Nonrigid Point Set Registration-based 3-D Human Pose Tracking from Depth Data

Dong-Luong Dinh[1], Nguyen Duc Thang[2], Sungyoung Lee[1], and Tae-Seong Kim[3]

[1]Department of Computer Engineering, Kyung Hee University, South Korea
[2]Department of Biomedical Engineering, International University, Vietnam
[3]Department of Biomedical Engineering, Kyung Hee University, South Korea

Abstract– **In this paper, we present a novel approach of recovering a 3-D human pose from a single human body depth silhouette using nonrigid point set registration. In our methodology, a human body depth silhouette is presented as a 3-D points set that is matched to the next 3-D points set through point correspondences between them. To recognize and maintain the body part labels, we first initialize the initial points set and their corresponding body parts, then transform them to the next points set according the point correspondences via nonrigid point set registration. Upon the point registration, we use the information of the transformed body labels of the registered pose to create a human skeleton model. Finally, a 3-D human pose is recovered by mapping the skeleton's position and orientation information to a 3-D synthetic human model. Our quantitative and qualitative evaluation on synthetic and real data show that complex poses could be tracked and recovered reliably.**

Keyword—***Human pose estimation, coherent point drift, depth image, point set registration.***

I. INTRODUCTION

Recently, 3-D human pose recovery from depth silhouettes has become an active research topic in computer vision, especially for complex human poses. This research work is triggered with an introduction of depth imaging devices which provide pixel-by-pixel distance images. Furthermore, from a sequence of depth images, a series 3-D poses, representing motion could be tracked and recovered. This research challenge is driven by many potential applications such as entertainment game, surveillance, sport science, health care technology, human computer interactions, motion tracking, and human activity recognition [1].

Many studies of this human pose recovery from depth silhouette have appeared in recent years [2]. To recover a 3-D human body pose from depth data, the techniques could be categorized into three, namely the graph-based, labeled body parts-based, or point set registration-based.

In the category of the graph-based, in [3], [4], and [5], to recover a 3-D human pose, they represented the depth data in a graph-based representation and then estimated geodesic distances of the graph to find the positions and orientations of primary human body parts such as head, hands, and feet. The computation cost of this technique is effective. However, these methods revealed some limitations. The number of detected body parts based on primary landmarks is limited and the detected parts do not identified left or right body parts. In addition, the graph topology is sensitive to occlusions of body parts in where geodesic distance could not find a continuous path since 3-D data is disconnected or interrupted, therefore, the results of detected body parts are unstable.

In the category of body parts labeling-based, in [6], [7], and [8], they proposed the effective method to human pose recognition in body parts from a single depth silhouette inferred from a per-pixel classification via some randomized decision trees. This approach allows efficient recognition of human body parts. It could recognize up to 31 body parts from a single human depth silhouette. However, these studies required a large database for training. The training database has to be created from prerecorded motion data for automatic pixel labeling. For this reason, misrecognitions will occur if the database used for training is not properly and adequate. In addition, in some complex human poses, which contain hands or legs crossing body parts, had a low recognition accuracy of these body parts.

In the third category, the point set registration is to find point correspondences between two different point sets of rigid or nonrigid objects. For registration of 3-D shape objects, many algorithms have been proposed in [9]. Iterative Closest Point (ICP) is one of the well-known fitting or registration algorithms between two sets and it has been widely used for several applications such as 3-D model fitting, shape registration, and human motion tracking [9]. For instance, in [10] and [11], they utilized ICP algorithm to fit the 3-D human body model on the 3-D articulation data in a hierarchical manner. However, the main drawback of ICP requires that the initial position of two given point sets is adequately close. Therefore, this method may return the local optima in some complex poses of a nonrigid object like the human body.

As the approaches mentioned above, to improve robust 3-D human pose recovery from depth silhouette including complex poses, we propose a new methodology to recover 3-D human pose from depth data by tracking human body parts using nonrigid point set registration as presented in Fig. 1. Coherence Point Drift (CPD) [12-14] is used for nonrigid point set registration. This technique allows recovered human poses to maintain their structure by preserving the motion coherence constraint during the matching process of this algorithm. In our approach, to find joint points of body parts for recovering 3-D human pose, we first initialize the initial points set and their corresponding body parts, then transform them to the next points set according the point correspondences via nonrigid point set registration. Based on the point registration, we use the body labels information of the registered pose to find the joint points and create a human skeleton model. Finally, a 3-D human pose is recovered by mapping the skeleton's joints positions and orientations to a 3-D synthetic body model.

The paper is structured as follows. In section II, we describe our proposed 3-D human pose recovery methodology. Section III presents experimental settings and obtained experimental results on both synthetic and real data. Conclusion remarks are given in sections IV.

II. THE PROPOSED 3-D HUMAN POSE RECOVERY METHOD

Fig. 1 describes the step by step of our processing framework. At initial step, a synthetic human depth map and corresponding body parts labeled map (f_o) are used to initialize the system. Given a human depth silhouette ($f_{i=1..n}$), which is extracted by removing the background, is uniformly down sampled and presented as a 3-D points cloud (f_i). The points set of this silhouette is then aligned with the points set (f_{i-1}) of previous pose to get point correspondences between them. By using the point correspondences and the body parts labeled map of the previous pose, human body parts of the given human depth silhouette are recognized. Joint position proposal is then applied on the recognized body parts. From determined joints, the known orientations and positions of the human body parts are applied on a 3-D synthetic human model [8].

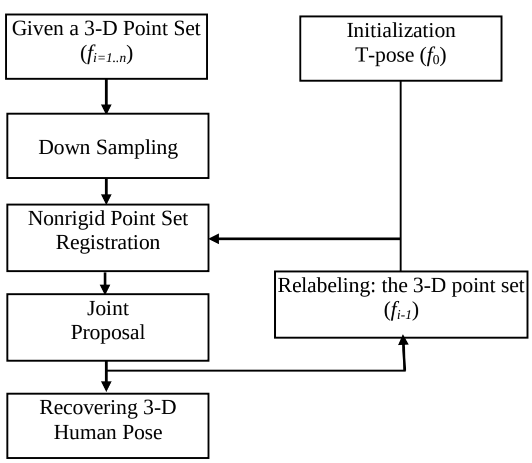


Fig. 1. The flowchart of our pose proposed 3-D human tracking framework from depth data

A. Initialization

Initialization is performed once in the beginning to help our system to identify the body part labels in the first coming depth silhouette.

B. Depth silhouette presentation

Let *X*, *Y*, *Z* be coordinates of points in 3-D space followed by x, y, and z dimensions, respectively. To convert 3-D data from depth image, the corresponding relationship between the coordinates of the scene points and these pixel of depth images are expressed as

$$X = c\frac{Z}{f}, \quad Y = v\frac{Z}{f}, \quad Z = Z, \tag{1}$$

where, the distance *f* is the focal length, *Z* the distance from camera to object (depth values), *c* and *v* are the column index and row index of the pixels in depth image.

C. Down sampling

To decrease computation cost and improve effective point set registration, we utilize a uniform down sampling method for this purpose as presented in Fig. 2.

(a) (b) (c)

Fig. 2. Depth silhouette presentation and down sampling, (a) depth data, (b) a 3-D points set, and (c) a uniform sampled points set.

D. Nonrigid point set registration

This part presents how to find point correspondences between two sets of points. We utilize the nonrigid point set registration method to find point correspondences between two complex point sets of human poses. To preserve the topological structure of human poses, we apply CPD [13], [14] to obtain the correspondences on the two point sets of human poses.

Given two 3-D points sets of human poses, the points set S_D of the previous pose and the points set S_C of the current pose. These two sets are considered the alignment as a probability density estimation problem in [13]. The CPD algorithm presents as the following:

CPD Algorithm

- Initialize parameters: β, λ
- Construct a Gaussian kernel matrix: *G*
- EM optimization, iterate until convergence
  - E-step: compute the posterior probabilities of GMM components P_r
  - M-step: replace current θ, δ

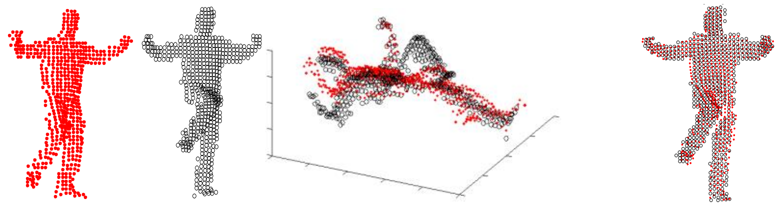
$$\theta, \delta \leftarrow \arg\min_{\theta', \delta'} Q(\theta', \delta' | \theta, \delta)$$

- The aligned point set is $S_C = S_{C\_init} + GW$
- The probability of correspondence is given by P_r

Where β is Gaussian smoothing filter size, λ is smoothness regularization weight, б is standard deviation, θ is a set of the transformation parameters, *G* is a Gaussian kernel matrix of S_C, P_r is a posterior probability, *W* is a matrix of coefficients. Fig. 3 shows the result of point set registration on two poses.

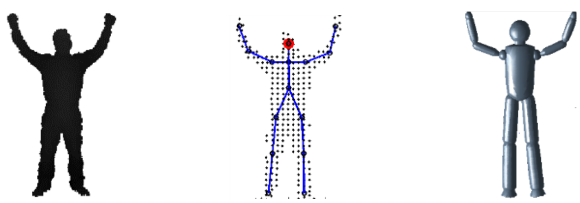
E. 3-D human pose recovery with joints proposals

From the point correspondences, we firstly label body parts of S_C based on the correspondences and the known label information of S_D. Secondly, the positions of joints are located by using mean shift algorithm [6] on each body part. From proposed joint positions, we create a human skeleton model. Finally, the orientation and position of each body part is determined from the skeleton model. The recovered 3-D human pose is presented in Fig. 4.



(a) (b) (c)

Fig. 3. Correspondences in two consecutive frames using non-rigid point set registration of two 3-D human points sets: (a) left: a previous pose, right: a current pose, (b, c) before and after pose illustrations using the point set registration, respectively.



(a) (b) (c)

Fig. 4. Illustrational of results in our proposed system. (a) a depth silhouette, (b) a skeleton model, and (c) a 3-D recovered human pose

F. Relabeling

To track and find the body part labels on the coming depth silhouettes, starting from the second depth silhouette, we use the labeled-parts information of the registered previous pose as target information to map on the depth silhouette of current pose using their correspondences. However, the fact that some points in the set of previous pose contain nonlabeled or mislabeled points. Therefore, these points relabel to correct their labels. In the work, we relabeled the mislabeled points based on distance from centroid points to neighbor pixels and its connectivity matrix.

III. EXPERIMENTAL RESULTS

In this section, we have evaluated our proposed methodology through the quantitative and qualitative assessments using synthetic and real data.

A. Experimental settings

In order to evaluate our proposed system quantitatively, we utilized synthetic depth silhouettes to test with the ground-truth information from the original 3-D body model. At each estimated 3-D human pose, we measured joint angles of few joints from the 3-D human body model and saved as the ground truth. Then, we derived the same joint angles from the reconstructed 3-D human pose and compared them the ground truth. In our experiment, we only focus on the evaluation of the two main joints including left-right elbows.

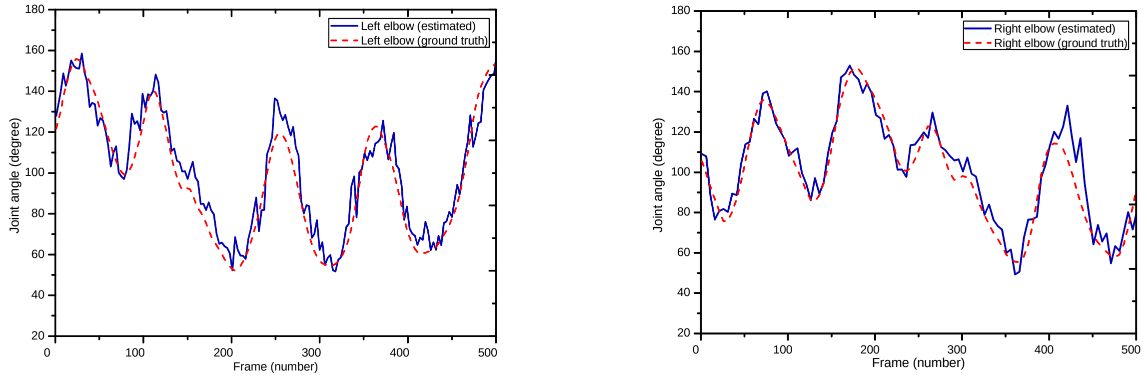
For qualitative assessment on real data, we utilized the coming depth silhouettes that were captured by a depth camera. Then, the human depth silhouettes were registered in our system to find point correspondences. From the point correspondences, the joints of the human body parts were determined. The orientation vectors of the body parts were estimated by joint pairs. These orientations were finally mapped on to the 3-D human body model similar to described in [8], resulting in the estimated 3-D human body pose. The testing process was run on a standard desktop PC with an Intel Pentium IV Dual-core, 2.5 GHz CPU, and 3G RAM.

B. Experimentation on synthetic data

We performed a quantitative evaluation using a series of 500 depth silhouettes containing various unconstrained movements. In this experiment, the evaluation results with the synthetic poses of our proposed methods are provided in Fig. 5. At each plot of Fig. 5 corresponds to an estimated joint angle by our proposed method. The solid and dashed lines indicate our estimated and its ground truth joint angles, respectively.

Based on the results of estimated joint angles and the ground truth joint angles, we have computed the average reconstruction error as

$$\varepsilon_\theta = \frac{\sum_{i=1}^{n} \left| \theta_i^{est} - \theta_i^{grd} \right|}{n} \quad (3)$$



(a) (b)

Figure 5: A comparison between the ground-truth and the estimated joint angles in synthetic data: (a) joint angle of left elbow and (b) joint angle of right elbow.

where n is the number of frames, i is the frame index, θ_i^{grd} is the ground-truth angle, and θ_i^{est} is the estimated angle. The average errors of the experiment at the two joints are given in Table 1.

Table 1: The average reconstruction error of the joint angles in degree

Evaluated angles	Left elbow	Right elbow
Average reconstruction error	6.08	6.02

C. Experimentation on real data

For qualitative assessment of real data, we asked the subject to perform some complex pose sequences of intersected body parts. Because the ground truth joint angles are not available for real data. We only performed by visual inspection of the results of recovered poses and RGB images. Fig. 6 shows sample results of our proposed method on depth images with the occlusion of arm or leg body parts. The 1st and 4th column are RGB images, the 2nd and 5th human depth silhouettes, and the 3rd and 6th recovered 3-D human poses.

IV. CONCLUSION

We have presented a novel approach of recovering a 3-D human pose from a single human body depth silhouette using nonrigid point set registration. The quantitative assessments indicated the average reconstruction error of 6.06 degree. The experiments on real data show that our system reliably performs on sequences containing occlusion movements of various appearance. This approach can also reconstruct some 3-D human complex pose recovery. Moreover, this method does not require any matching or training data and it is able to tracking arbitrary movements. However, the computational cost of the CPD algorithm is still high.

ACKNOWLEDGMENT

This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency (NIPA-2013-(H0301-13-2001)).

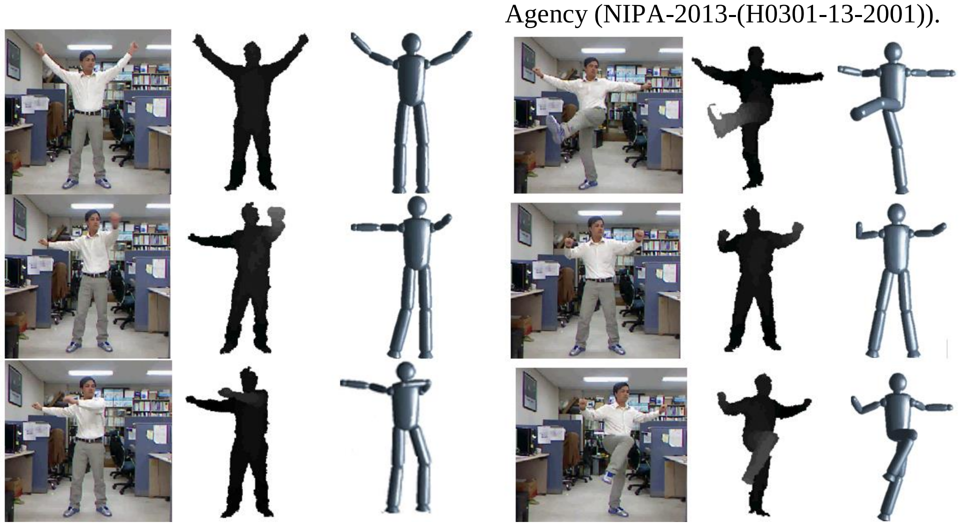


Fig. 6. Sample illustrations of our proposed 3-D human pose recovery method on depth images with the occlusion of arm or leg body parts

REFERENCES

1. Poppe R (2007) Vision-based human motion analysis: An overview, Computer Vision and Image Understanding, Vol. 108, No. 1–2, pp. 4–18
2. Chen L, Wei H, and Ferryman J (2013) A survey of human motion analysis using depth imagery, Pattern Recognition Letters. ISSN 0167–8655
3. Plagemann C, Ganapathi V, Koller D, and Thrun S (2010) Real-time identification and localization of body parts from depth images, In Proc. ICRA, pp. 3108–3113
4. Schwarz L A, Mkhitaryan A, Mateus D, and Navab N (2012) Human skeleton tracking from depth data using geodesic distance and optical flow, Image and Vision Computering, Vol. 30, No. 3, pp. 217–226
5. Kalogerakis E, Hertzmann A, and Singh K (2010) Learning 3D mesh segmentation and labeling, In Proc. ACM SIGGRAPH 2010, ACM, New York, NY, 775 USA, pp. 102–114
6. Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, and Blake A (2011) Real-time human pose recognition in parts from single depth images, In Proc. CVPR
7. Taylor J, Shotton J, Sharp T, and Fitzgibbon A (2012) The Vitruvian Manifold: Inferring Dense Correspondences for One-Shot Human Pose Estimation, In Proc. CVPR, pp. 103–110
8. Dinh D L, Han H S, Jeon H J, Lee S Y, and Kim T S (2013) Principal direction analysis-based real-time 3D human pose reconstruction from a single depth image, In Proc SoICT'13, pp. 206–212
9. Tam G K L, Cheng Z Q, Lai Y K, Langbein F C, Liu Y H, Marshall D, Martin R R, Sun X F, and Rosin P L (2013) Registration of 3D Point Clouds and Meshes: A Survey from Rigid to Nonrigid, IEEE Transactions on Visualization and Computer Graphics, Vol. 19, No. 7, pp. 1199–1217
10. Kim D W and Kim D J (2008) A novel fitting algorithm using the ICP and the particle filters for robust 3D human body motion tracking, In Proc. ACM workshop on Vision networks for behavior analysis, pp. 69–76
11. Mundermann L, Corazza S, and Andriacchi T (2007) Accurately measuring human movement using articulated ICP with soft joint constraints and a repository of articulated models, In CVPR
12. A. L. Yuille and N. M. Grzywacz (1998) A Mathematical Analysis of the Motion Coherence Theory," Int'l J. Computer Vision, Vol. 3, No. 2, pp. 155–175
13. Myronenko A and Song X (2010) Point Set Registration: Coherent Point Drift, IEEE Trans. on PAMI, Vol. 32, No. 12, pp. 2262–2275
14. Ye M, Wang X, Yang R, Ren L, and Pollefeys M (2011) Accurate 3D pose estimation from a single depth image, In Proc. ICCV, pp. 731–738