International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2021)
J. Maiero, M. Weier, and D. Zielasko (Editors)

Poster

Estimating the Pose of a Medical Manikin for Haptic Augmentation of a Virtual Patient in Mixed Reality Training

David Scherfgen[1] and Jonas Schild[1]

[1]Hochschule Hannover University of Applied Sciences and Arts, Hannover, Germany

Figure 1: *Estimating the pose of a medical manikin can be used for haptic support in medical training.*

Abstract

Virtual medical emergency training provides complex while safe interactions with virtual patients. Haptically integrating a medical manikin into virtual training has the potential to improve the interaction with a virtual patient and the training experience. We present a system that estimates the 3D pose of a medical manikin in order to haptically augment a human model in a virtual reality training environment, allowing users to physically touch a virtual patient. The system uses an existing convolutional neural network-based (CNN) body keypoint detector to locate relevant 2D keypoints of the manikin in the images of the stereo camera built into a head-mounted display. The manikin's position, orientation and joint angles are found by non-linear optimization. A preliminary analysis reports an error of 4.3 cm. The system is not yet capable of real-time processing.

CCS Concepts

• ***Human-centered computing*** → ***Mixed / augmented reality; Virtual reality;***

1. Introduction

Medical emergency training through means of virtual reality (VR) allows users to safely train relevant medical aspects like diagnostics and treatment of varying patient conditions in complex situations. In an augmented virtuality (AV) approach, users experience a virtual scenario through a head-mounted display (HMD) and can haptically feel selected objects that need to be tracked within the real world. We focus on social interaction with a virtual patient, since we expect that observing an empathetic reaction to a haptic touch may lead to an increased engagement in training.

2. Related work

Hanzaki et al. [HB20] combined a medical simulation manikin with a virtual character and compared this setting in an MR and VR environment. The MR scenario achieved higher believability scores and higher realistic experience scores. Girau et al. [GMB*19] created a virtual counterpart of a simulation manikin by 3D-scanning a real person. They integrated the physical manikin and virtual character at the same position inside the MR environment, however the manikin's pose had to be determined in a lengthy manual process.

3. Body pose estimation via optical keypoint detection

The advent of deep learning and convolutional neural networks (CNNs) has led to considerable progress in image-based human body pose estimation. The open-source library OpenPose [CHS*19] can reliably identify the pixel coordinates in an image where certain keypoints (eyes, nose, shoulders, etc.) of a human body are located. Pose detection CNNs also work on human-like medical manikins. Most HMDs feature integrated cameras that can provide the input images to the CNN. The process of extracting a 3D body pose from 2D keypoints is described in the following. Except for a HMD with a camera, it does not require any additional equipment, neither does the manikin have to be altered in any way.

DOI: 10.2312/egve.20211334

3.1. Preparation of the virtual patient model (offline)

For each body keypoint (OpenPose uses 25 keypoints), the system must know where the corresponding point is located in the virtual patient model. The model is a standard skinned mesh, i.e. its geometry is affected by the rotation of virtual bones. For each keypoint, an invisible "marker" object is inserted into the model's bone hierarchy at the appropriate location, e.g. the marker for the "nose" keypoint is placed at the tip of the nose and attached to the "head" bone. Additionally, the bones corresponding to the manikin's movable limbs must be identified. Each is manually assigned some rotational degrees of freedom (rotation axes and min./max. rotation angles), based on the physical properties of the manikin's joints.

3.2. Camera calibration (offline)

In order to make absolute measurements from a camera image, the camera's intrinsic parameters must be known (e.g. focal length, lens distortion). For 3D information, we also must know the camera's position and orientation in space at the point in time when the image was captured. Even though the HMD is tracked, we do not know the camera's position and orientation relative to the HMD, i.e. where exactly it is mounted relative to the coordinate frame used by the tracking system. We modified a standard camera calibration technique where a chessboard pattern is observed while moving the tracked HMD. By comparing the camera poses determined by the calibration algorithm to the HMD poses determined by the tracking system, the camera's HMD-relative position and orientation can be found.

3.3. Image capture (online)

We capture images of the manikin from different perspectives, storing the camera position and orientation with each image. Usually, 5-10 images allow for good results. Since several images from different views are required, the system currently does not allow for real-time tracking of the manikin. We use the system in a "static" way in our medical training project, i.e. the manikin is captured once before starting a session, or on demand after being moved.

3.4. Initial estimation of position and orientation (online)

Before being able to find the manikin's full pose, an initial estimate is required, as otherwise the subsequent non-linear optimization won't converge on a good solution. We found that estimating only the manikin's position and orientation is sufficient, i.e. we do not need joint angle estimates. We triangulate the 3D location of each body keypoint that was observed at least twice. We then find a translation and rotation for the virtual patient model (in some neutral default pose) so that the distances between its keypoint markers (see 3.1) and the manikin's 3D body keypoints are minimized.

3.5. Non-linear optimization of all parameters (online)

Finally, we apply non-linear optimization to find the values of all parameters (position, orientation and joint angles) that minimize the 2D distances between the observed keypoints in the camera images and the projected keypoint markers in the model. The min./max. joint angles act as constraints. We then apply the solution to the virtual patient model, setting its position and orientation and rotating the bones according to the angles found (see Figure 2).

Figure 2: *Camera image of the medical manikin (left) and the corresponding HMD view of the virtual patient after body pose estimation using keypoint detection (right).*

4. Preliminary results

We used OpenPose 1.7.0 and Ceres Solver 2.0.0 for non-linear optimization. With a GeForce GTX 1060, detecting keypoints took an average of 116 ms per image. Non-linear optimization time depends on the number of keypoint observations. With an Intel Core i7-7700HQ, we measured an average of 0.9 ms per keypoint observation. Using a Valve Index HMD with a built-in stereo camera ($2 \times 960 \times 960$ px), we estimated the solution accuracy by measuring distances between keypoints on the manikin and on the virtual patient. Median/mean distances were 4.3/4.8 cm. Accuracy must be improved in order to allow for a convincing haptic experience that supports precise interaction. We use the system for social interaction and are planning to examine effects on social presence due to the haptic presence of the manikin, where it matters if practitioners lay a hand on the patient's shoulder to provide comfort.

5. Conclusion

We presented our manikin pose estimation method based on detecting body keypoints of the manikin in HMD camera images and finding suitable pose parameters using non-linear optimization. At this time we do not recommend using the system for precise medical interactions, but rather for supporting social interaction in VR. Currently, the processing time does not allow for real-time tracking. We are confident that the approach based on keypoint detection is advantageous regarding supporting a versatile use with different models of medical manikins, virtual patients and environments.

References

[CHS*19] CAO Z., HIDALGO MARTINEZ G., SIMON T., WEI S., SHEIKH Y. A.: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence* (2019). doi:10.1109/TPAMI.2019.2929257. 1

[GMB*19] GIRAU E., MURA F., BAZURRO S., CASADIO M., CHIRICO M., SOLARI F., CHESSA M.: A Mixed Reality system for the simulation of emergency and first-aid scenarios. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (2019), 5690–5695. doi:10.1109/EMBC.2019.8856777. 1

[HB20] HANZAKI M. R., BOULANGER P.: Proxy Haptics for Surgical Training. In *Proceedings - 2020 22nd Symposium on Virtual and Augmented Reality, SVR 2020* (2020), pp. 134–143. doi:10.1109/SVR51698.2020.00033. 1