Random Tree Walk toward Instantaneous 3D Human Pose Estimation

Ho Yub Jung[1], Soochahn Lee[2], Yong Seok Heo[3], Il Dong Yun[1]
[1]Div. of Comp. & Elect. Sys. Eng., Hankuk University of Foreign Studies. [2]Dept. of Elect. Eng., Soonchunghyang University. [3]Dept. of Elect. & Comp. Eng., Ajou University.

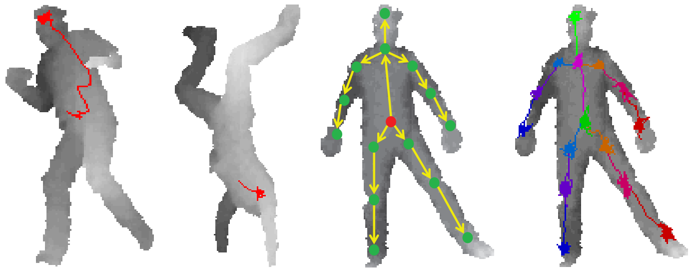

Figure 1: The red lines represents the random tree walks trained to find the head position. The random walk starts from the body center in (a). In (b), the head position is found with fewer steps by starting from the chest, which is much closer than the body center. (c) illustrates the kinematic tree implemented along with RTW. The adjacent joint positions can be used as the starting positions for new RTW. (d) shows the RTW path examples.

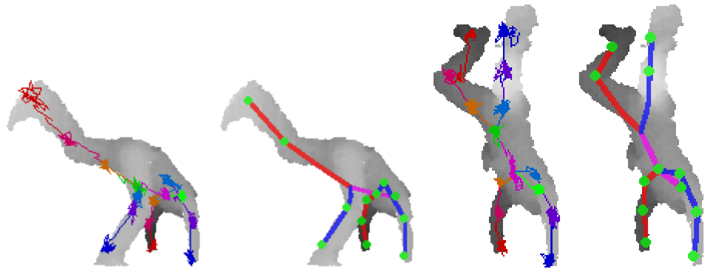

Figure 2: Example results of the RTW from EVAL db [1]. Proposed approach achieves the state-of-the-art accuracy without using the temporal prior. 64 RTW steps are taken for each joint to estimate human pose from a single depth image. The RTW paths are drawn, and the expectations of RTW steps are used to find the joint positions. The pose estimation from a single frame takes less than 1 millisecond.

The availability of accurate depth cameras have made real-time human pose estimation possible; however, there are still demands for faster algorithms on low power processors. This paper introduces 1000 frames per second pose estimation method on a single core 3.20 GHz CPU with no additional use of GPU or SIMD operations. Considering the omission of parallel computing, the proposed method is over 100 times more efficient than previous methods. A large computation gain is achieved by random walk sub-sampling. Instead of training trees for pixel-wise classification, a regression tree is trained to estimate the probability distribution to the direction toward the particular joint, relative to the current position. At test time, the direction for the random walk is randomly chosen from a set of representative directions. The new position is found by a constant step toward the direction, and the distribution for next direction is found at the new position. The continual random walk through 3D space will eventually produce an expectation of step positions, which we estimate as the joint position. A regression tree is built separately for each joint. The number of random walk steps can be assigned for each joint so that the computation time is consistent regardless of the size of body segmentation. The experiments show that even with large computation gain, the accuracy is higher or comparable to the state-of-the-art pose estimation methods.

As with the methods of [2, 3, 4], the proposed method is also based on randomized trees, and human pose estimation from single depth image. In previous methods, the joint position is estimated by aggregating pixel-wise tree evaluations. Since the body part sizes are all different, an excessive number of tree evaluations are often made for a larger body part. Similar to supervised descent approach [5], we learn to estimate the relative direction to the joint. Then, at test stage, an initial starting point is moved towards the joint position by random walk in the direction estimated from the trained randomized regression trees. We term this process as *random tree walks* (RTW). We note that, the specific vector that guides the walk is randomly selected among a set of representative vectors in each leaf node of the random tree. By reconstructing the joint position estimation using random walks, we minimize the number of required samples. Fig. 1 shows an example of the proposed RTW process to estimate the head position. We can see the path of the walk as the regression tree guides the step direction at each point.

The 15 joints in skeletal frame are as follows: head, chest, belly, L/R hip, L/R shoulders, L/R elbows, L/R wrists, L/R knees and L/R ankles. The placement of the each joint is illustrated in Fig. 1 (c). The joint positions are determined sequentially according to the typical skeletal topology. The process is illustrated in Fig. 1 (c) and (d). First, the RTW for belly position starts from the body center. Then, the expectation of RTW $\bar{p}$ is used as the q_0 of next adjacent joint. An example of the sequential RTW is shown in Fig. 1 (d).

In this paper, RTW approach for 3D pose estimation problem is proposed, which gives a large computation gain with a state-of-the-art accuracy on EVAL db [1]. The difficult poses like hand-stand are accurately estimated without the initialization from previous frames and temporal prior. See Fig. 2. The proposed approach moves away from pixel-wise classification, and applies a supervised gradient descent and MCMC like random sampler in the form of Markov random walks. By initializing the starting point to the adjacent joint according to kinematic tree, we demonstrate a robust and super-real-time pose estimation algorithm.

[1] Varun Ganapathi, Christian Plagemann, Daphne Koller, and Sebastian Thrun. Real-time human pose tracking from range data. *Proc. European Conf. Computer Vision*, 2012.

[2] Ross Girshick, Jamie Shotton, Pushmeet Kohli, Antonio Criminisi, and Andrew Fitzgibbon. Efficient regression of general-activity human poses from depth images. *Proc. Int'l Conf. Computer Vision*, 2011.

[3] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. *Proc. Conf. Computer Vision and Pattern Recognition*, 2011.

[4] Min Sun, Pushmeet Kohli, and Jamie Shotton. Conditional regression forests for human pose estimation. *Proc. Conf. Computer Vision and Pattern Recognition*, 2012.

[5] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 532–539. IEEE, 2013.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.