



Anatomical Augmented Reality with 3D Commodity Tracking and Image-space Alignment - Supplemental Material

Armelle Bauer^{a,b}, Debanga Raj Neog^c, Ali-Hamadi Dicko^{b,d}, Dinesh K. Pai^e, François Faure^{b,d}, Olivier Palombi^{b,c}, Jocelyne Troccaz^a

^aUniv. Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, F-38000 Grenoble, France

^bUniv. Grenoble Alpes, CNRS, Grenoble INP, INRIA, LJK, F-38000 Grenoble, France

^cGrenoble Alpes Hospital, LADAF, F-38000 Grenoble, France

^dAnatoScope SA, F-34000 Montpellier, France

^eDepartment of Computer Science, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

ARTICLE INFO

Article history:

Received September 30, 2017

Keywords: User-specific Anatomy, Augmented Human, Real-Time, Motion Capture, Augmented Reality, Markerless Device, Handled occlusion and self-occlusion

ABSTRACT

This document is Supplemental Material for the paper titled *Anatomical Augmented Reality with 3D Commodity Tracking and Image-space Alignment*. We decided to create a hierarchical body tracking system with anatomically constrained joint orientations using the *Kinect* body tracking skeleton as input and allowing us to correct significantly *Kinect* rough data. We conducted experiments to have a better understanding of *Kinect* body tracking limitations as a way to evaluate more precisely our system results.

© 2017 Elsevier B. V. All rights reserved.

To simplify this document, we will annotate as *Kinect* the Microsoft V2.0 *Kinect*. *Kinect* is a RGB-D camera composed of a HD RGB camera (30Hz, 1920x1080 pixels), an infrared transmitter (light non-dependent), a depth captor (30Hz, 512x424 pixels), and four microphones.

The *Kinect* sensor gives in real-time, at each frame, and for each pixel of an image, a RGB color and depth value. The *Kinect* SDK provides tools to go from image space (2D) to camera space (3D). It is also able to track up to 6 users at the same time. For each user, the *Kinect* SDK gives an animation skeleton composed of 25 joints (position and orientation): the *Kinect* body tracking skeleton.

Appendix A. Optimal *Kinect* conditions

Thanks to the infrared sensor, using the *Kinect* does not require any specific light condition for optimal user tracking results. By avoiding light sources pointing in the direction of the *Kinect* color sensor, and by favoring natural light, the global

quality of the color map given by *Kinect* is improved and usable.

According to Pfister et al. (2014), to obtain more precise results concerning the user's lower limb joints, the *Kinect* needs to be 60cm up ground. This height will allow a better tracking of the motions done while crouching.

We observed that if the user's body is reflected, for example on the ground or in a mirror, the *Kinect* usually consider the reflection as part of the user's silhouette. This problem can induce errors, for instance lower limb lengths errors in the case of the user's reflection in floor tiles.

The infrared rays provide depth informations but do not allow to visualize objects that are behind others and thus imply occlusion problems. User occlusion may drastically reduce the overall quality of the *Kinect* body tracking system results. These occlusions can be of two different types: part of the user is behind an obstacle (fixed or mobile), or, part of his body is occluded by himself (e.g self-occlusion) from the *Kinect* point of view.

Appendix B. Kinect body tracking skeleton

As we mentioned in the abstract, the *Kinect* is able to track 6 users, and for all of them it gives a body tracking skeleton composed of 25 joints. For simplicity, in our work we consider these 25 joints as being the principal human body joints. Despite the fact that the limited number of joints will limit motion range, we decided to keep this simple structure to represent the user's body.

To avoid occlusion problems, the *Kinect* SDK estimates joint positions, along with a state variable representing the reliability. A joint can be defined as tracked, inferred (by looking at adjacent joints) or not tracked (if there is not enough information to infer it).

Note that the users silhouettes may be wrong due to reflective surfaces, background objects, large clothing and even some hair styles.

Appendix C. Body orientation from the Kinect point of view

To better understand the limitations in term of body orientation, we did two experiments. The first experiment (see Fig. C.1.a), we asked a user in T-pose and facing the *Kinect* to turn by 15 degrees every two seconds until coming back facing the *Kinect* sensor. The second experiment (see Fig. C.1.b) is the same with U-pose; we also asked the user to perform a squat motion at each step of 15 degrees.

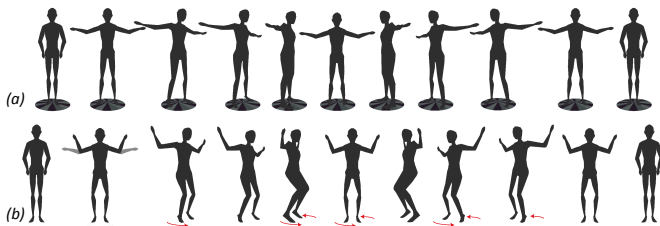


Fig. C.1. Schema of two *Kinect* sequences done to define the critical limits in terms of body orientation from the *Kinect* point of view.

Using these two sequences and the state variable of the joints, we determined the critical body orientation from the *Kinect* point of view. We admit the results consistent when more than half the user's joints are considered tracked by the *Kinect*. This is based on the *Kinect* state variable evaluation for each joint and at each step of the sequence. We found the results acceptable as long as the user orientation remains less than 60 degrees away from facing the *Kinect*. The range of trackable motion decreases along with the angle, due to self-occlusion.

Note that a user seen from the back is considered as facing the *Kinect* by the sensor. In this extreme case, all the orientations will be considered tracked by the sensor but the values will not be exploitable (e.g knee bending the wrong way).

Appendix D. Quality of the Kinect body tracking skeleton

To evaluate the robustness of the skeleton modeling, we observe the variation of the length in upper limbs segments.

We designed a sequence during which the user bends the left arm (rotation of the elbow joint of 160 degrees) step by step, by staying still for two seconds each 35 degrees. This sequence is executed with three variations: with a closed fist (see Fig. D.2.a), with a flat hand oriented in the frontal plane (see Fig. D.2.b), and with a flat hand oriented in the transverse plane (see Fig. D.2.c).

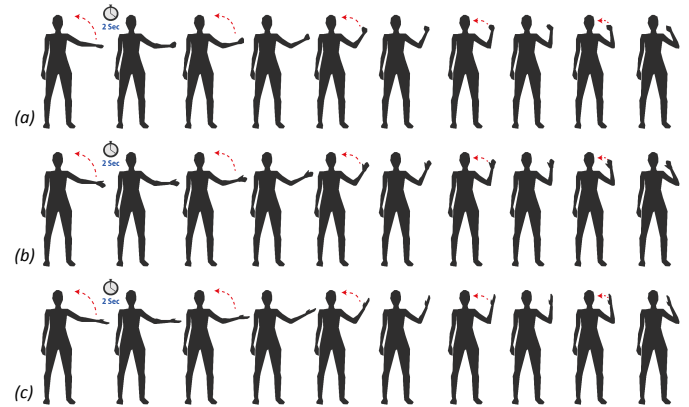


Fig. D.2. Schema of three *Kinect* sequences done to observe the variation of upper limb segments length.

Table D.1 displays the lengths of two segments for each sequence. The blue line (titled "reference value") represents the value measured directly onto the user's bent arm and is considered as ground truth.

Shoulder-Elbow Segment Length

	Fig. D.2.a	Fig. D.2.b	Fig. D.2.c
Reference value	21.5		
Mean value	20.44	22.02	19.42
Max variation	6.12	4.73	5.06
Standard deviation	0.5	1.2	1.1

Elbow-Wrist Segment Length

	Fig. D.2.a	Fig. D.2.b	Fig. D.2.c
Reference value	22.3		
Mean value	21.18	21.37	19.89
Max variation	7.23	7.78	7.79
Standard deviation	1.6	1.2	2.0

Table D.1. Measurements (in cm) obtained from the sequences presented in Fig. D.2.

We infer from these results that the distances are mostly fluctuating depending on the hand. For instance, the segments length obtained in the third sequence (see Fig. D.2.c) are 2 to 2.5cm smaller than the real value. This can be due to the fact that a flat hand represents less pixels and this could induce errors in the user captured silhouette. On the one hand pose, we can see that the standard deviation is quite small (between

0.5cm and 2.0cm), while on the other the fluctuations can be really high (up to 7.79cm). This confirms that the use of rough *Kinect* body tracking data without distance constraints between joints can generate visual artifacts near the joints of the 3D anatomical model (e.g bones separation or inter-penetration).

In the final experiment, we observed the trajectories of hand joints in space: wrist, hand palm, thumb and fingers tip. To do so, we gave the user a rigid bar to hold allowing him to keep a constant distance of 50 cm in-between his hands. The user performed a set of motions. Some of them presented in *Fig. D.3*: swinging arms from bottom to top, rowing motion, swinging arms from left to right, and so on.

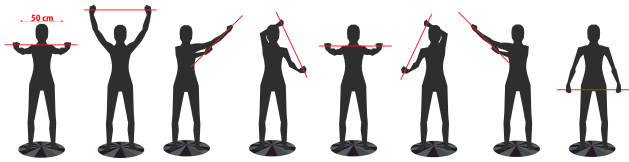


Fig. D.3. Schema of a *Kinect* sequence to evaluate repeatability and precision of joints positions (e.g wrist joint, hand palm joint, finger tip joint and thumb joint).

Table *D.2* displays the average value, the maximum variation and the standard deviation for each distance of interest (e.g distance between wrists, distance between hand palms, distance between fingers tip and distance between thumbs).

Hand to hand distance measured at:	Mean	Max variation	Std deviation
Wrists	55.78	25.03	3.1
Hand palms	57.08	29.41	2.0
Finger tips	56.91	27.2	4.5
Thumbs	56.91	36.37	2.1

Table D.2. Measurements (in cm) obtained from the sequence presented in *Fig. D.3*.

We observe that most often the average value is 6 to 7cm larger than expected. Considering that the user's fist is 8.5cm wide, we can explain these errors. We can also estimate, for a motion sequence, that the standard deviation is acceptable. Moreover, we can see that variations are really important in these joints, which implies a lot of noise in the tracking data.

References

Pfister, A, West, AM, Bronner, S, Noah, JA. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *Journal of medical engineering & technology* 2014;38(5):274–280.