2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)

TutAR: Semi-Automatic Generation of Augmented Reality Tutorials for Medical Education

Daniel Eckhoff*
Nara Institute of Science and Technology, Japan

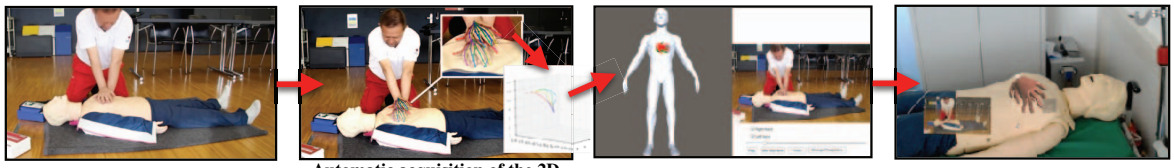
Christian Sandor†
Nara Institute of Science and Technology, Japan

Denis Kalkofen‡
Graz University of Technology, Austria

Ulrich Eck§
Technical University of Munich, Germany

Christian Lins¶
OFFIS - Institute for Information Technology, Oldenburg, Germany

Andreas Hein‖
Carl von Ossietzky University of Oldenburg, Germany



a) Input video. b) Automatic acquisition of the 2D trajectory and 3D posture of up to two hands. c) Manual authoring to reconstruct a 3D trajectory. d) Augmented Reality Tutorial.

Figure 1: The author chooses the input video (a). TutAR extracts relevant motion from the video (b), estimates the hand posture and creates an animation. With an authoring tool, the author can reconstruct a 3D trajectory relative to the human body (c). The animation of the 3D hand will be displayed on the registered place and plays synchronously with the motion in the video on an OST-HMD (d).

Abstract

With Augmented Reality (AR) on Optical-See-Through-Head-Mounted Displays (OST-HMD), users can observe the real world and computer graphics at the same time. In this work, we present TutAR, a pipeline that semi-automatically creates AR tutorials out of 2D RGB videos. TutAR extracts relevant 3D hand motion from the input video. The derived motion will be displayed as an animated 3D hand relative to the human body and plays synchronously with the motion in the video on an OST-HMD.

Index Terms: Augmented Reality—Video tutorials—Motion extraction

1 Demo Description

What makes it unique and special **TutAR** is capable of automatically transferring 2D videos into 3D AR applications with minimal user input. TutAR extracts 3D objects and their movement from 2D videos and images. The system displays the 3D movement in the real world through an OST-HMD. The user can change the viewpoint without degrading the quality of the AR experience.

Figure 2 shows a pipeline of the proposed solution. We are aiming to automate as many steps as possible. Due to some limitations, we still rely on input from an author of the AR tutorial. In the figure, automated steps are colored blue, whereas steps which require author input are colored red. During authoring, the author divides the video tutorial into segments for each task. In each segment, TutAR acquires motion of up to two human hands. For acquiring hand motion, we are using OpenPose [1]. OpenPose is a library for body keypoint detection allowing acquiring the position for 21 keypoints for each hand in image space. To retrieve the hand posture of each hand, we use Hand3D [6].

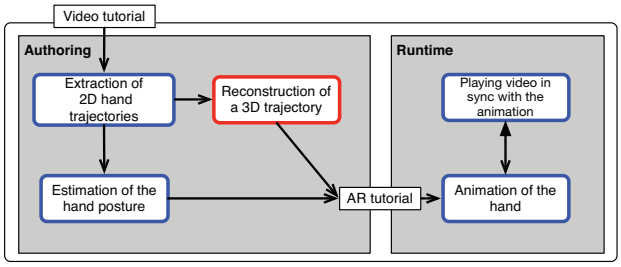


Figure 2: The pipeline of TutAR. Automated steps are colored blue, whereas steps which require author input are colored red.

However, just extracting 2D hand motion in a video in image space is not enough to create a useful 3D AR application. Also, it is difficult to estimate the pose of an object in 3D space in a 2D image with an unknown environment. Thus, an authoring tool is needed. Since we are focussing on tutorials for medical education which are using hands as the main tool on the human body, the authoring tool involves around creating a 3D trajectory relative to the human body.

An authoring tool in which the author is asked to place the hand relative to the manikin is shown in Figure 1 (c). After adjusting the hands of interest, the author can add the pose of the hands of the current frame as a keyframe. The author can add new keyframes when the position of the hand should be corrected. Between each keyframe, the pose will be interpolated. TutAR will create an AR application with the reconstructed 3D motion for the Microsoft HoloLens, a popular OST-HMD. The system renders the 3D hands relative to a user-placed spatial anchor. The motion of the hands will be played in sync with the input video which will also be displayed on the HMD.

*e-mail: daniel.eckhoff@gmail.com
†e-mail: sandor@is.naist.jp
‡e-mail: kalkofen@icg.tugraz.at
§e-mail: ulrich.eck@tum.de
¶e-mail: christian.lins@offis.de
‖e-mail: andreas.hein@offis.de

DOI 10.1109/ISMAR-Adjunct.2018.00131

Explanation of the novelty TutAR is creates medical AR tutorials for education semi-automatically out of RGB videos.

The system profoundly differentiates from existing work. It can reconstruct a 2D hand trajectory including the 3D hand posture. With additional author input, a 3D trajectory can be reconstructed and transformed into a 3D trajectory.

Automatic approaches to generate AR applications go back to the pioneering work of Feiner et al. [2] from 1993. They describe a prototype of a scripting-based AR system for maintaining laser printer based on a see-through HMD [5].

A popular approach consists of overlaying video material on the user's view. Some works, including [4], merely overlay a 2D RGB video of a reference performance directly onto the user's view on the AR display. Overlaying the original video has the advantage of showing unaltered content. This approach also does not require a lot of authoring. However, by just overlaying a 2D video, the user will not be able to have a free choice of the viewpoint limiting the practical value. The approach of his work is mostly inspired by the work of Mohr et al. [3]. However, while their system creates AR tutorials with surface contact from 2D RGB videos, TutAR creates tutorials out of acquired hand movement from tutorial videos.

Why will it draw a crowd? Our demo is very enjoyable. It will show a demonstration of a semi-automatic generated cardiopulmonary resuscitation (CPR) tutorial. ISMAR attendants will be able to perform a TutAR assisted CPR on a manikin.

REFERENCES

[1] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Computer Vision and Pattern Recognition*, pages 7291–7299, 2017.

[2] S. Feiner, B. Macintyre, and D. Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, pages 53–62, 1993.

[3] P. Mohr, D. Mandl, M. Tatzgern, E. E. Veas, D. Schmalstieg, and D. Kalkofen. Retargeting video tutorials showing tools with surface contact to augmented reality. *ACM Conference on Human Factors in Computing Systems*, pages 6547–6558, 2017.

[4] N. Petersen and D. Stricker. Learning task structure from video examples for workflow tracking and authoring. *International Symposium on Mixed and Augmented Reality*, pages 237–246, 2012.

[5] J. Zauner, M. Haller, A. Brandl, and W. Hartmann. Authoring of a mixed reality assembly instructor for hierarchical structures. In *International Symposium on Mixed and Augmented Reality*, pages 237–246, 2003.

[6] C. Zimmermann and T. Brox. Learning to estimate 3d hand pose from single rgb images. In *Computer Vision and Pattern Recognition*, pages 4913–4921, 2017.