Evaluating 2D Human-Posture-Estimation Accuracy Using 3D Motion-Capture Data

Student

Kai Erdin

Introduction: Human posture estimation (HPE) is a computer vision technique yielding the joint positions on an image of a human. This technique enables a variety of modern applications like giving feedback to users while they perform physical exercises. However, today's HPE neural models still offer a limited accuracy. Reasons for this are, a huge variability in the appearance of humans on videos, and manually and thus imprecisely labeled videos used for their training. An approach for the accuracy evaluation of HPEs is developed in this work. We introduce rigidly aligned reflective markers (Figure 2), enabling precise and automated 3D localization of joint centers. The precision and accuracy of this method make it ideally suited for HPE-model performance evaluations conducted in this work.

Approach / Technology: The aim of this work is to compare 2D joint positions estimated by a HPE neural model from an RGB images with a valid reference. Therefor, we use true 3D joint positions captured and recorded during the execution of physical exercises. Specifically, we use an OptiTrack motion capture system with 8 infrared (IR) cameras and one RGB camera together with a motion capture suit carrying reflective markers (Figures 1, 3).

To compare the 2D joint positions, estimated by the HPE from the RGB image with the captured 3D joint points, two metrics have been defined. A first metric, expressing the minimum possible Euclidean deviation in 3D space in meters, and a second metric, expressing the Euclidean deviation on the 2D image in units of pixels.

We supplied the estimated 3D marker positions on the surface of the motion capture suit to the Motive application, a tool offered by OptiTrack to estimate the joint positions of the skeleton. Tests uncovered frequent estimation errors, often larger than 10 cm, making the estimated joint positions useless as a reference. As an alternative, a rigid body holding multiple reflective markers was constructed in this work (Figure 2), so that the geometric center of all markers is well aligned with the elbow center position. Using this rigid body, the 3D position estimate of the elbow joint, projected to the RGB video, appears to be exactly at the true joint center position by visual inspection (see Figure 3).

Result: The reference elbow joint center positions, obtained from the rigidly aligned markers, were compared with the outputs of the two HPE models MediaPipe and YOLO-Pose, using the minimal spatial distance metric. The resulting deviations are as large as 8.1 and 8.6 cm RMS, respectively. These large values indicate that state-of-the-art HPE-models still have substantial potential for improvement. Specifically, physiotherapy applications, giving automated feedback on the execution of physical exercises conducted in front of a camera, could benefit from an improved HPE accuracy. A key to such improvements can be automatically labeled videos by using highly accurate joint position measurements obtained by the proposed rigidly aligned marker technique. Alternatively, accurately measured joint positions enable the animation of avatars for the generation of synthetic physical exercise videos with extremely accurate joint-center labels.

Figure 1: 3D human posture tracking using an OptiTrack motion capture system with 8 IR-cameras.
Own presentment

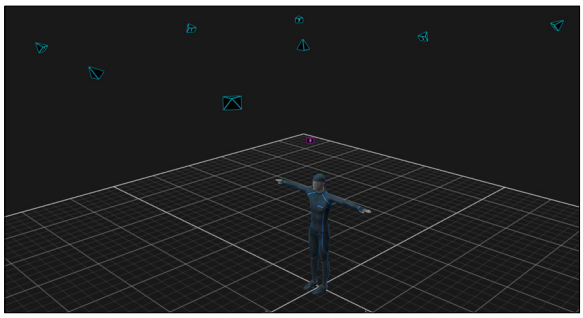


Figure 2: Different detection methods. Top-Right: Rigidly aligned markers for accurate elbow joint-center detection.
Own presentment

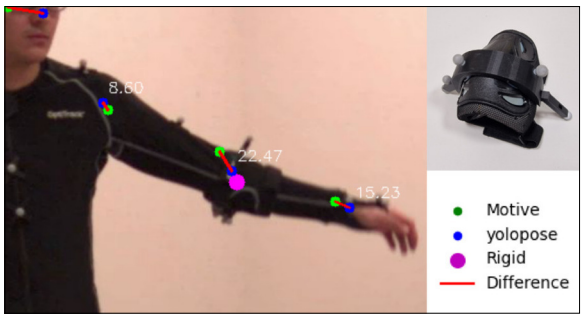
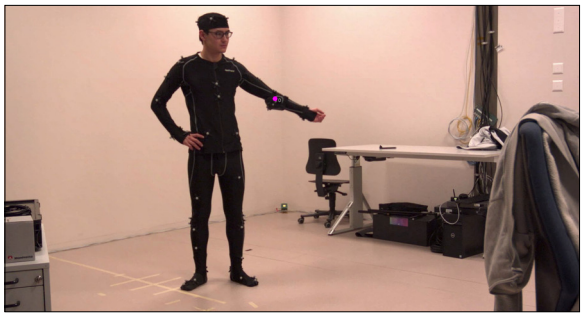


Figure 3: 3D elbow joint-center position projected onto an RGB video frame and indicated by a red bullet.
Own presentment



Advisors
Prof. Dr. Martin Weisenhorn, Marc Benz

Subject Area
Image Processing and Computer Vision

Project Partner
ICOM Institute für Communication Systems, OST - Ostschweizer Fachhochschule, Rapperswil, SG