**Overview**

HiFECap is a novel neural human performance capture approach capturing human poses, facial expressions, and hand gestures, and high-frequency details, such as deforming wrinkles on the clothes, simultaneously just from a single RGB video.

- The first monocular 3D human performance capture approach enabling joint tracking of body pose, the non-rigidly deforming surface, hand gestures, and facial expressions.
- A visibility- and rigidity-aware vertex displacement network to enable the capture of high-frequency geometric details of the dynamic human surface.
- A multi-stage training process for surface recovery and a face and hand model integration.

**Method**

- **Input:** a single segmented image
- **Output:** the corresponding 3D human mesh.
  - **PoseNet** estimates the 3D skeletal pose as joint angles and a global rotation.
  - **EDefNet (Coarse Deformation)** captures coarse skin and clothing details by predicting the deformation on the embedded graph.
  - **DisplaceNet (Fine Deformation)** refines the results with high-frequency details with a vertex displacement field.
  - **Face & Hand Prediction** replaces the corresponding template parts with parametric hand and face models and then predicts those parameters.

**Results**

- Precisely overlays onto the input images
- Captures the wrinkle patterns nicely.
- Occluded regions look plausible in the back views.

**Subjects with different types of apparel, poses, backgrounds**

---

**Comparisons**