Unsupervised 3D Human Pose Recognition from a Single Depth Human Silhouette Using a Geodesic Map and Kinematic Body Model

Hyun Jae Jeon
Dept. of Biomedical Engineering
Kyung Hee University
Yongin, S. Korea
hjjeon1106@khu.ac.kr

Sang Beom Nam
Dept. of Biomedical Engineering
Kyung Hee University
Yongin, S. Korea
sbnam@khu.ac.kr

Sung Un Park
Dept. of Biomedical Engineering
Kyung Hee University
Yongin, S. Korea
supark@khu.ac.kr

Jun Hyuk Park
Dept. of Biomedical Engineering
Kyung Hee University
Yongin, S. Korea
psmt2655@khu.ac.kr

ChangKyoo Yoo
Dept. of Environmental Science & Engineering
Kyung Hee University
Yongin, S. Korea
ckyoo@khu.ac.kr

Jeong Tai Kim
Dept. of Architectural Engineering
Kyung Hee University
Yongin, S. Korea
jtkim@khu.ac.kr

Tae-Seong Kim
Dept. of Biomedical Engineering
Kyung Hee University
Yongin, S. Korea
tskim@khu.ac.kr

ABSTRACT

Recently, human pose recognition (HPR) in 3D using only a single depth sensor without any optical markers has become an active research topic. Till now, most existing HPR approaches are based on supervised recognition of human body parts, requiring a classifier trained with a proper database. In this paper, we propose a novel unsupervised 3D HPR utilizing a geodesic distance map (GDM) of human depth silhouette and a 3D kinematic body model which requires no training and database. From each GDM, we derive a set of landmarks of human body joints and fit the joint landmarks of a kinematic body model to them to reconstruct its corresponding pose in 3D. Our numerical evaluation results of our proposed methodology indicate a range of errors from 0.01 to 33.45 mm in the Euclidean distance of the joints to their true location in 3D. Experimental results with real data demonstrate that the proposed technique could perform HPR in 3D with a reasonable accuracy and reliability.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems – Human Information Processing.

General Terms

Human Factor, Experimentation, Performance

IMCOM'15, January 8–10, 2015, Bali, Indonesia.

http://dx.doi.org/10.1145/2701126.2701172

Keywords

Geodesic Distance Map (GDM), Landmark Localization, Kinematic Model Fitting

1. INTRODUCTION

Human pose recognition (HPR) refers to the process of estimating the configuration of the underlying body positions and its skeletal structure of a human body in 3D [1]. Recently, HPR has become an active research topic in human computer interaction (HCI), as gesture interface technology was selected as one of top 10 emerging technologies in 2011 [2]. HPR is generally considered as a challenging problem due to the complexity and diversity of human movements.

Recently, depth sensors [3] have been introduced which provide images of distance between a sensor and an object. These new imaging devices are actively employed for HPR to be used in smart home, rehabilitation, and health care systems. The depth sensor-based HPR methodologies can be categorized into two different approaches. One is performed by recognizing human body parts (i.e., classifying each depth pixel into each body part) through supervised classification with training a classifier and the other is via unsupervised classification by analyzing the depth map itself without training.

For the supervised approach, a specific classifier is trained in the learning phase with a large set of training data. Once the classifier is trained, all pixels in every input silhouette is classified into the body parts and assigned labels. Then from the recognized body parts, their joint positions and angles are identified and mapped into a synthetic body model to recover a 3D pose of a subject [3]. One representative example of this kind approach is the work of Shotton et al. [3] in which 31 human body parts are recognized and labeled from a depth silhouette using Random Forests (RFs) classifier. This

methodology requires training of a classifier with a suitable set of training data.

As for the unsupervised approach, recently HPR using Geodesic distance map (GDM) of human depth silhouettes is getting an attention. One related study is the work of Correa et al. [4] in which five primary joint locations or landmarks (head, two hands and feet) found from GDM. They utilized the Dijkstra algorithm which has an advantage of finding the five primary landmarks. One of the limitations of this methodology is that it is difficult to find the landmarks reliably. Especially, the secondary landmarks such as elbows, knees, etc. are more difficult to find from GDM.

In this paper, we present a novel approach to reconstruct a 3D human pose from GDM of human body silhouettes captured by a single depth camera. We extend the work of Correa et al. [4] by twofold: 1) we propose a more stable algorithm to find the primary and secondary landmarks through Principal Component Analysis (PCA) of GDM and 2). We propose fitting of a kinematic human body model with the length constraints of body parts to the landmarks found from GDM to overcome the limitation of the landmark detection. We have tested the proposed algorithms in qualitative and quantitative ways with synthetic and real data. Our experimental evaluations demonstrate that the proposed technique is effective and could be applied in real-time HPR in an unsupervised way.

2. METHODS

Our proposed methodology for HPR in 3D is summarized in Fig. 1, consisting of the following five main stages. First, in the preprocessing stage, we obtain a depth image from a depth camera. From the depth image, we extract a human depth silhouette. Then, we initialize our system with a T-pose and identify the key joints and set the lengths and labels of the body parts. Second, we compute GDM from each human depth silhouette which represents the distance map from the body centroid to all other pixels in the silhouette.

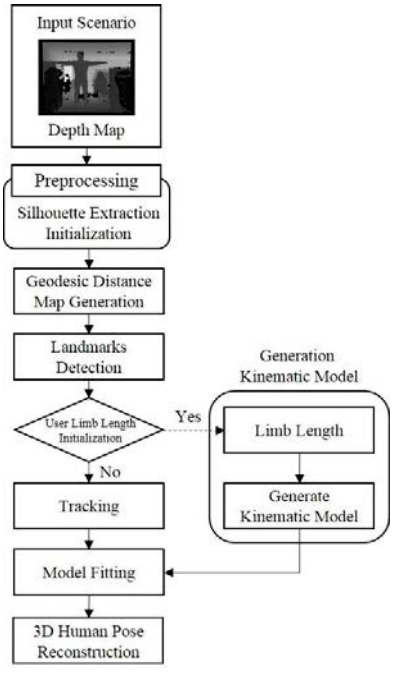


Figure 1. Flow Diagram of Our Approach

Third, detection of body landmarks is performed. To find the landmarks, we derive the shortest path from the centroid to all other pixels on the human depth silhouettes by applying the Dijkstra algorithm to GDM [5]. The end points of the shortest path indicate the primary landmarks including two hands, two feet, and the head. Then, the secondary landmarks are found by intersecting two eigenvectors of the shortest path. Fourth, a set of the same landmarks derived from a kinematic body model is fitted to the 12 landmarks found from GDM by optimizing the joint angles of the kinematic model. To identify the label of each joint, we track the label of each landmark from the initialization by utilizing the k-Nearest Neighbor tracking algorithm [6]. Finally, we reconstruct a 3D human pose by applying the optimized joint angles to an ellipsoid 3D body model.

2.1 Depth Imaging

We obtain a depth image from a depth camera [3]. The imaging parameters of the depth camera are of an image size of 640x480, field of view 57.5, 45, 69 degree (horizontal, vertical, and diagonal), frame speed of 30fps, and operation range 0.8~3.5m.

2.2 Preprocessing and Initialization

We extract a human body silhouette from a depth image by removing background. Details of our background removal are available in [3]. In the initialization stage, we use a T-pose to identify the initial joint positions of human depth silhouettes and the length of the body parts. Based on this information, we create a 3D forward kinematic model using the Robotics Toolbox [7].

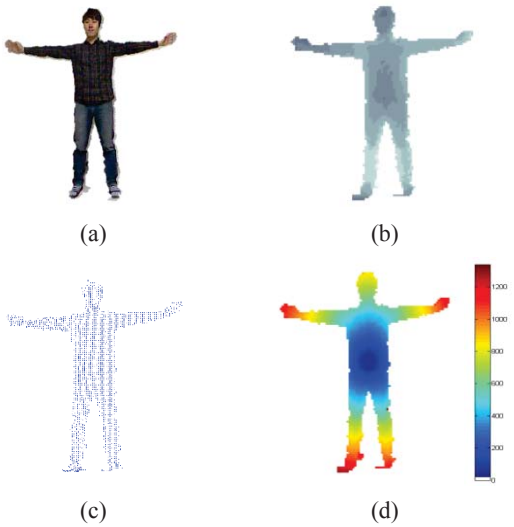


Figure 2. (a) A human RGB silhouette, (b) human depth silhouette, (c) 3D point cloud, and (d) Geodesic distance map. The color bar reflects distance from the body centroid.

2.3 Geodesic Distance Map Generation

We generate GDM from the body centroid G_c to all other pixels on the silhouette. With the 3D point clouds of the human depth silhouette, GDM is formed by Eq. (1).

$$G_t = (Z_t, E_t) \quad (1)$$

where, Z_t are the positions in 3D as vertices and E_t the edges between the vertices at frame *t*. Fig. 2 shows the silhouettes and GDM.

2.4 Landmarks Localization

We obtain a set of 12 joint landmarks consisting of 5 primaries (head, Right(R)/Left(L) hands, and R/L feet) and 7 secondaries (neck, R/L shoulders, R/L elbows, and R/L knees). In order to estimate the primary landmarks, we remove pixels whose distances are less than the estimated length of torso from the initialization stage. Then we obtain the 5 isolated groups of pixels, head and four limbs. For each group, a point having the longest geodesic distance becomes a primary landmark such as the head, two hands and two feet.

Our process of the secondary landmarks localization performs the following steps. We derive paths from the body centroid to the primary landmarks using the Dijkstra algorithm, resulting in the five main paths which are the shortest paths from the centroid to the five primaries. In order to extract the upper and lower part of each main path, we divide the five main paths and call them as the subpath (head, spine, R/L shoulders, R/L upper-arms, R/L lower-arms, R/L upper-legs, and R/L lower-legs). We utilize a 3D line fitting algorithm using PCA to the twelve subpaths for detection of the secondary landmarks. First, in order to estimate the directional information of each subpath, we derive a covariance matrix. Second, from the covariance matrix, twelve eigenvectors with the maximum eigenvalue for each subpath is estimated using PCA. Using these eigenvectors, candidates of the secondary landmarks are detected by finding an intersection point of the two neighboring eigenvectors (e.g., intersection of eigenvectors of the upper and lower arms). Lastly, we relocate the intersecting point to the middle of region of each secondary landmark. Here, we finally obtain the 5 primary landmarks and 7 secondary landmarks in total.

2.5 Tracking

In order to track the label of five primary landmarks (head, R/L hands, and R/L feet) from the initialization frame to the following frames, a k-Nearest Neighbor (k-NN) classifier is applied [6]. The tracked labels of the landmarks are used to determine the current labels of the primary landmarks.

2.6 Fitting of a 3D Kinematic Human Model

As the last step, we fit a kinematic human body model. The kinematic model has a total of thirteen joints (torso, neck, head, R/L shoulders, R/L elbows, R/L hands, R/L knees and R/L feet) with two degree-of-freedom (DOF) per each joint (i.e., each joint has 2 angles of pitch and roll). Our model fitting is done by optimizing a set of joint angles of the model by minimizing the sum of Euclidean distance errors between the landmarks of the model and those identified from GDM. Eq. (4) shows the cost of this fitting process.

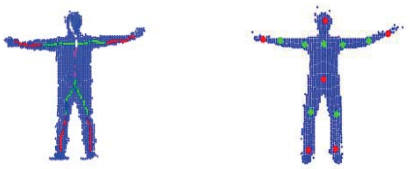
$$\mathrm{Err}(\theta_{cand}) = \frac{1}{N}\sqrt{\sum(L_i - M_i)^2}\ (i = 1, 2, \dots, N) \quad (4)$$

where θ_{cand} is the parameters of 24 joint angles. Err() is a cost function of minimizing the Euclidean distances between the landmarks and the model's joints. N is the number of landmarks. i is the index of joints. L_i and M_i are 3D position of the landmark (L) and model (M). With the error function, a set of angles estimated by Eq. (5).

$$\arg\min_{\theta} Err(\theta_{cand}) = \theta^*_{joints}\{\theta^*_1, \theta^*_2, \dots, \theta^*_{24}\} \quad (5)$$

where θ^*_{joints} are the estimated joint angles.

We utilized a Newton's optimization method [8]. Finally, we reconstruct a 3D pose by mapping the optimized angle set θ^*_{joints} to our ellipsoid 3D body model.



(a) (b)

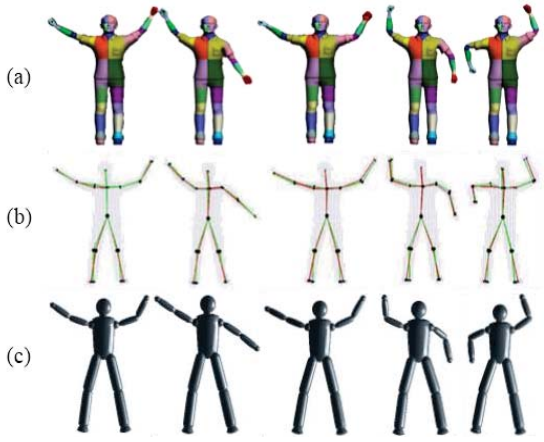
Figure 3. (a) The lines indicate the shortest paths from the body centroid to the extremities and (b) the primary landmarks (red dots) and the secondary landmarks (green dots) superimposed on the depth silhouette.

3. RESULT

3.1 3D Pose Estimation with Synthetic Data

We created a set of synthetic pose data using Autodesk 3Ds Max [9] as some sample poses are shown in Fig. 4 (a). The primary and secondary landmarks obtained from these poses are used as the ground-truth for quantitative analysis. Fig. 4 (b) shows the ground-truth skeletal poses in red and the estimated poses in green superimposed on their corresponding silhouettes. Fig. 4 (c) shows the reconstructed poses with the ellipsoid body model in 3D.

To evaluate our proposed methodology quantitatively, we examined the Euclidean distance error between the landmark positions from the synthetic ground-truth poses and the optimized poses with randomly selected 50 poses. Table 1 shows the evaluation results: it shows the range of error from min. of 0.01mm to max of 33.45mm. The average error for the twelve landmarks is 6.63mm. One should note two factors regarding these results: one is that the depth sensor has its own resolution error of 12mm. The other is the landmarks are found on the surface of the body since the depth silhouette reflects only the surface. This makes also the kinematic skeletal model is fitted to the surface of the depth silhouette. However, there is an error caused by the positional difference between the surface and skeleton in the body volume. This means that the actual location of the skeleton is within the body volume. In addition, different volume of each body part contributes to the errors differently. A correction routine might be needed if one hopes to fit the model to the true volumetric center of the body part, but this requires an estimation routine to the true center of the landmark from the depth body part.



(a)
(b)
(c)

Figure 4. Results of 3D pose estimation with synthetic data: (a) poses of a synthetic model, (b) estimated skeleton poses (truth in red and estimated in green lines) superimposed on the depth silhouettes, and (c) estimated 3D ellipsoid poses

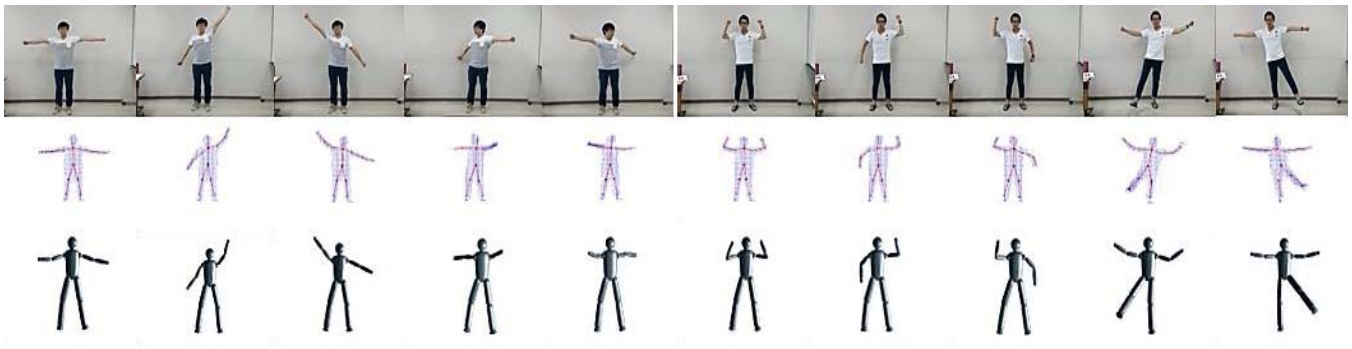


Figure 5. Results of HPR with real data with the proposed methodology: (a) RGB images, (b) estimated 3D skeletal poses superimposed on the depth silhouettes, and (c) reconstructed 3D poses with an ellipsoid body model.

Table 1. Euclidean Distance Errors Between the true locations of the landmarks to the detected landmarks on the depth silhouettes

Landmarks	Neck	Head	Right Shoulder	Right Arm	Right Hand	Left Shoulder
Min/Max (mm)	0.01/ 24.83	0.01/ 15.73	0.01/ 2.15	0.01/ 18.61	0.01/ 0.37	0.01/ 5.76
Mean (mm)	14.14	6.60	0.25	7.47	0.08	1.79
Standard Deviation (mm)	8.57	5.06	0.49	5.78	0.10	1.92
Landmarks	**Left Arm**	**Left Hand**	**Right Leg**	**Right Foot**	**Left Leg**	**Left Foot**
Min/Max (mm)	0.01/ 9.58	0.01/ 5.21	0.05/ 19.54	0.01/ 33.45	0.03/ 18.48	0.02/ 29.87
Mean (mm)	3.43	0.91	6.67	16.34	6.07	15.79
Standard Deviation (mm)	2.97	1.50	6.20	11.80	5.47	9.05

3.2 3D Pose Estimation with Real Data

We performed the experiments with two subjects asked to pose freely. To the captured depth silhouettes, we applied the proposed methodology and evaluated the results only qualitatively with only the reconstructed pose, since the true landmarks cannot be detected. The experimental results are shown in Fig. 5. Fig. 5 (a) shows a set of RGB images, Fig. 5 (b) their corresponding depth silhouettes with the estimated skeletal poses in red lines, and Fig. 5 (c) the reconstructed 3D poses with the ellipsoid model. The left five columns are from Subject 1 and the right five from Subject 2. The results show that our proposed methodology could recover the poses reliably in an unsupervised way without some trained classifiers.

4. CONCLUSION

In this work, we present an unsupervised 3D human pose reconstruction methodology from human depth silhouettes. Our proposed methodology can be applied for 3D HPR without trained classifiers. A future direction is to extend the proposed methodology to hand pose estimation. We expect that our proposed methodology could be utilized in the fields of rehabilitation and health care.

5. ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2008-0061908).

6. REFERENCES

[1] G. O. Young, “Synthetic structure of industrial plastics”, *New York: McGraw-Hill*, 1964.

[2] W.-K. Chen, “Linear networks and systems: algorithms and computer-aided implementations,” *World Scientific*, 1990.

[3] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, "Real-time human pose recognition in parts from a single depth images," *Communications of the ACM*, vol. 56, pp.116-124, 2013.

[4] P. Correa, F. Marques, X. Marichal, B. Macq, “3D Posture Estimation using Geodesic Distance Maps,” *Multimedia Tools and Applications*, vol. 38, pp.365-384, 2008.

[5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, “Introduction to Algorithms,*” MIT Press and McGraw-Hill*, 2001.

[6] R. O. Duda and P. E. Hart, “Pattern Classification,” *New York: John Wiley & Sons*, 1974.

[7] Robotics Toolbox: http://www.mathworks.com/matlabcentral/linkexchange/links/2961-robotics-toolbox-for-matlab

[8] J. C. Gilbert and C. Lemarechal, “Some Numerical Experiments with Variable-Storage Quasi-Newton Algorithms,” *Mathematical Programming*, vol. 45, pp.407-435, 1989.

[9] Autodesk 3Ds MAX 2012 : http://www.autodesk.co.kr/