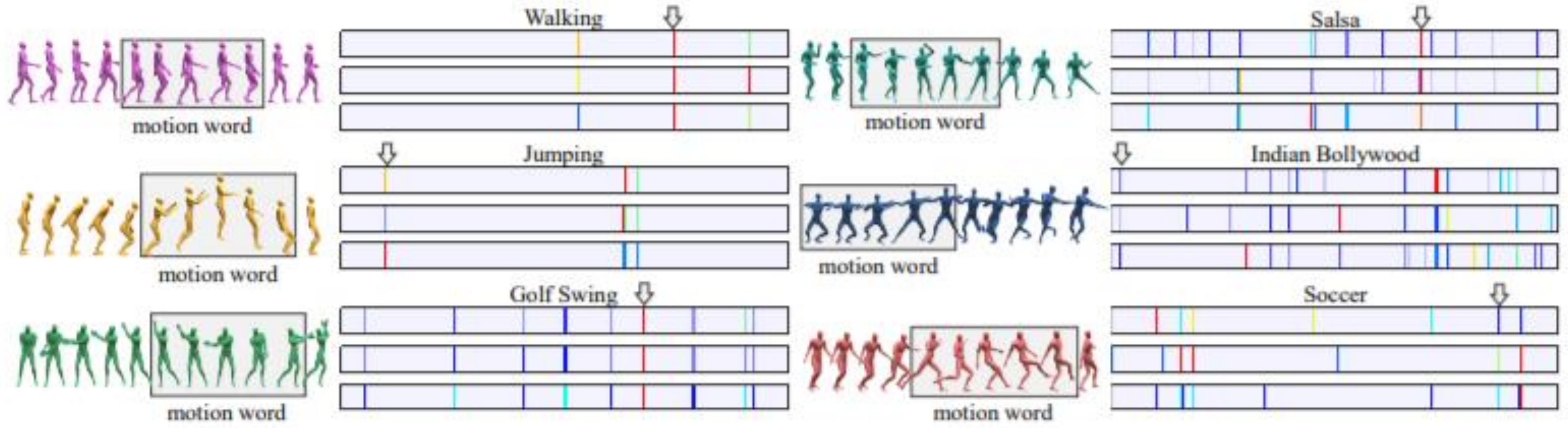


Fig. 1. Our motion signatures are defined using a deep analysis of motion words and selection of motion-motifs. Each signature is represented by a horizontal bar that shows the frequency of motion-motifs using color coding from red (high) through blue (low) to gray (zero). Note that the signatures represent distributions and not time evolution - the horizontal axis is not temporal. Three signatures of sequences are shown for each motion type – as can be seen, motions of similar type produce similar signatures where many motifs align. The rectangles in the sequence of motion to the left of the signatures illustrate motion words associated with the motifs shown by the corresponding arrow above the signature.

Many analysis tasks for human motion rely on high-level similarity between sequences of motions, that are not an exact matches in joint angles, timing, or ordering of actions. Even the same movements performed by the same person can vary in duration and speed. Similar motions are characterized by similar sets of actions that appear frequently. In this paper we introduce *motion motifs* and *motion signatures* that are a succinct but descriptive representation of motion sequences. We first break the motion sequences to short-term movements called motion words, and then cluster the words in a high-dimensional feature space to find motifs. Hence, motifs are words that are both common and descriptive, and their distribution represents the motion sequence. To cluster words and find motifs, the challenge is to define an effective feature space, where the distances among motion words are semantically meaningful, and where variations in speed and duration are handled. To this end, we use a deep neural network to embed the motion words into feature space using a triplet loss function. To define a signature, we choose a finite set of motion-motifs, creating a bag-of-motifs representation for the sequence. Motion signatures are agnostic to movement order, speed or duration variations, and can distinguish fine-grained differences between motions of the same class. We illustrate examples of characterizing motion sequences by motifs, and for the use of motion signatures in a number of applications.

CCS Concepts: • **Computing methodologies** → **Motion capture**; **Motion processing**;

Additional Key Words and Phrases: Animation, Motion Word, Motif, Motion Signature, Convolutional Network, Triplet Loss.

ACM Reference Format:
Andreas Aristidou, Daniel Cohen-Or, Jessica K. Hodgins, Yiorgos Chrysanthou, and Ariel Shamir. 2018. Deep Motifs and Motion Signatures. *ACM Trans. Graph.* 37, 06, Article 187 (November 2018), 13 pages. https://doi.org/10.1145/3272127.3275038

1 INTRODUCTION

The availability of human motion data in big repositories is growing with the emergence of simpler motion capture devices [Mehta et al. 2017; Pavlakos et al. 2017]. Content-based techniques and searching methods become essential to facilitate the use of such data. However, motion data is not always annotated or parameterized, hindering the semantic analysis of motions, the search in motion datasets, and the comparison between motion data. Working directly with the motion sequences is challenging due to the high-dimensional, temporal, nature of the motion, their large variations

Authors' addresses: Andreas Aristidou, The Interdisciplinary Center, Kanfei Nesharim, Herzliya, Israel, 4610101, a.aristidou@ieee.org; Daniel Cohen-Or, Tel-Aviv University, Ramat Aviv, Tel-Aviv, Israel, 6997801, cohenor@gmail.com; Jessica K. Hodgins, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, USA, PA 15213, jkh@cs.cmu.edu; Yiorgos Chrysanthou, University of Cyprus & RISE Research Center, 75, Kallipoleos, Nicosia, Cyprus, 1678, yiorgos@cs.ucy.ac.cy; Ariel Shamir, The Interdisciplinary Center, Kanfei Nesharim, Herzliya, Israel, 4610101, arik@idc.ac.il.

0730-0301/2018/11-ART187 $15.00
https://doi.org/10.1145/3272127.3275038

Introduction

Contextual motion analysis

https://youtu.be/weSvQCGuTvU

Dance is “a performing-art form consisting of purposefully selected and controlled rhythmic sequences of human movements”. These movements have aesthetic and often symbolic value.

S. H. Fraleigh, Dance and the Lived Body: A Descriptive Aesthetics.
University of Pittsburgh Press, 1987.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Dance is "a performing-art form consisting of purposefully selected and controlled rhythmic sequences of human movements". These movements have aesthetic and often symbolic value.

S. H. Fraleigh, Dance and the Lived Body: A Descriptive Aesthetics. University of Pittsburgh Press, 1987.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

The premise of our work

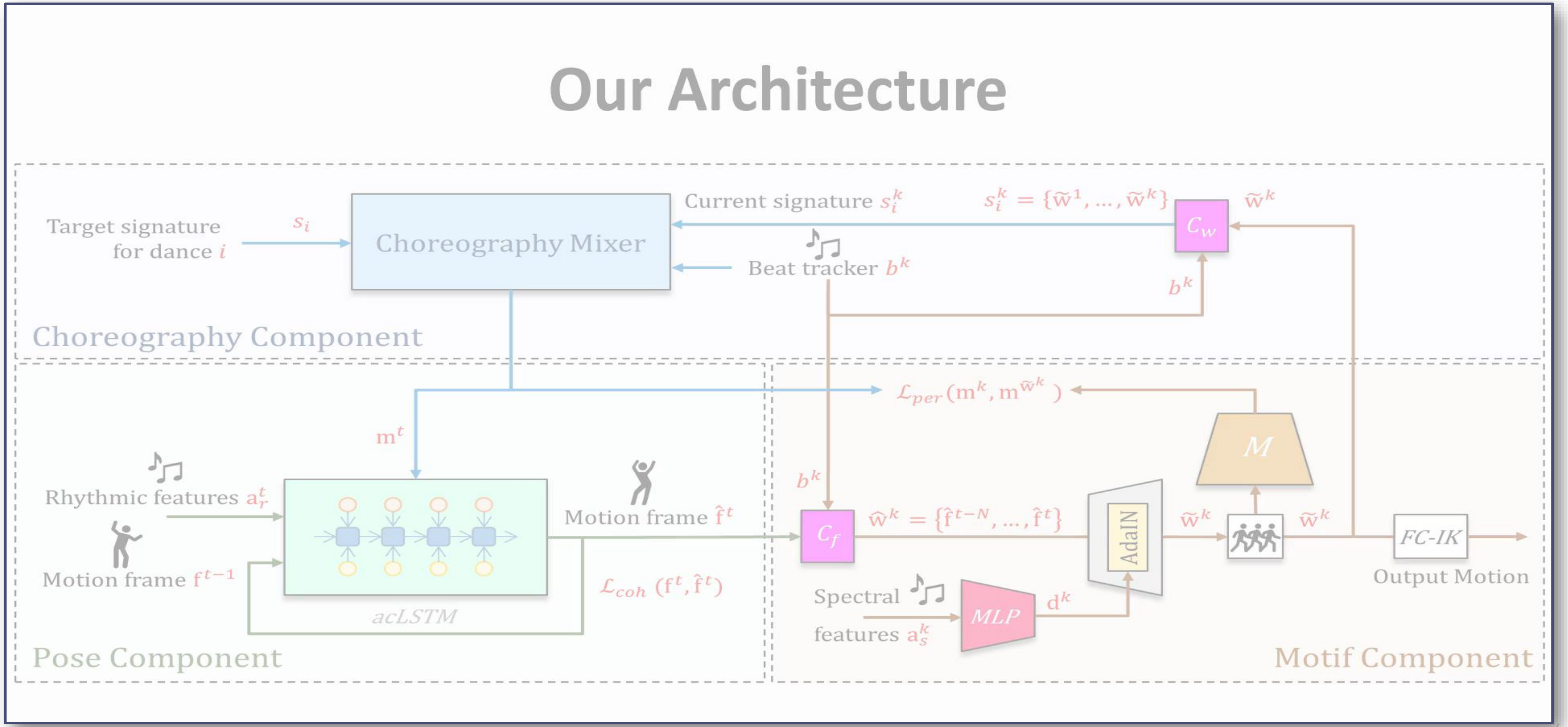
Music-driven motion synthesis

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Our network

Music-driven motion synthesis

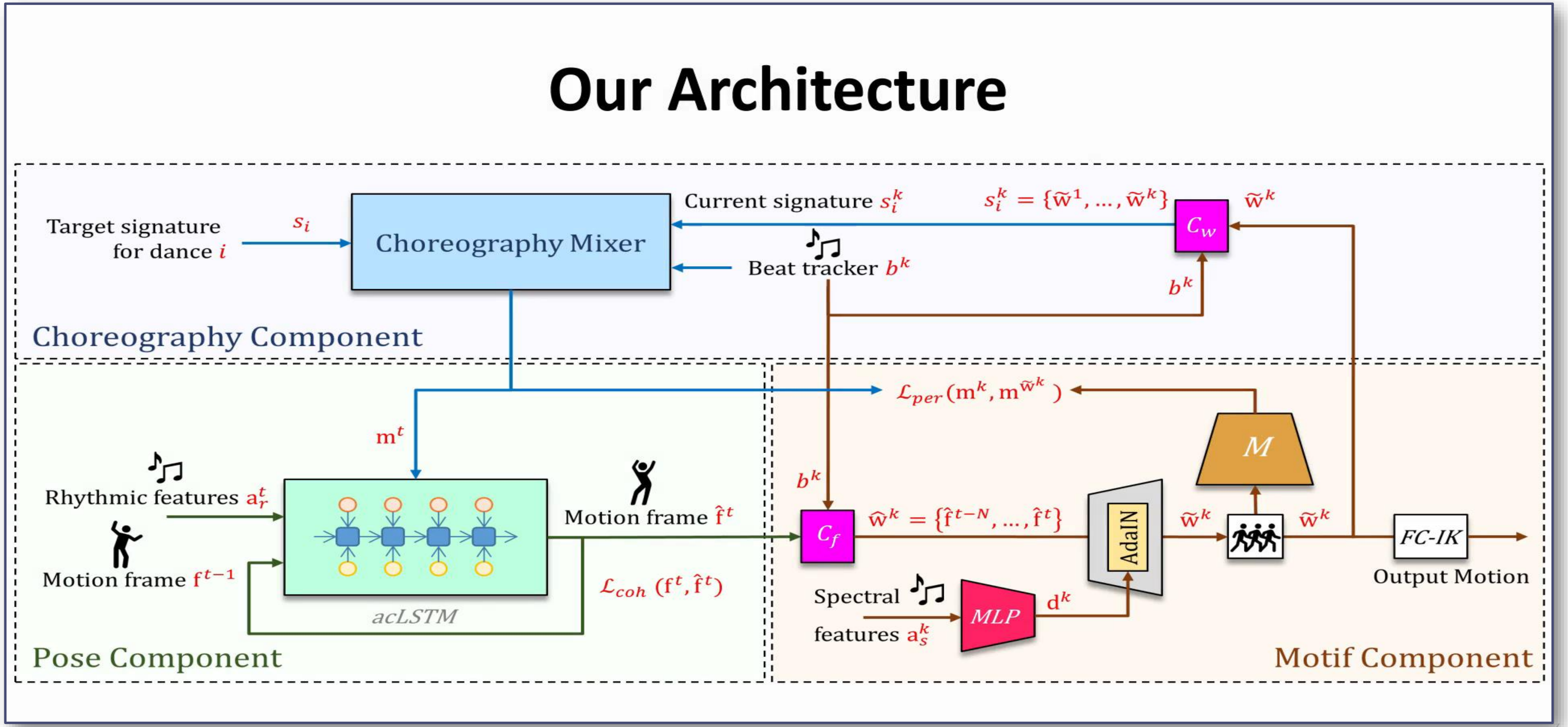


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Our network: *Pose component*

Music-driven motion synthesis

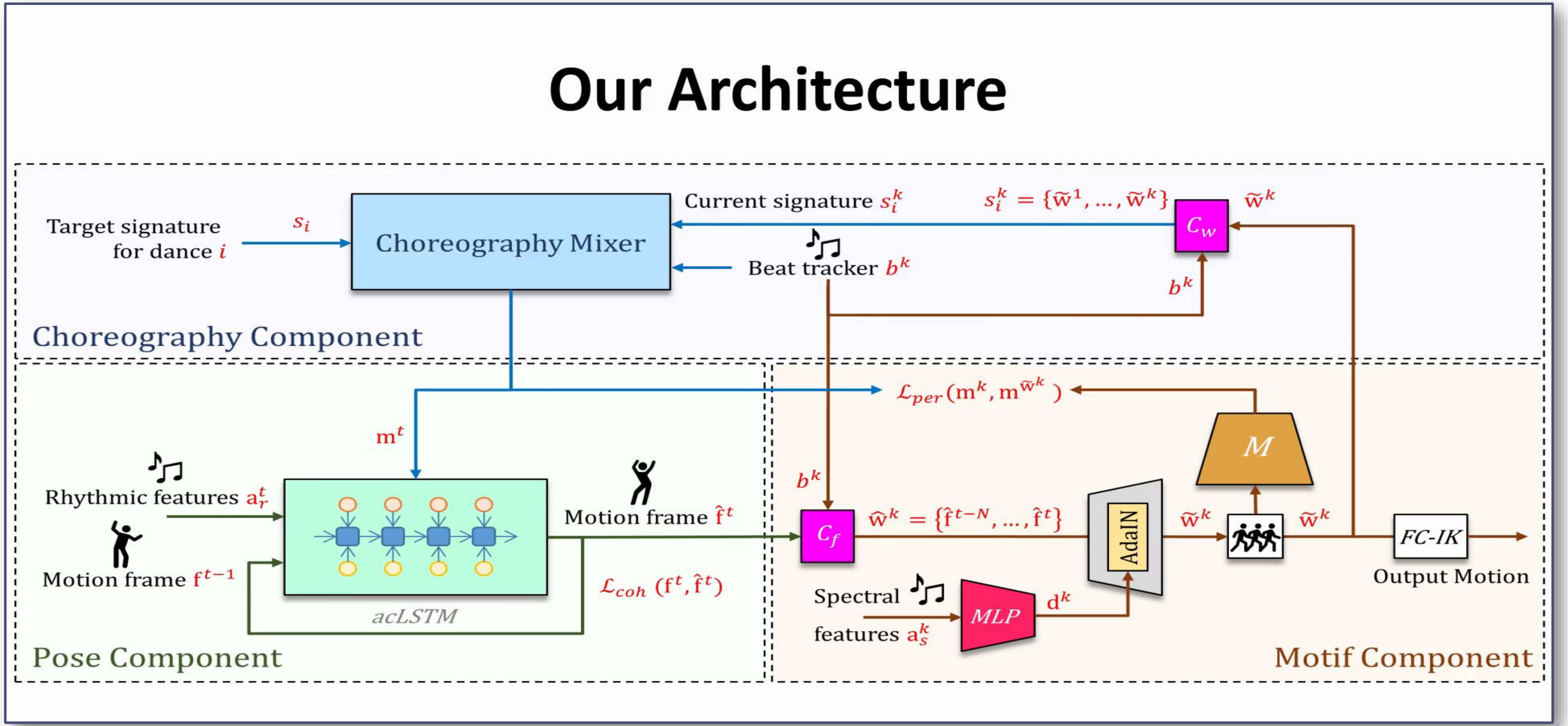


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Our network: *Motif component*

Music-driven motion synthesis

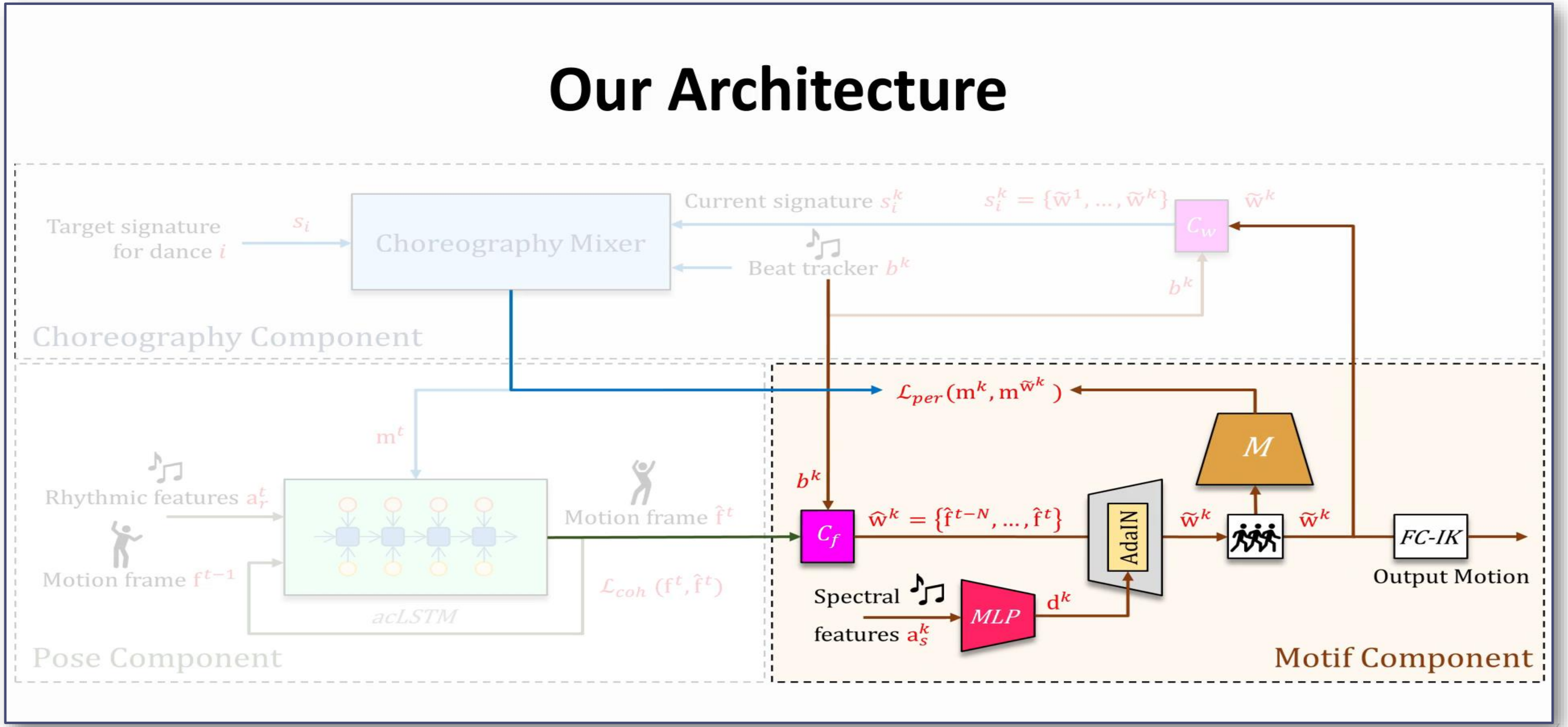


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Our network: *Motif component*

Music-driven motion synthesis

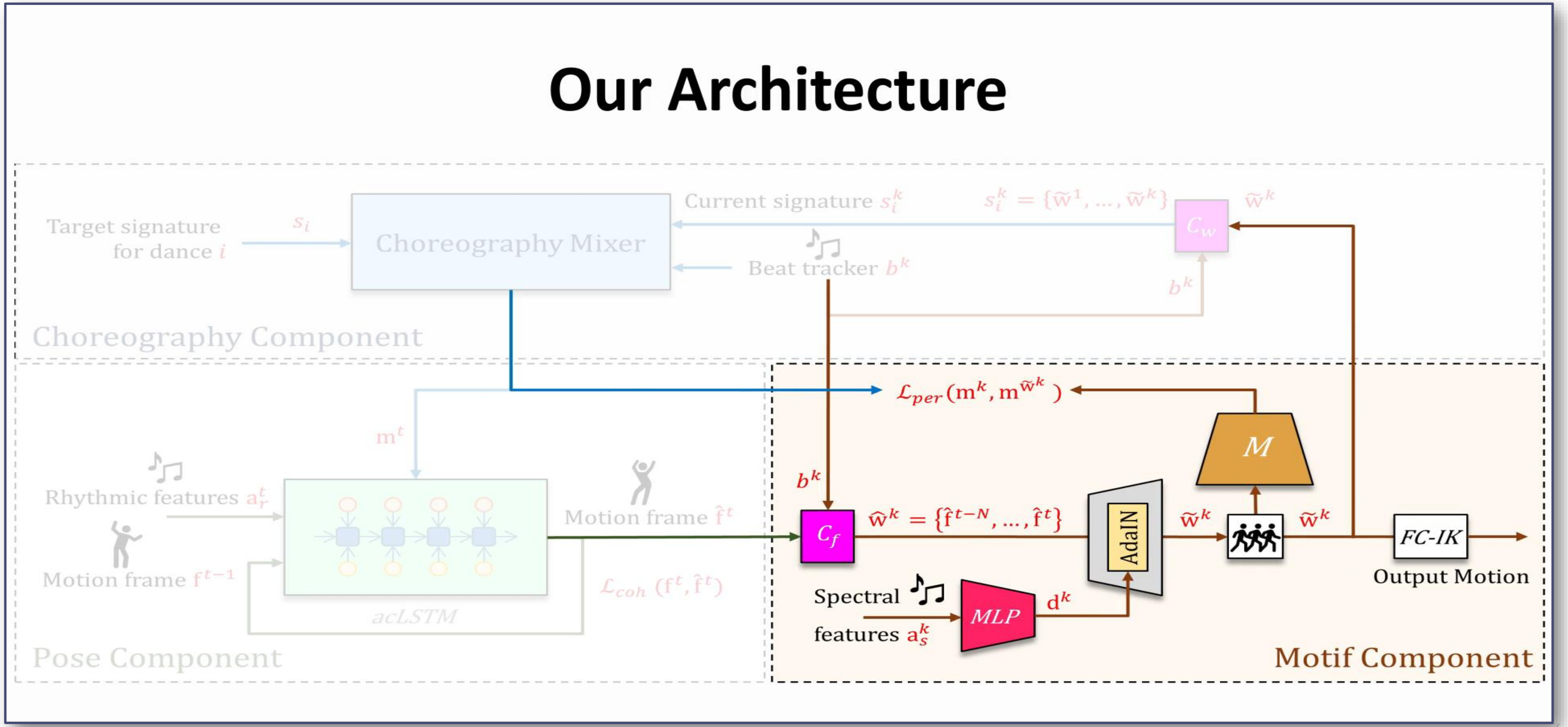


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Our network: *Motif component*

Music-driven motion synthesis



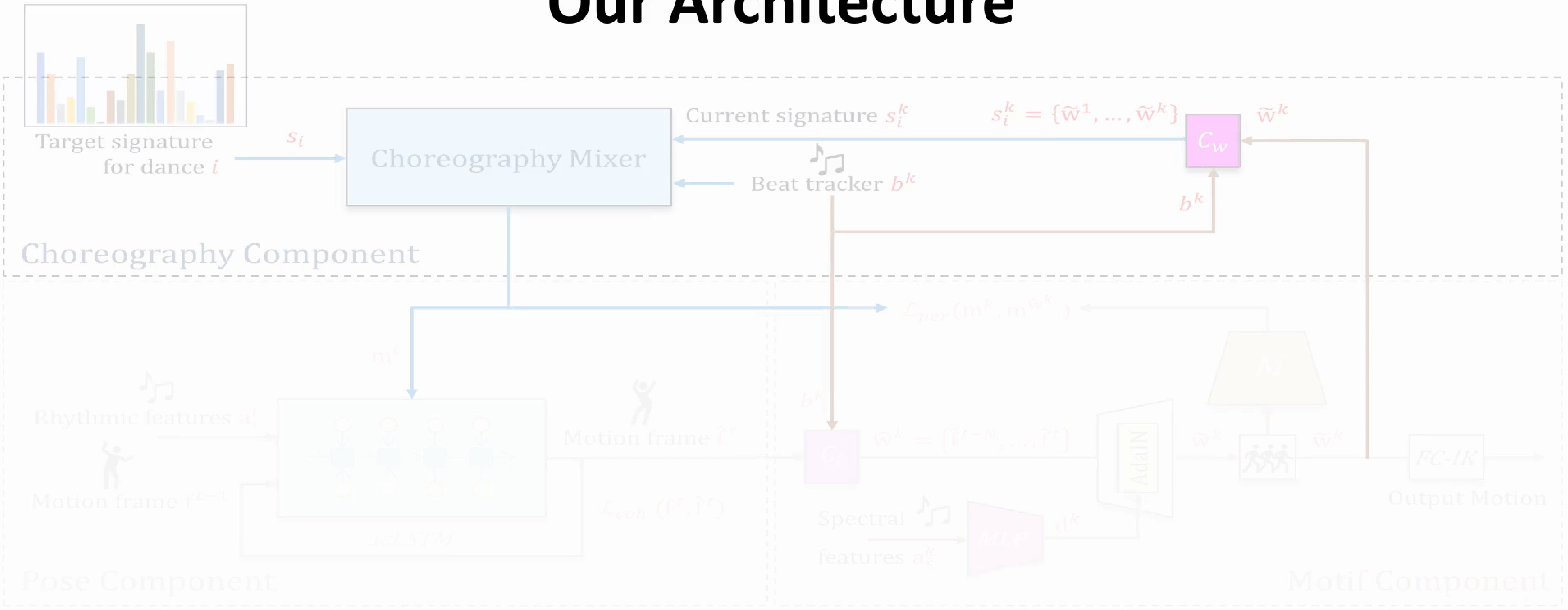
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Our network: *Choreography component*

Music-driven motion synthesis

Our Architecture



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Our network: *Losses and other important parameters*

Music-driven motion synthesis

- Audio representation (Librosa Library [Ellis 2007]):
  - Rhythmic features $\mathbf{a}_r^t \in \mathbb{R}^4$
  - Spectral features $\mathbf{a}_s^t \in \mathbb{R}^{87}$
- Pose representation: $\mathbf{f}^t = [\mathbf{f}_t, \mathbf{f}_q] \in \mathbb{R}^{3+4J}$
  - the root displacement $\mathbf{f}_t \in \mathbb{R}^3$
  - joint rotations in unit quaternions, $\mathbf{f}_q \in \mathbb{R}^{4J}$, for $J = 31$ joints
- Motif representation:
  - $\mathbf{m}^t \in \mathbb{R}^d$, where $d = 184$ universal features
  - motion words are segmented on the beat; time-scaled to 13 frames

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Music-driven motion synthesis

- The input to the network at time t is:

$$\mathbf{n}^t = [\mathbf{a}_r^t, \mathbf{m}^t, \mathbf{f}^t, \mathbf{c}^t] \in \mathbb{R}^{4+d+4J+2}$$

- where $\mathbf{c}^t \in \{0,1\}^2$ is a binary vector representing the left and the right foot contact labels

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Music-driven motion synthesis

- **Foot Sliding Cleaning (pose level)**
  - predict foot contact labels
- **Motion Diversity (motif level)**
  - AdaIN layer to inject style variation using $\mathbf{a}_s^t$
- **Motion Perceptual-Loss (motif level)**
  - controls the content of motion words
- **Motif Transition matrix (choreography level)**
  - describes probability of the temporal connectivity between consecutive motion motifs
- **Signature difference (choreography level)**
  - compares the current signature to the target signature

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results

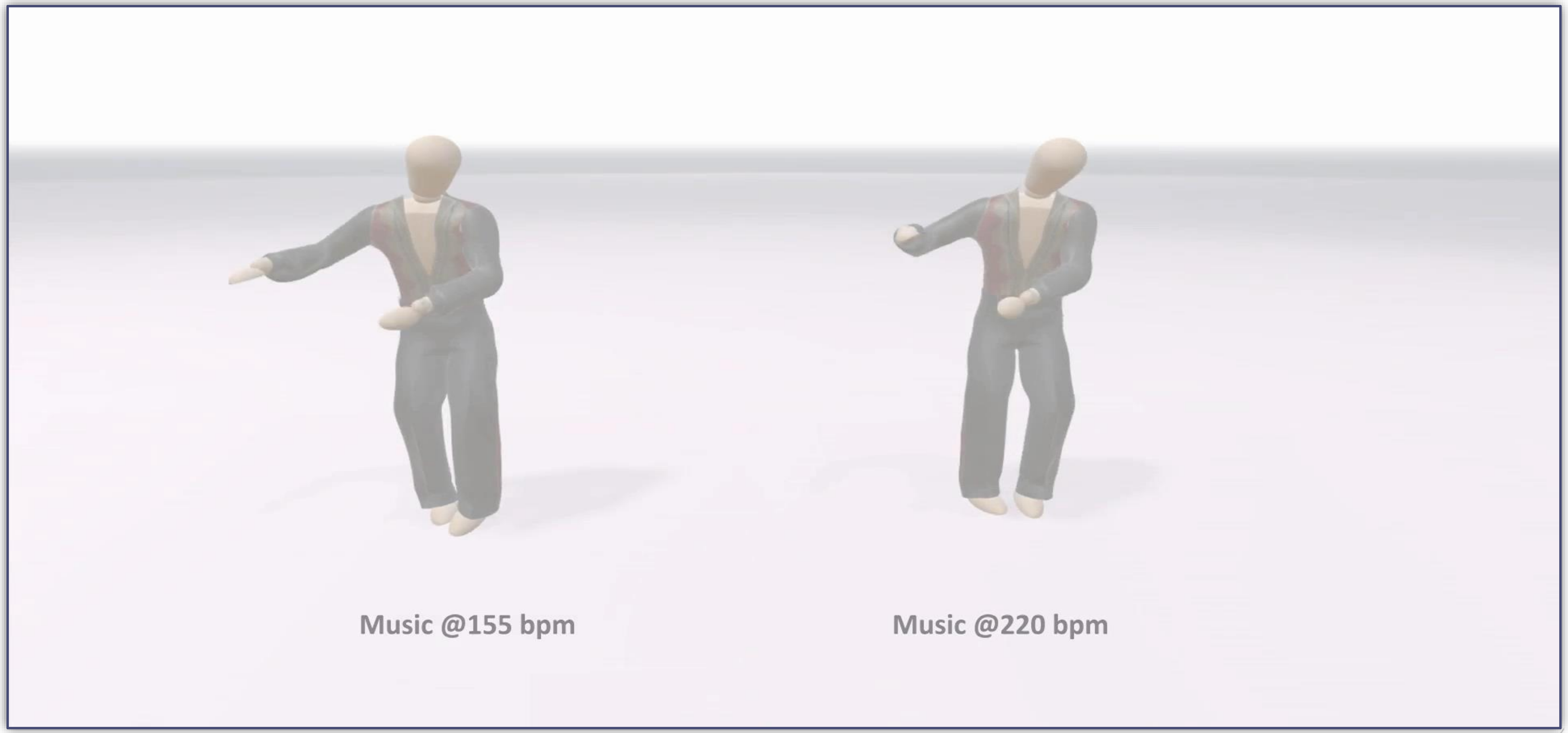
Music-driven motion synthesis

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results

Dance synthesis at different bpm

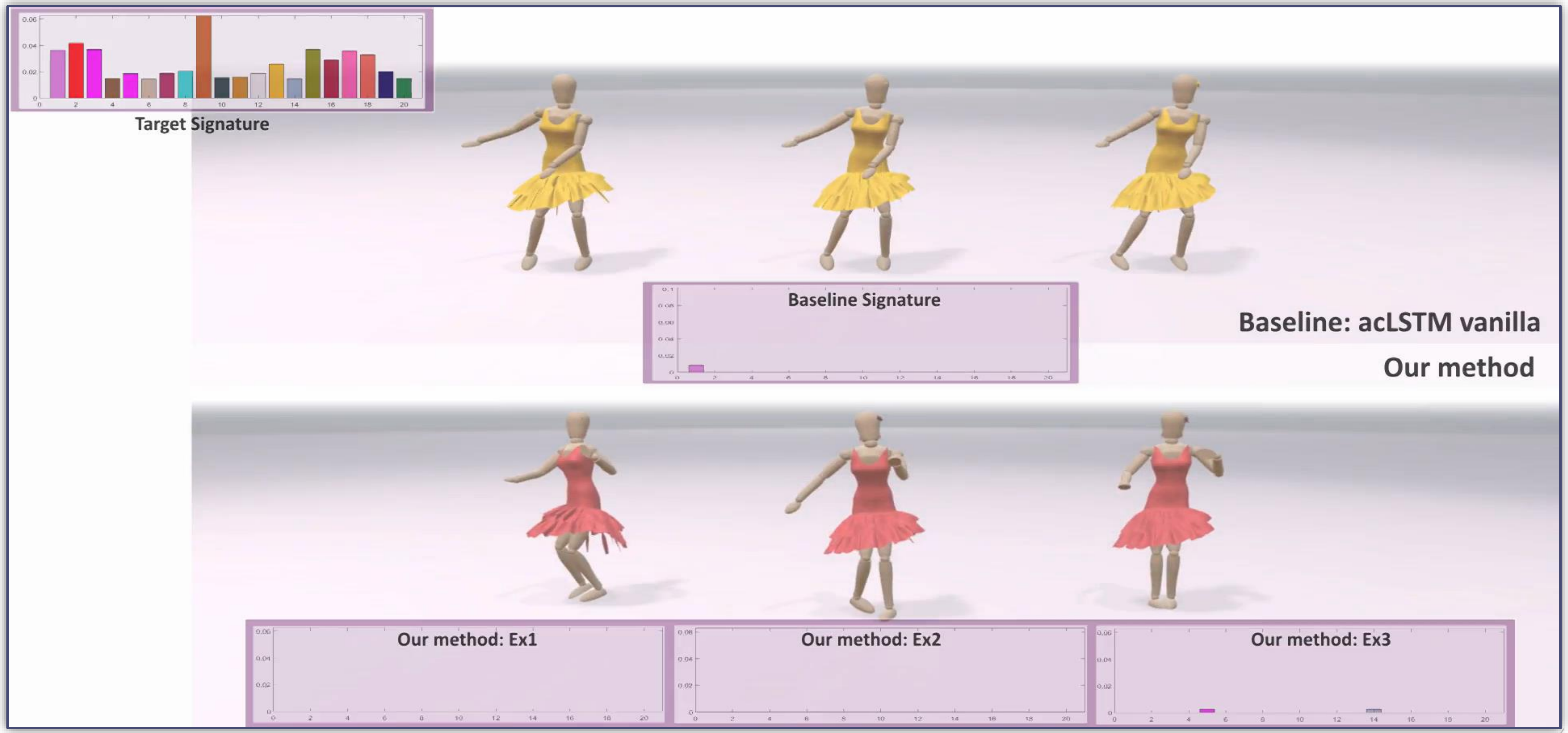


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results

Dance synthesis with variation



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Results: *Ablation study*

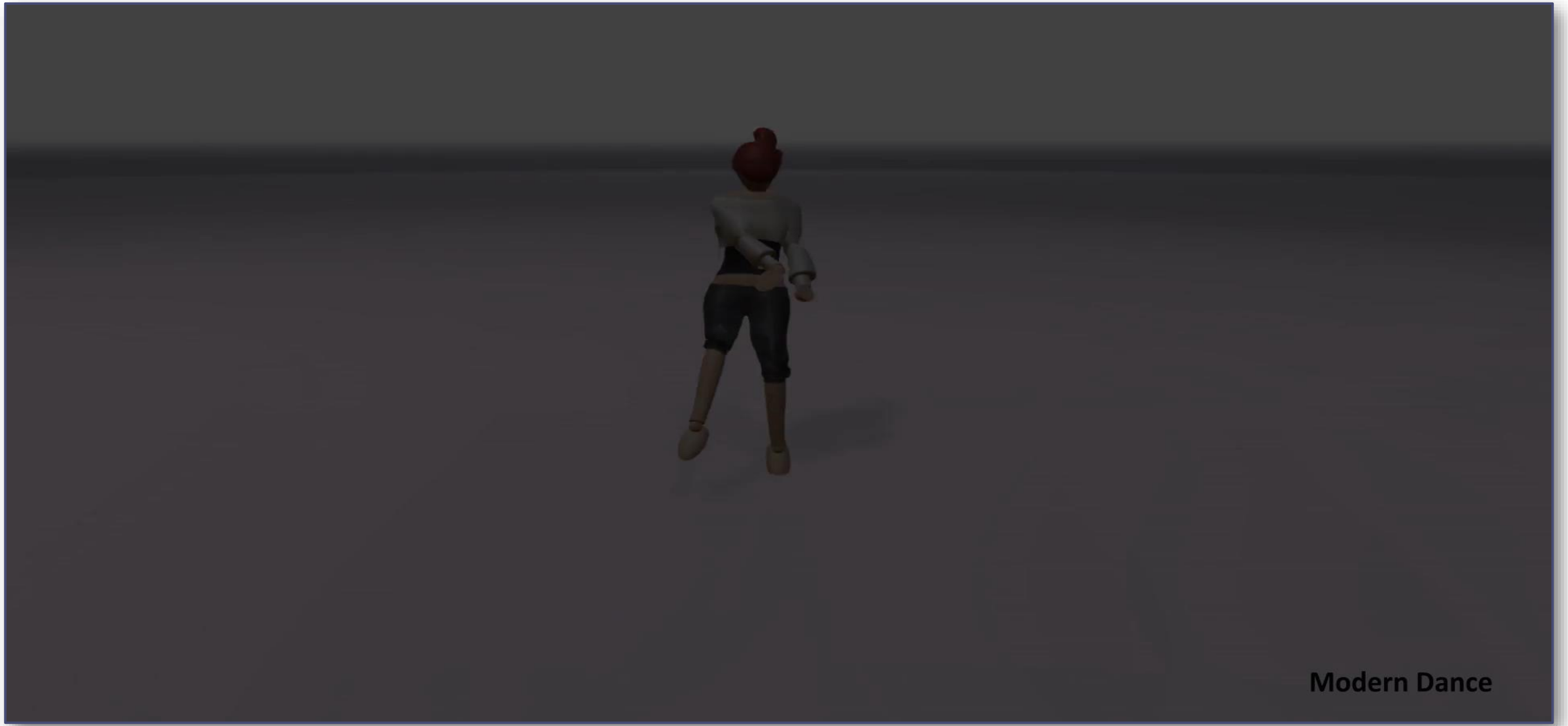
Spectral audio for subtle variations

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

More Results

Music-driven motion synthesis

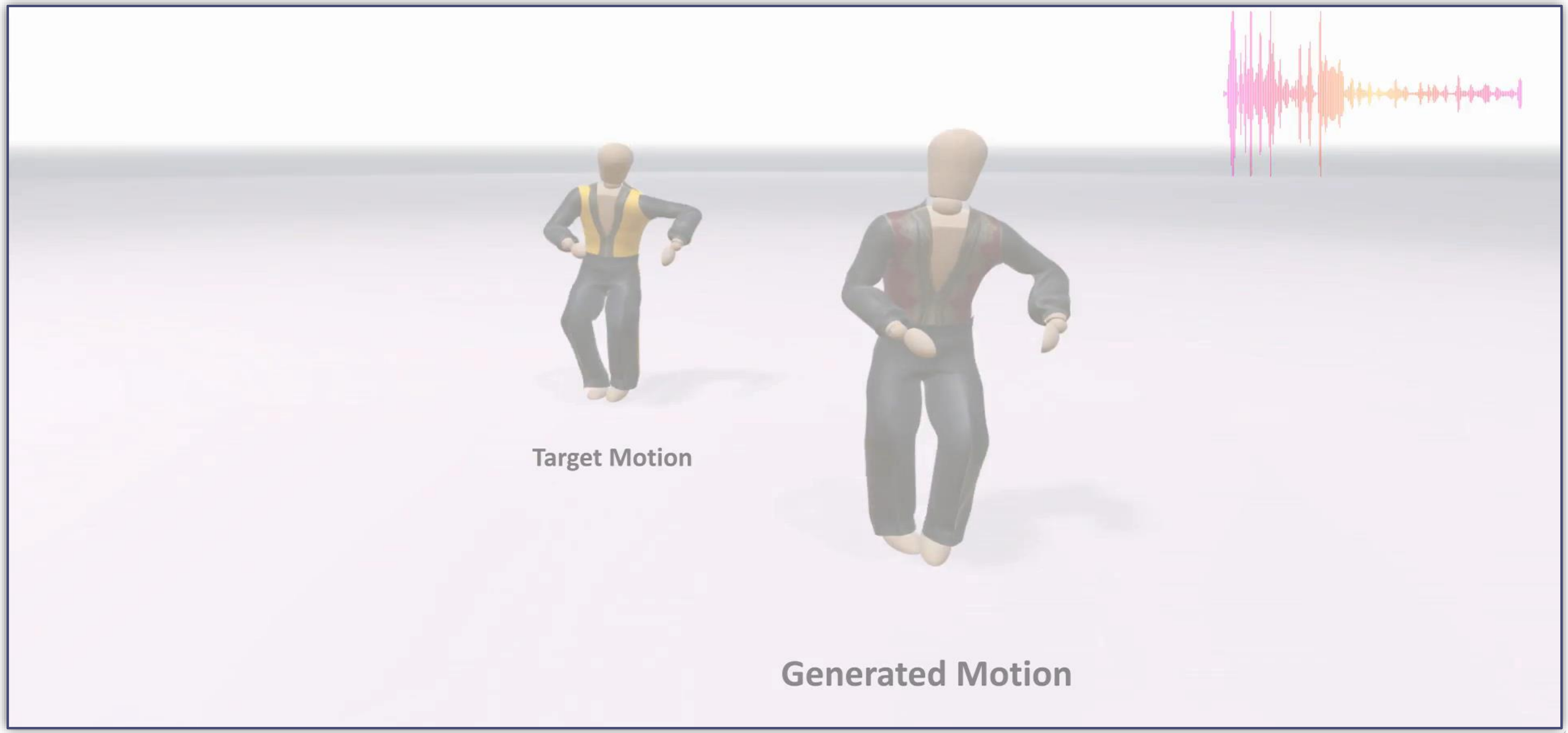


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Applications

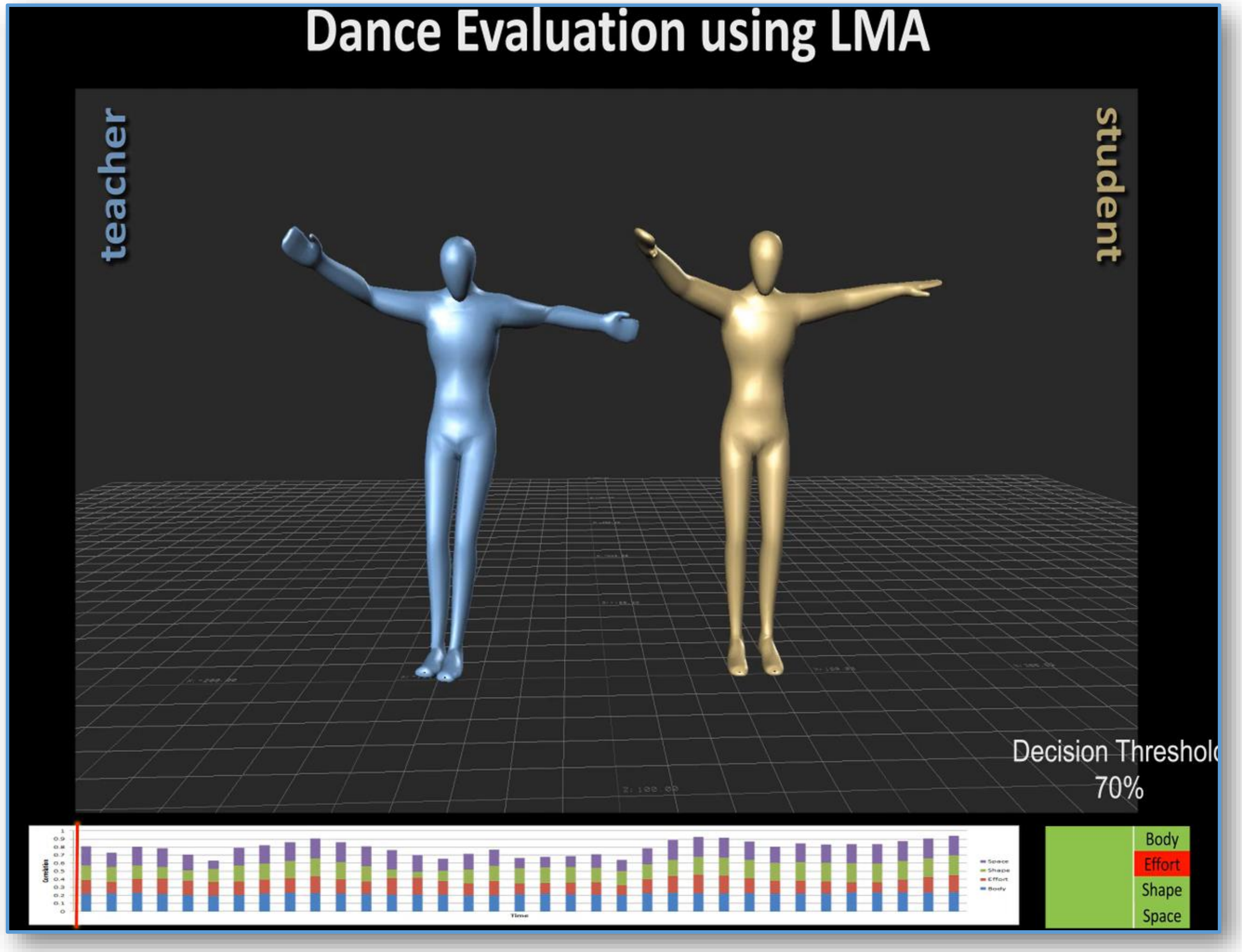
Recreate an existing dance

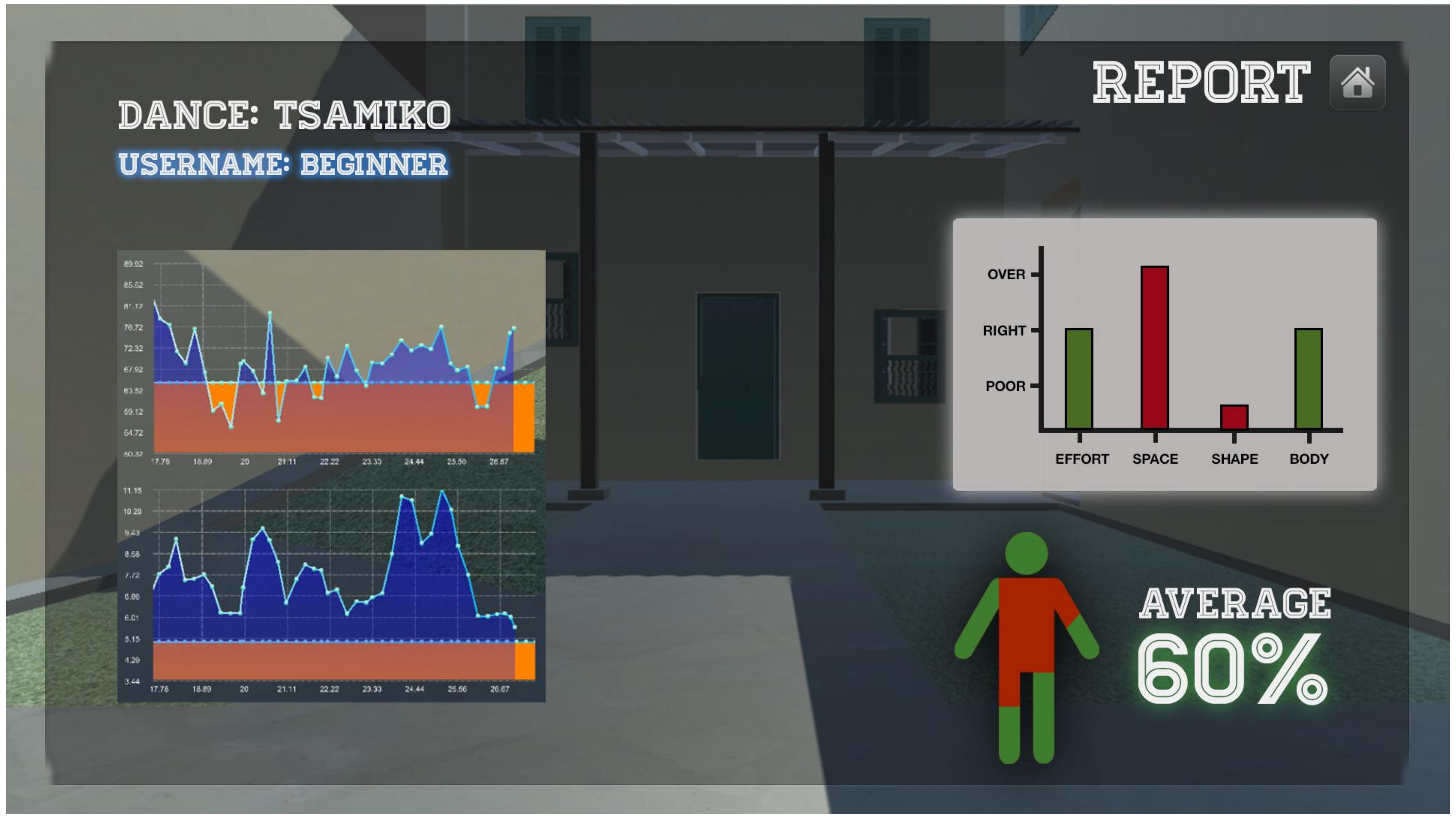


Co-financed by the European Union
Connecting Europe Facility

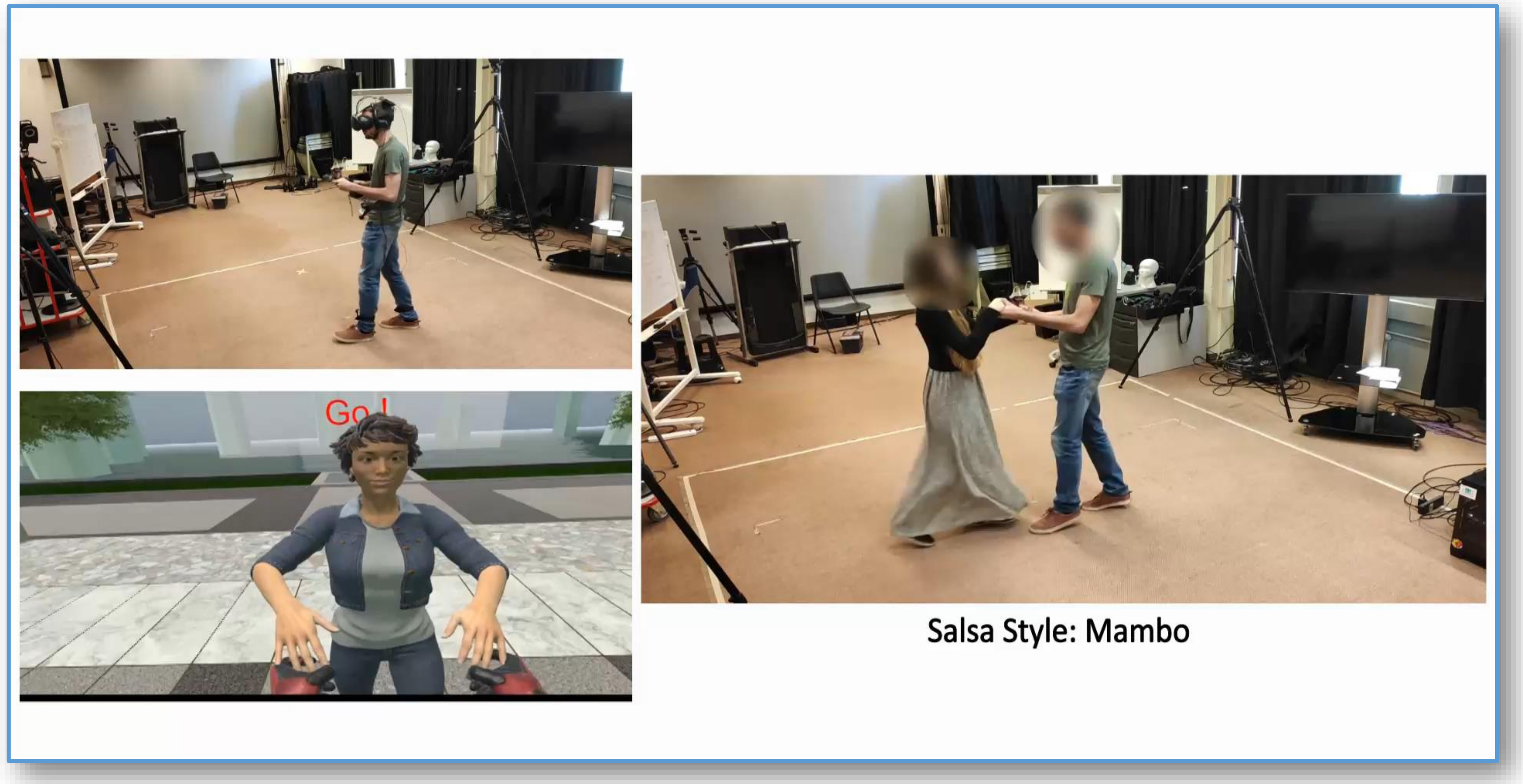
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Other Applications

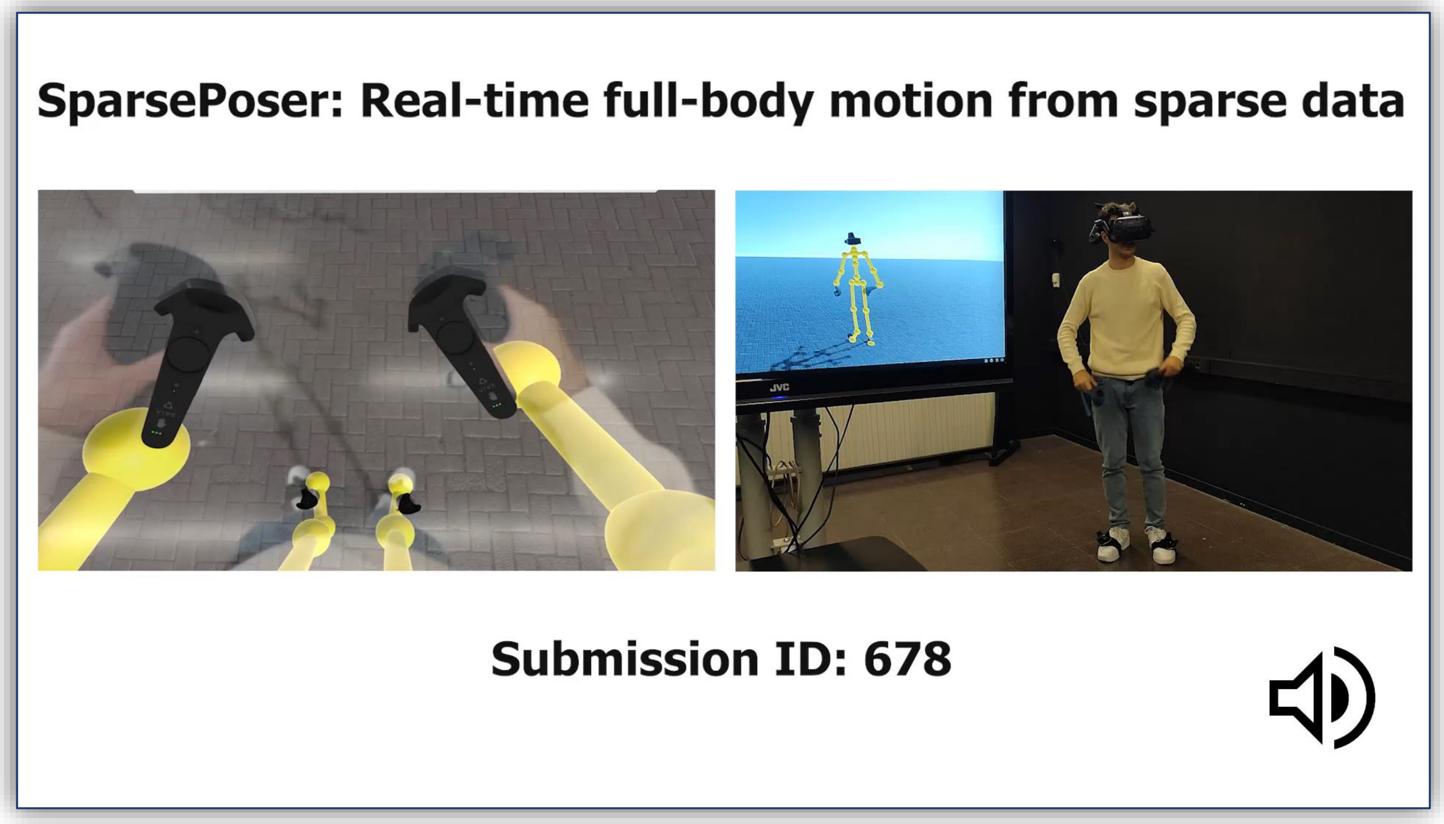


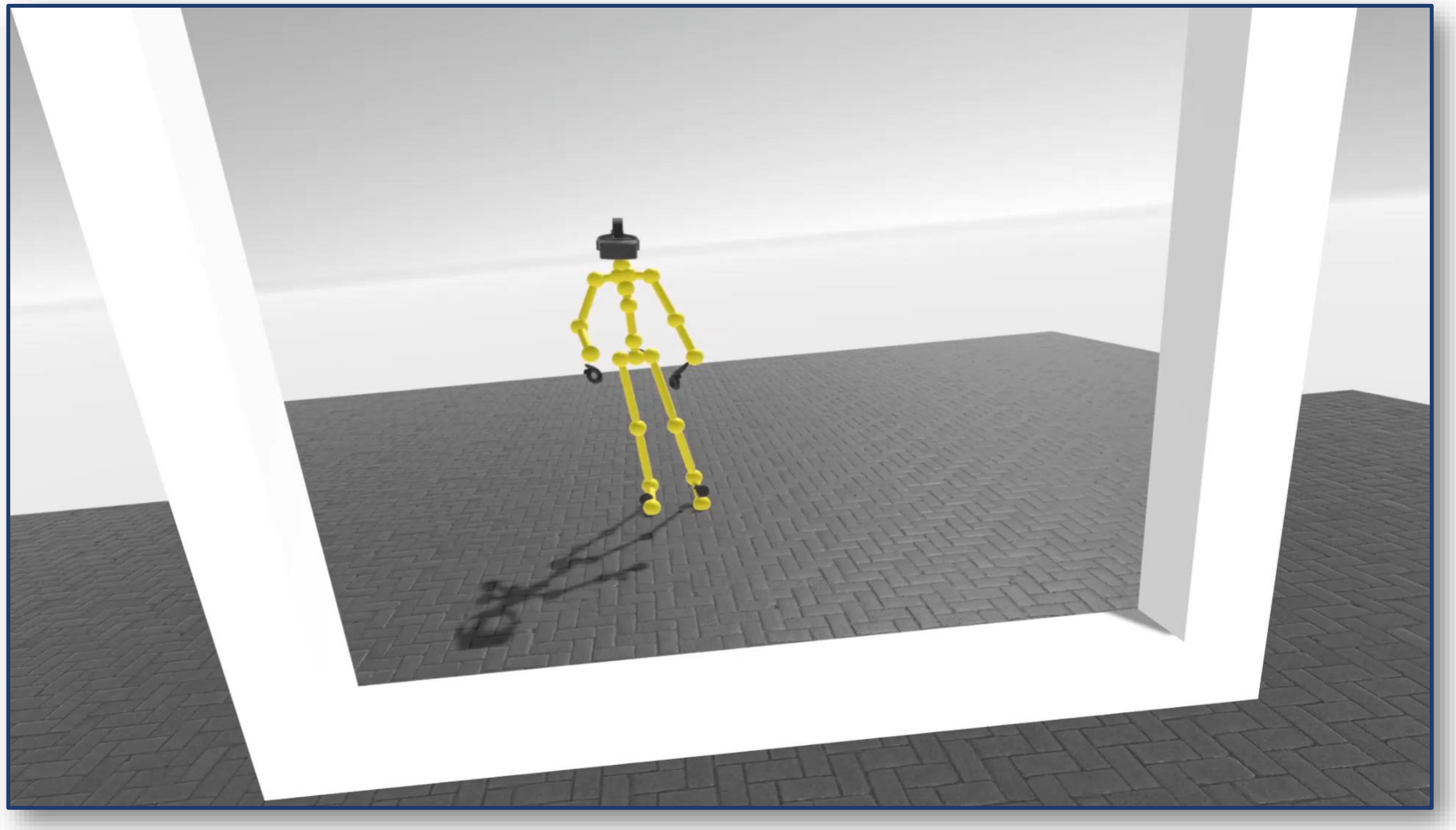


Other Applications



Other Applications





Other Applications

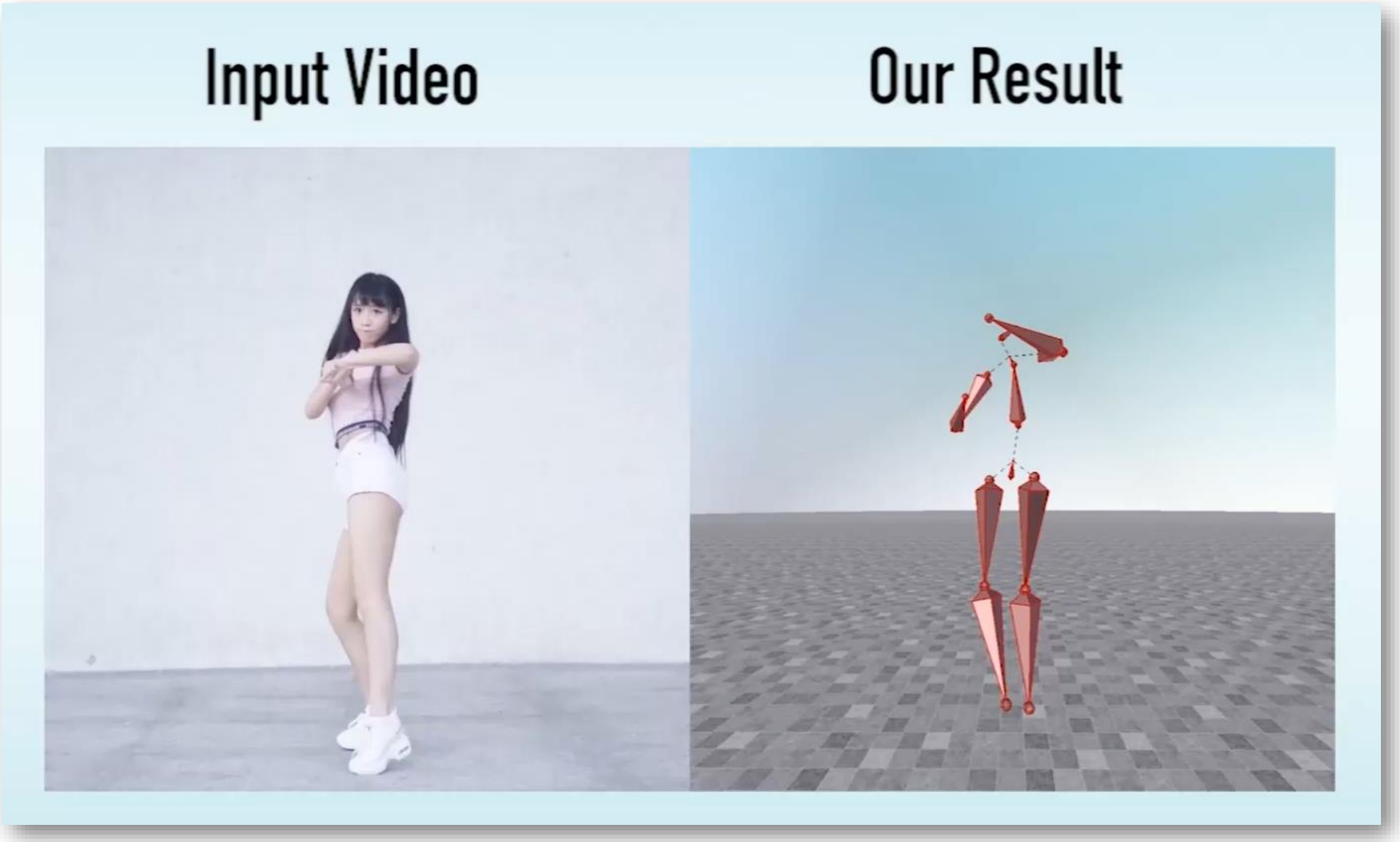
Dance Central 4 - Shape of You

Just Dance® 2019 - Me Me Me

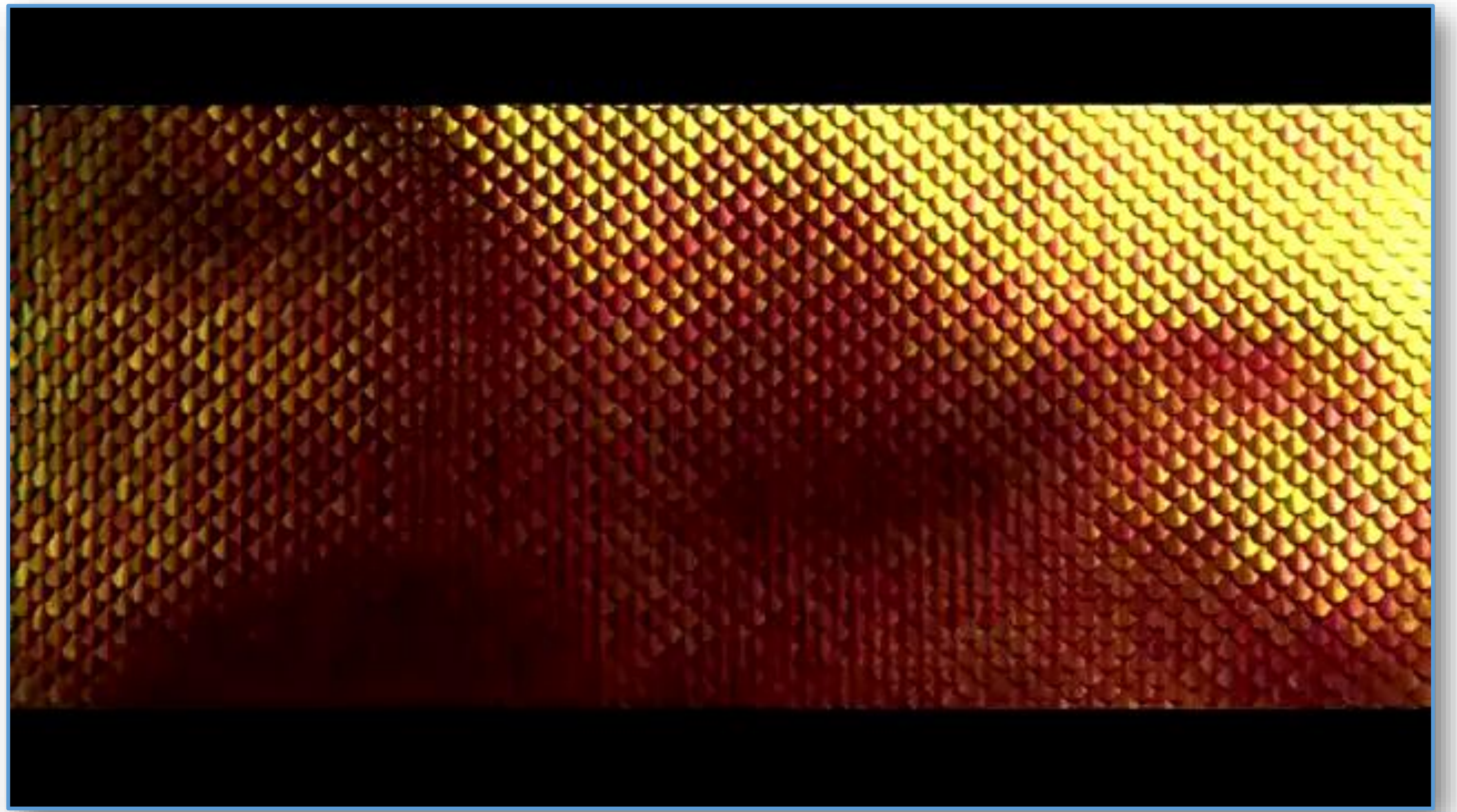
Other Applications

CYENS
CENTRE OF EXCELLENCE

北京电影学院
BEIJING FILM ACADEMY



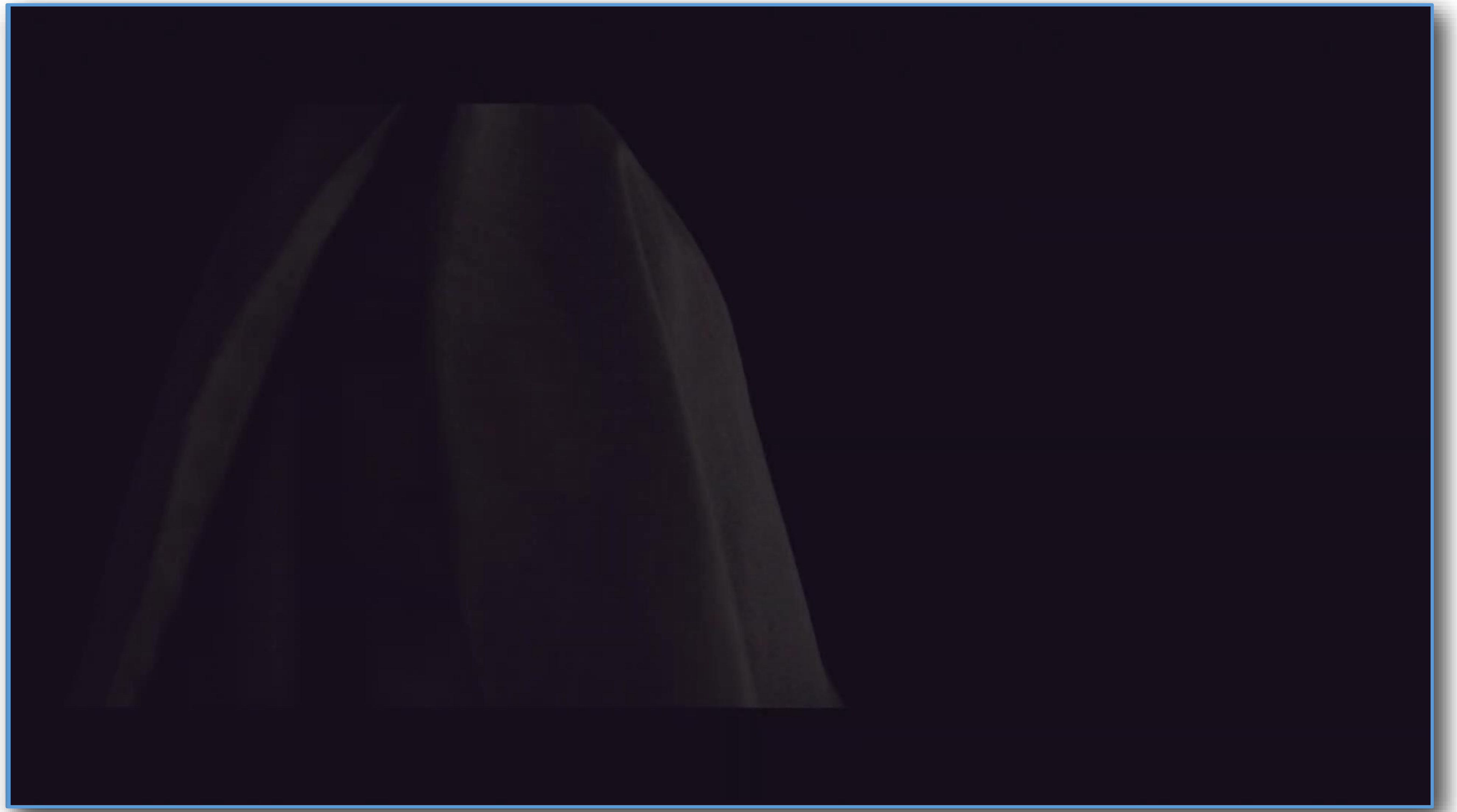
Other Applications



AICP sponsor reel by Method Studios
https://youtu.be/fd_9qwpzVBQ

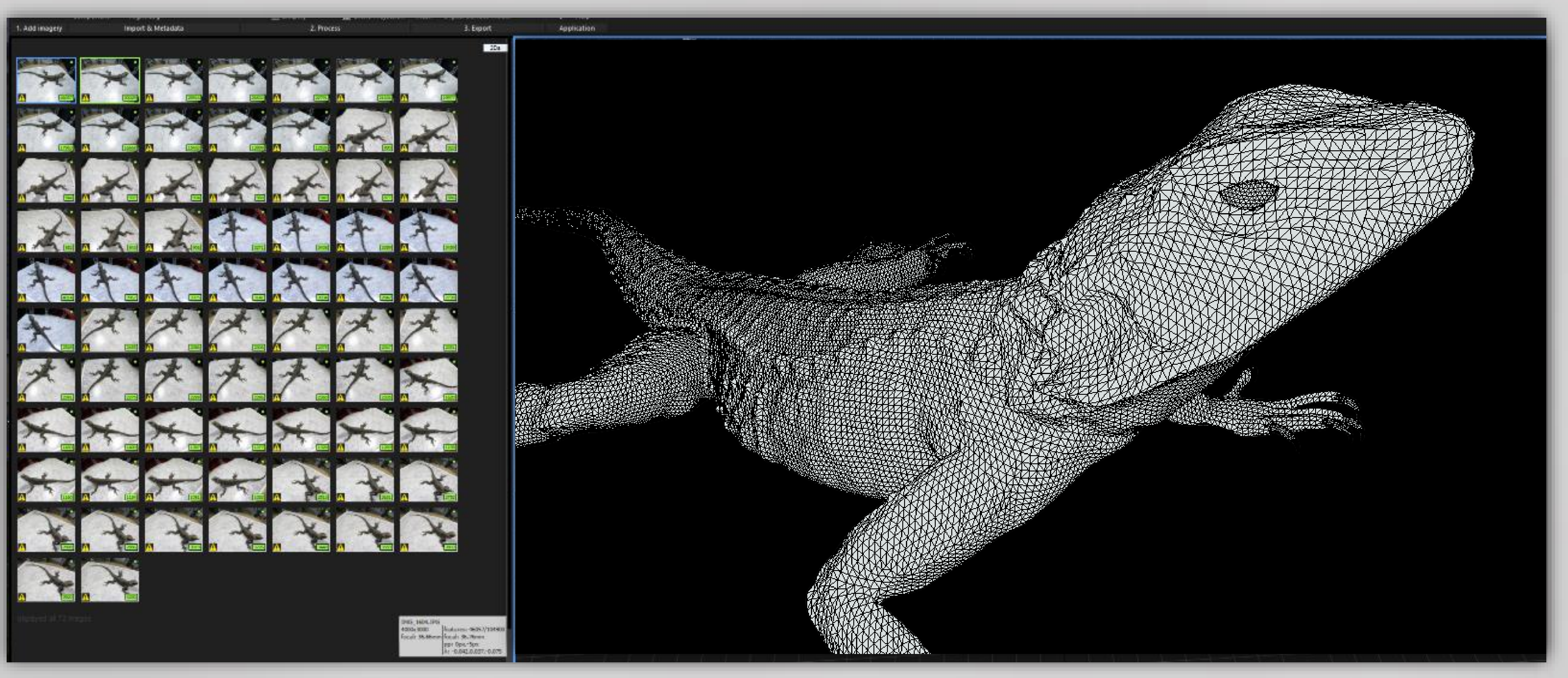
Dancing Phantoms by Kiyan Forootan
https://youtu.be/Ig7A6fZrWyM

Other Applications

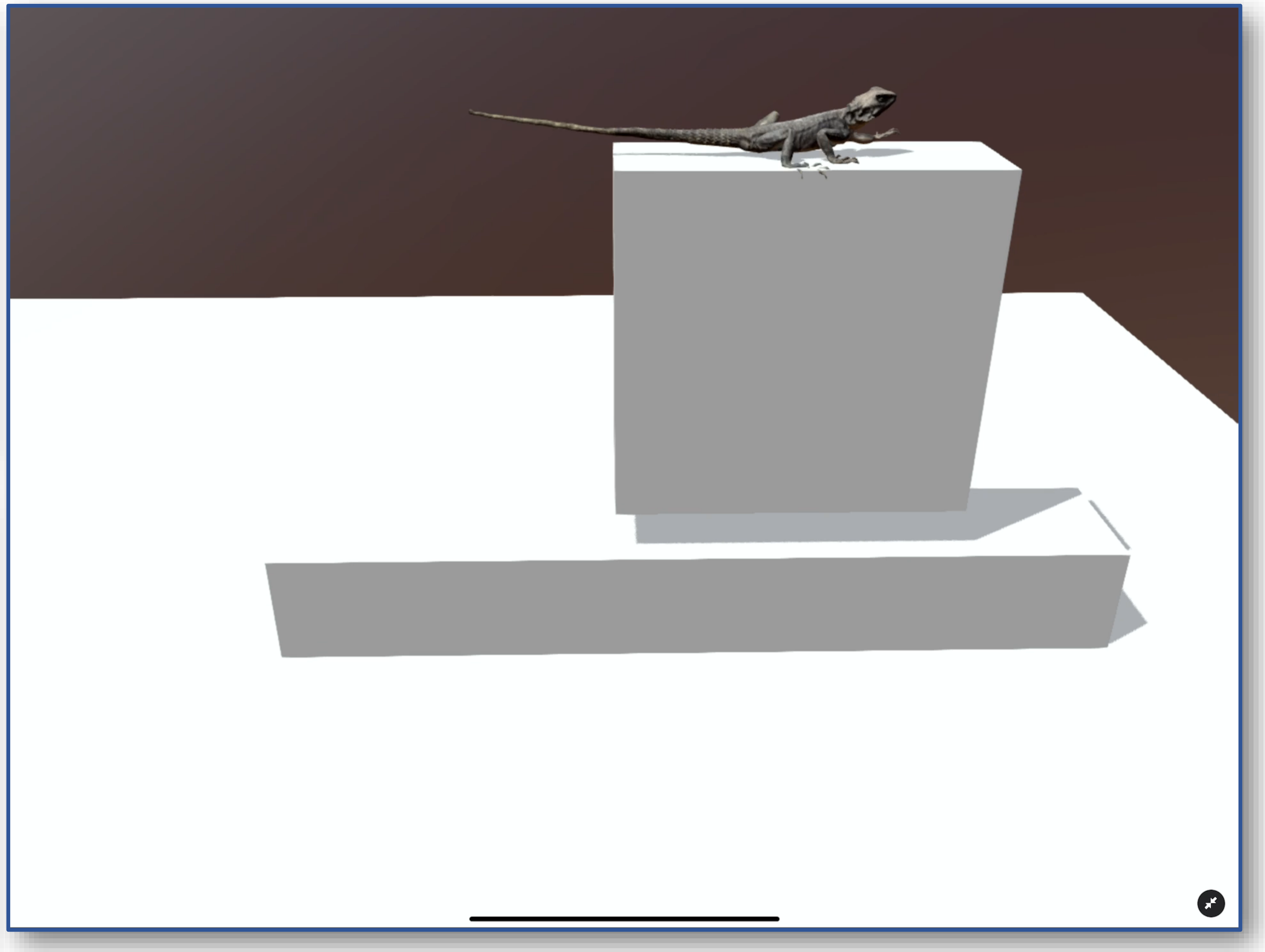


PARKER • Become the Fool
https://youtu.be/5oJUfpB4f90

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423


1. Add imagery
Import & Metadata
2. Process
3. Export
Application

Other Applications



Other Applications

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Join our team at the *Graphics & Extended Reality Lab*

Andreas Aristidou
Assistant Professor

Office: FST01, Room B113

Office hours: Only after appointment

email: andarist@ucy.ac.cy

Research Interests:

Machine Learning, Deep Learning and its applications in Computer Graphics and Character Animation, Virtual/Augmented Reality, Digital Heritage

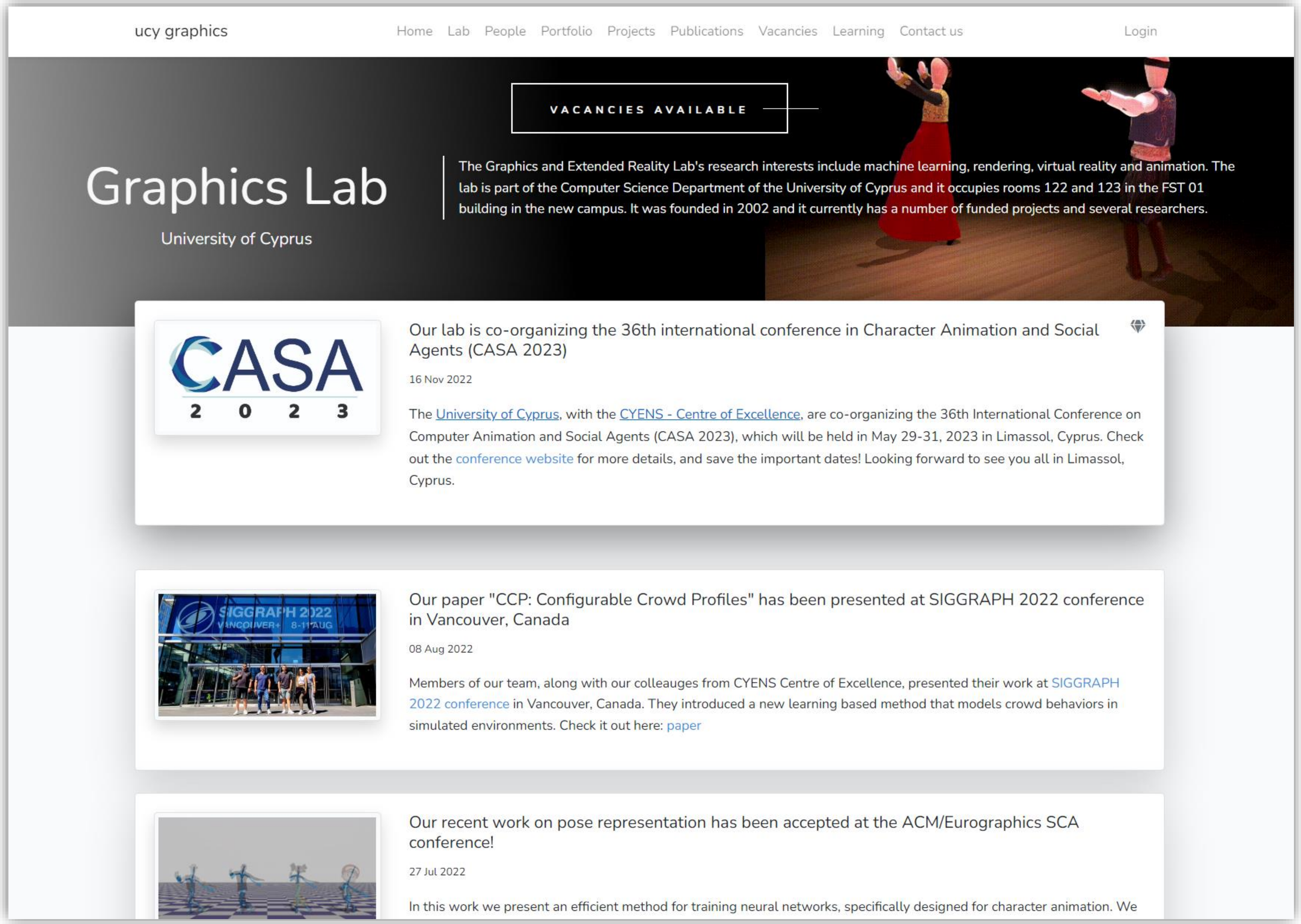
https://www.cs.ucy.ac.cy/~andarist

GRAPHICS & EXTENDED REALITY LAB

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Join our team at the *Graphics & Extended Reality Lab*



The **Graphics and Extended Reality Lab** at the University of Cyprus, part of the Computer Science Department, conducts research in areas such as machine learning, rendering, virtual reality and animation.

Founded in 2002, the lab is located in rooms 122 and 123 of the FST 01 building on the university's new campus, and is staffed by two faculty members and twelve research associates. It also has several active, funded projects.

Website: https://graphics.cs.ucy.ac.cy/

Paper Presentations

- A Style-Based Generator Architecture for Generative Adversarial Networks
- DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation
- GANimator: Neural Motion Synthesis from a Single Sequence
- Human Motion Diffusion as a Generative Prior

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thank you!

That’s all folks!!!

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423