Master Thesis

3D human interaction synthesis for action recognition data augmentation

Anders Bredgaard Thuesen

March 7, 2024

Abstract

Introduction

Methods

SMPL & Human Mesh Reconstruction

The Skinned Multi-Person Linear (SMPL) model is a parametric body shape model that accurately represents a wide range of human bodies and poses. It is built upon a foundation of linear blend skinning enhanced with corrective blend shapes, which are derived from a large dataset of body scans. The model captures the subtle deformations that occur with different body shapes and poses and can easily be rendered due to its compatibility with existing graphics pipelines. Since its publication, several extensions such as DMPL, incorporating dynamic soft-tissue deformation and SMPL-X, also modelling hands and facial expressions have been introduced. The model is parameterized by $\vec{\beta}$, capturing the variations from a mean body shape and $\vec{\theta}$, specifying the axis-angle rotation of 23 of the template skeleton joints. Mathematically, the model can be expressed as:

$$M(\vec{\beta}, \vec{\theta}) = W(T_P(\vec{\beta}, \vec{\theta}), J(\vec{\beta}), \vec{\theta}, \mathcal{W}) \tag{1}$$

where $T_P(\vec{\beta}, \vec{\theta})$ returns the vertices of the rest pose, incorporating the deformations from the body shape and pose and is given by:

$$T_P(\vec{\beta}, \vec{\theta}) = \bar{\mathbf{T}} + B_S(\vec{\beta}) + B_P(\vec{\theta}) \tag{2}$$

$J(\vec{\beta})$ returns the 3D joint locations from the shaped template vertices using a learned regression matrix $\mathcal{J}$ and is given by:

$$J(\vec{\beta}) = \mathcal{J}(\bar{\mathbf{T}} + B_S(\vec{\beta})) \tag{3}$$

W is the skinning function (e.g. Linear Blend Skinning (LBS) or Dual-Quaternion Blend Skinning (DQBS)) and $\mathcal{W}$ is the blend weights.

Tracking & Matching

To track the prediction over time we naively compare each prediction point-wise with the latest track of people in the scene according to the following loss function;

$$\mathcal{L}_{\text{match}}(a, b) = \alpha \left\| a_{\text{3D kpts}} - b_{\text{3D kpts}} \right\|_2 + \beta \left\| a_{\text{class}} - b_{\text{class}} \right\|_\infty , \tag{4}$$

incorporating both the Euclidean distance between the 3D joint keypoints as well as the predicted patient class, with α and β weighting the influence of each term. The predictions are then greedily assigned to the tracks and new tracks are created for unassigned detections.