Meysam Madadi
Advisors: Sergio Escalera, Jordi Gonzàlez, F. Xavier Roca, and Felipe Lumbreras

Depth-based Multi-part Body Segmentation

Introduction

Soft biometrics are traits of the human body which can be used to describe a person like height, weight, and skin. They have been used in video surveillance to track people [2], in combination to hard biometrics to increase reliability and accuracy [3], person re-identification [5], and supported diagnosis in clinical setups [6], just to mention a few.

In this paper, we propose a body segmentation approach in 3D space applying Kinect to compute accurate geometrical soft biometrics such as arm and leg lengths, and neck, chest, stomach, waist and hip sizes. Pixel labels are extracted in a model based multi-part approach. To compute and compare the results accurately, a user must stand face-to-Kinect such that the whole body can be seen.

Methodology

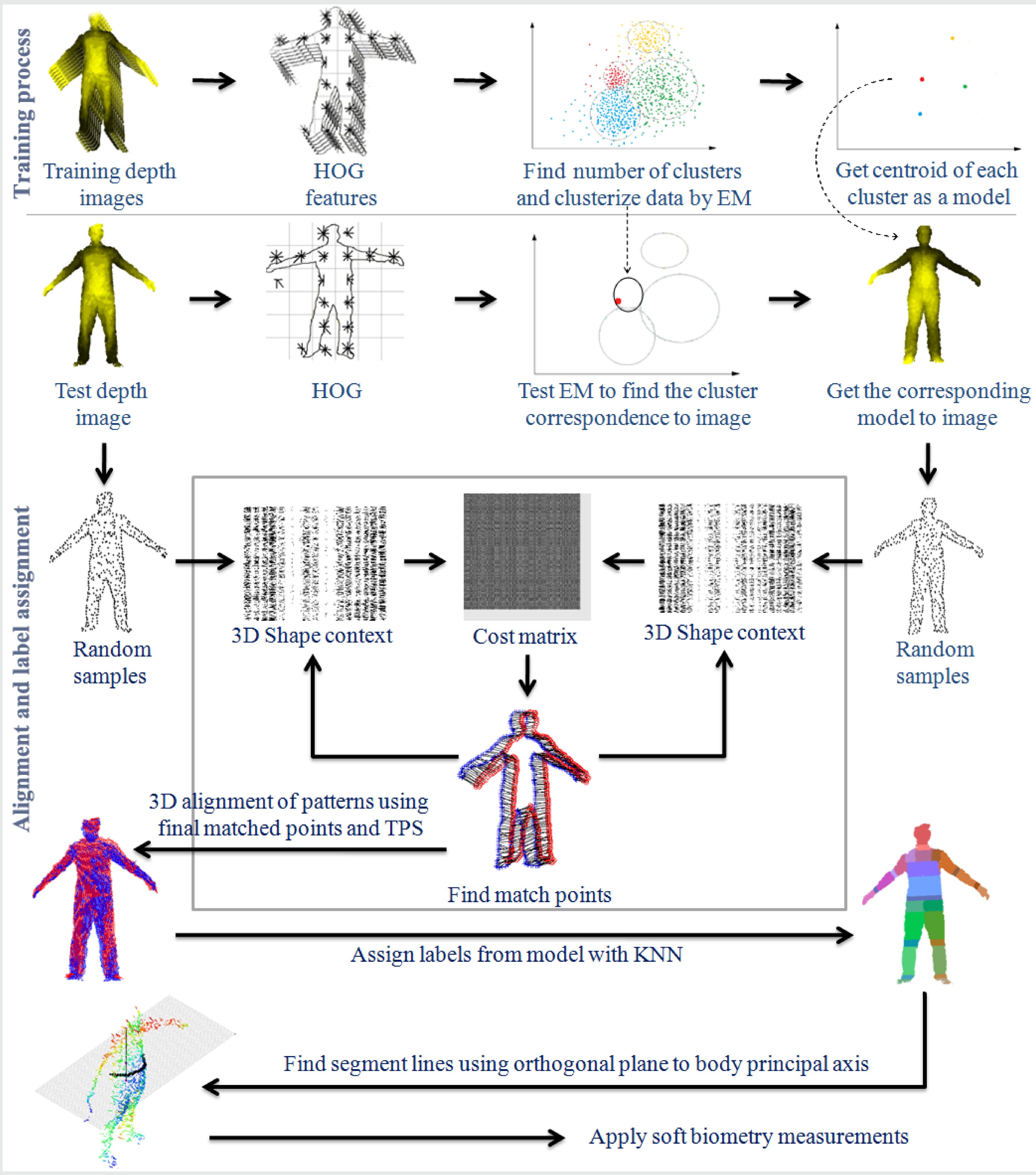
To extract body pixel labels in the point cloud, we use an iterative approach which aligns the test point cloud to the nearest model in the training set.

In the training step, we clusterize all poses using EM algorithm applying HOG descriptor [7] on depth images and keep the centroids as representative models. We optimize the number of mixtures using a combination of EM and k-means [4]. Finally, we keep EM parameters of clustering for future pose estimation as a classifier.

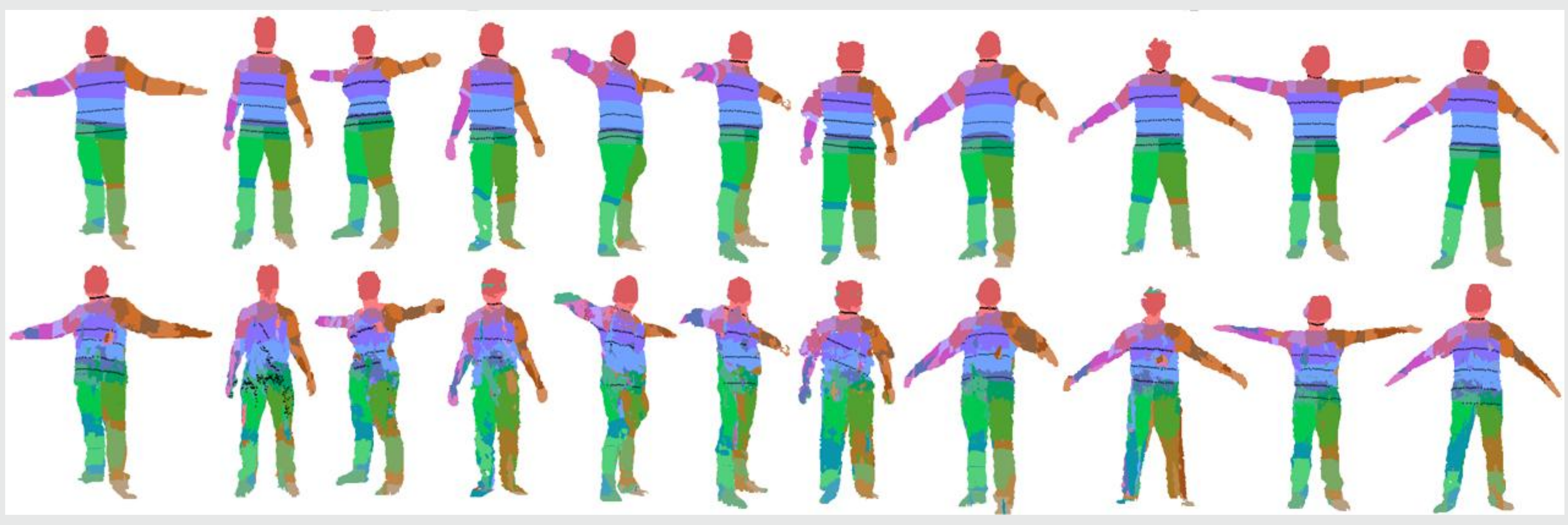
In the test step, we find the nearest model to the test image using HOG descriptor and the trained EM. In the next step, a random pixel selection is applied on test and model depth image and an iterative 3D alignment is performed applying the following process [1]:

a. 3D shape context descriptors are computed for selected points,
b. The cost matrix between all point descriptors is computed after adding some dummy points,
c. The best match points are extracted using linear assignment problem using Jonker-Volgenant algorithm,
d. The best match points are aligned using 3D thin plate spline algorithm.

This process is repeated to refine the matching points. Pixel labels are assigned from the nearest model pixels after final 3D alignment.



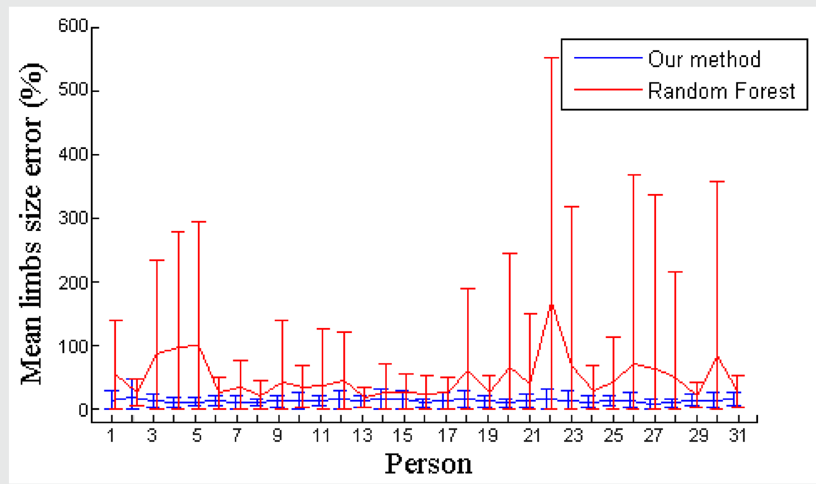
Process diagram of the system.



Qualitative results. First row our method, and second row RF labeling approach. Black points correspond to segment lines. It can be seen that segment lines accuracy has a direct relation with the segmentation accuracy and purity.

Soft biometry measurement

Having an accurate body segmentation, we are able to compute limb sizes using the hitting points of the orthogonal plane to the principal axes of the limb crossing from the mean points and body hull. The interpolation of such points makes the curve which can be used to measure the size. The length of the limbs like arm can be directly computed from the summation of distances between joint points.



Overall size error per person.

Discussion

We created a dataset containing 1061 frames of 31 individuals using Kinect to evaluate our method. We used a 10-fold cross validation over all frames to generate the results. We compare our results to the standard random forest pixel labeling approach [8]. We could get pure segmentation in edges showing an accurate segmentation also applicable in pose recovery. Our approach shows a low sensitivity to the number of training data vs. random forest.

Future work and challenges

The most critical part in this approach is point cloud registration. A more accurate registration will cause a better alignment and consequently better segmentation. In the future work, we apply a pose retrieval system to find nearest model besides using a global vs. local descriptor for registration. We apply our method on more complicated poses.

Meysam Madadi received his Bachelor degree in Software Engineering at BuAli Sina university of Hamedan and M.S. degree in Computer Vision and Artificial intelligence at Universitat Autònoma de Barcelona (UAB) in 2007 and 2013, respectively. He has started his research activities by focusing on information retrieval and data mining since his bachelor project, continuing in master specifically on computer vision and image processing. He gave a special attention to pose recovery and human behavior analysis from his master thesis. He is interested in generating and developing new algorithms in these topics applying the knowledge in computer vision and retrieval systems besides machine learning, algorithms design in artificial intelligence, statistics, linear algebra, different geometries and many other relevant areas.

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. PAMI, 24:509–522, 2002.
[2] S. Denman, A. Bialkowski, C. Fookes, and S. Sridharan. Identifying customer behaviour and dwell timeusing soft biometrics. Springer, 409:199–238, 2012.
[3] G. Guo, G. Mu, and K. Ricanek. Cross-age face recognition on a very large database: The performanceversus age intervals and improvement using soft biometric traits. ICPR, pages 3392–3395, 2010.
[4] Y. Lee, K. Y. Lee, and J. Lee. The estimating optimal number of gaussian mixtures based on incrementalk-means for speaker identification, 2006.
[5] A. Møgelmose, A. Clapés, C. Bahnsen, T. Moeslund, and S. Escalera. Tri-modal Person Re-identificationwith RGB, Depth and Thermal Features. CVPR, 2013.
[6] M. Reyes, A. Clapés, J. Ramírez, J. R. Revilla, and S. Escalera. Automatic digital biometry analysis basedon depth maps. Computers in Industry, 2013.
[7] G. Shakhnarovich, P. Viola, and T. Darrell. Fast pose estimation with parameter-sensitive hashing. ICCV,2:750–757, 2007.
[8] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Realtimehuman pose recognition in parts from single depth images. CVPR, 2011.