Visual Feedback

A visual feedback system for core training using a monocular camera image. To support the user in maintaining the correct postures from target poses, by adopting 3D human shape estimation for both the target image and input camera video.

Its aim to provide a user interface for visual feedback in sports training using state-of-the-art pose estimation approaches. In contrast to 3D sensing using laser sensors or depth cameras, the proposed system only requires a single image from a standard web camera.

The framework consists of two main components: pose estimation and visual feedback.

To obtain 3D human shape and pose from a single image, the proposed system adopts pose estimation based on OpenPose to realize the bounding box of the human region. It also employs human mesh recovery to estimate the 3D human pose. For the generation of 3D human mesh, an SMPL model is used.

OpenPose can be used to detect the multiple persons' pose from an input image using multi-stage CNN. SMPL model can represent various human postures and shapes, which are defined by the triangular meshes $M(\theta, \beta) \in R^{\to 3XN}$. $N = 6, 980$ denotes the number of mesh vertices. In an SMPL model, human pose θ is defined by the skeleton rig with $K = 23$ joints and $\theta \in R^{\to 3Xk+3}$. Therefore, pose θ has $3\times K+3 = 3\times 23+3 = 72$ parameters including 3 for each part and 3 for root orientation. $\beta \in R^{\to 10}$ denotes shape parameters for 3D human mesh model in 10 dimensions of shape space using principal component analysis.

Human mesh recovery (HMR) model is a deep learning based approach for reconstructing 3D human model from a single RGB image. The human mesh model SMPL can be generated from 9 DoF camera pose (translation, rotation, and scaling) and SMPL parameters pose θ and shape β, which can be evaluated using a 3D regression module. In the discriminator network, the natural human shapes are distinguished if unnatural joint bending and unnatural body shapes such as too thin exist.

For visual feedback, a runtime user interface is proposed to guide the user on the differences between target pose and the current pose from the captured frame. Both target and current poses are estimated from a single image using the pose estimation method discussed above.

The user can predefine the easy-tosee viewpoints on the pre-training user interface. The user can first, select the target image with a standard pose to be used for core training. Afterward, the system generates a 3D human model in two view windows using the proposed pose estimation approach. For each view window, the user can select the good viewpoints in his/her judgment. Finally, the system saves the predefined set of viewpoints.

At the user interface the 3D human model that denotes the target 3D pose is represented by blue color while the 3D human model that represents the current pose from the camera image is represented by red color. The current pose model and the target pose model are displayed in a superimposed manner. Different colors of markers are used to indicate the differences between current and target models using for training guidance.

They used red, orange, yellow and green-yellow colors in their implementation to represent distances ranging from far to close.

For representing the differences between target and current 3D models, they adopt color visualization on predefined 10 marker positions on the human body. They set the marker positions at both hands and elbows, shoulders, knees, and ankles because the differences of the two models are more apparent at the end parts than the human trunk. They did not choose hip joints on the body because the differences in hip points are not apparent due to coincided waist positions.

They used shoulders, elbows, and hands positions for a good representation of the upper body postures, which are the most important parts in core training.

To obtain the marker positions using SMPL-based human model, they superimposed the aforementioned joint positions with SMPL mesh model to determine the start and end points for each marker.

To obtain the mesh vertices of markers, they substituted the transformation matrix of perspective projection from the 3D model to the output image. The relation existing in the perspective projection can be expressed as follows.

$$\begin{pmatrix} X_{image} \\ Y_{image} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{500 \cdot X_{model}}{Z_{model} + 2} + 332.5 \\ \frac{500 \cdot Y_{model}}{Z_{model} + 2} + 325 \\ 1 \end{pmatrix}$$

Finally, they calculate the difference, d_e in vertex positions with four colors. The color of marker point is set to be red color if $d_e \in [0.5, \infty)$; orange color if $d_e \in [0.25, 0.5)$; yellow color if $d_e \in [0.1, 0.25)$; and green-yellow if $d_e \in [0, 0.1)$.

To verify the efficiency and feasibility of the proposed Visual feedback system, the researchers compared the proposed system with the usual visual feedback on skeletal information.

They asked users to use both systems and answer a questionnaire, the results were that the proposed system is more effective in correcting postures when performing core training and requires less time to achieve the correct posture from the usual visual feedback on skeletal information.

The system has tested only on eight participants all of whom are male graduate students around 25 years old, so the results is limited.

In addition, the current execution time of the proposed visual feedback system is about 2 seconds for one frame. Although it may be suitable for core training because the user has to maintain the pose over a long period, it will be bottle-neck for other sports such as gymnastics and ball sports where movement is in high speed.

Another disadvantage that the standard SMPL model was used in this study and the actual human body cannot be captured in real-time.

The system is not available for download and is not provide as an open source library.