Accurate, 360-degree measurement of multi-person posture using Kinect cameras & special dress design & analysis for accurate 3D tracking using a single camera

Nilaksh Agarwal, Jay Paranjape, Manas, Prof. Rahul Garg, Prof. Prathosh A. P., Prof. Varsha Singh

COMPUTER SCIENCE AND ENGINEERING IIT DELHI

ABSTRACT

In this study, we aim to provide accurate measurement of 3D keypoints (joints) from a Kinect Camera input (RGB + Depth), to obtain an accurate multi-person posture, including missing/occluded parts. We wish to do a real-time 360 degree tracking of the same.

We also aim to design a special dress and develop an analysis algorithm for tracking using a single camera + dress

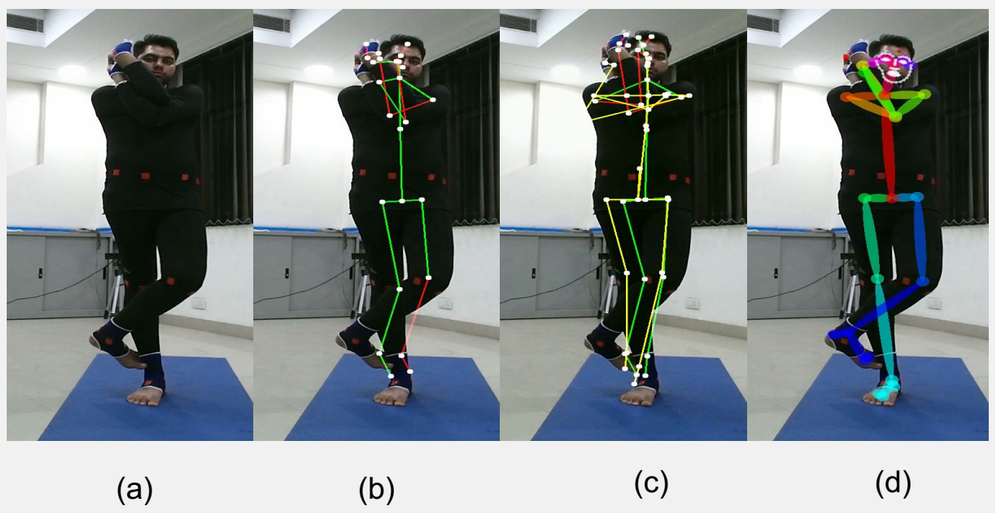
POSTURE TRACKING

Target is to make a deep-learning pipeline to localize 2D and 3D keypoints (joints) from a given RGB+Depth image.

Failure of Kinect : Tracking the Yoga asanas can be challenging due to unusual body extensions and high amount of self-occlusion in the frames. Kinect SDK fails in such cases [Figure 1(b) & Figure 2(b)]

Earlier Work: We worked on a model which integrated RGB and Depth images, and obtained better results than the Kinect [Figure 1(c) & Figure 2(c)]

2D Prediction Pipeline : Replicating works of *Cao et al.* and *Martinez et al.* we used a pipeline to predict 19 body keypoints, 21 hand keypoints, 3 foot keypoints and 70 facial keypoints. [Figure 1(d) and 2(d)]



(a) (b) (c) (d)

Figure 2 - Garudasana: (a) Original image, (b) Kinect Predictions (c) Earlier Predictions (d) Our Model Predictions

Future Work: Next steps include mapping 2D joints and depth maps to 3D joints. Using the limb-length, joint angle constraints and previous frames of video priors into the predictor's learning, and training on the generated animation data with ground truth labels

Potential Applications: Posture detection has applications in Video Surveillance, Action recognition, Animation, Gaming, Sports coaching, Filming and CGI etc.

MOCAP SUIT

We aim to make a novel motion capture suit which can track **any** body point in real time.

Algorithm : Suit comprises of red and blue patterns with different amounts of intensity. Thus, a 3X3 matrix as captured by a simple high resolution camera will be sufficient to uniquely label a point.

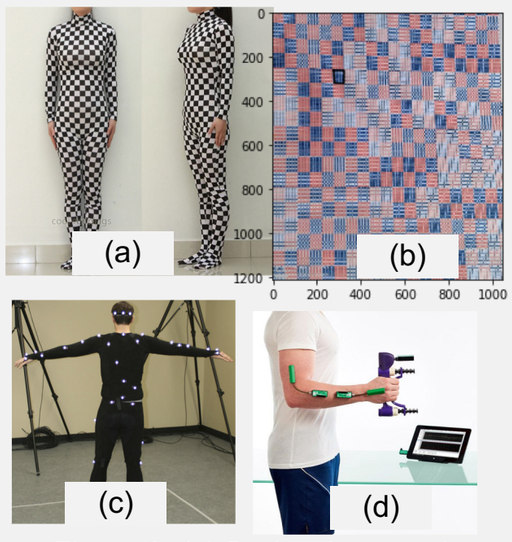


Figure 3: (a) Bodysuit example, (b) Pattern used, (c) Modern Mocal Suit, (d) Goniometer

Potential Applications : This is a real time high accuracy tracking which will improve the current motion capture process. This has high applications in sports, animation industry, filming, posture studies, recommender and selection systems. We plan to complete the algorithm, prepare the dress and start experiments. If successful, we will be applying for a patent

Cost of present day tracking is at least 1.5 lakh, with high precision goniometers and sensors . Our model aims to bring this down drastically along while increasing number of tracked points

OUTCOMES & TIMELINES

Year 1	• Data collection on 25 Subjects (2 cameras) and 50 Subjects(4 cameras) across 18 asanas (videos + depth images) • Work on localization of 2D and 3D keypoints from RGB + Depth Data • Created animation from joint predictions to clean inconsistent data
Year 2	• Single Kinect Camera single person tracking (both 2D and 3D) • Design of the dress and tracking algorithm with dress + Filing a patent • Project titled "Machine Learning model for early diagnosis of stroke in resource limited setting" jointly with AIIMS submitted to DBT • Paper titled "Mind and Body in Balance: Assessing Yoga to Demystify its Effects on Cognitive Performance" accepted at ACCS 2019 • Enabled a cognitive study for a PhD student, NRCVEE on "*Experience and Stability in Yogaasana"* and *"Yoga Experience and Anxiety"*
Year 3	• Accurate multi-person tracking using multiple Kinect cameras • Production of tracking dress & using the dress for tracking in-the-wild • Data collection for RGB+Depth images in-the-wild
Year 4	• Using Video data (additional temporal information) to improve accuracy • Finally using only RGB Video (inferring depth) to predict pose

REFERENCES

- J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, Real-time human pose recognition in parts from single depth images, 2011
- Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M. J., Laptev, I., & Schmid, C., Learning from synthetic humans, 2017
- Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., & Black, M. J., Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image, 2016
- Martinez, J., Hossain, R., Romero, J., & Little, J. J., A simple yet effective baseline for 3d human pose estimation, 2017
- Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2018). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields., 2018
- Simon, T., Joo, H., Matthews, I., & Sheikh, Y. Hand keypoint detection in single images using multiview bootstrapping. 2017.
- http://www.biometricsltd.com
- https://optitrack.com

For further queries, Contact: Nilaksh.Agarwal.ph115@physics.iitd.ac.in