Multimedia Systems (2017) 23:369–380
DOI 10.1007/s00530-015-0497-y

REGULAR PAPER

3-D human pose recovery using nonrigid point set registration and body part tracking of depth data

Dong-Luong Dinh[1] · Sungyoung Lee[2] · Tae-Seong Kim[3]

Received: 8 August 2014 / Accepted: 22 October 2015 / Published online: 19 November 2015

Abstract In this paper, we present a novel approach for recovering a 3-D pose from a single human body depth silhouette using nonrigid point set registration and body part tracking. In our method, a human body depth silhouette is presented as a set of 3-D points and matched to another set of 3-D points using point correspondences. To recognize and maintain body part labels, we initialize the first set of points to corresponding human body parts, resulting in a body part-labeled map. Then, we transform the points to a sequential set of points based on point correspondences determined by nonrigid point set registration. After point registration, we utilize the information from tracked body part labels and registered points to create a human skeleton model. A 3-D human pose gets recovered by mapping joint information from the skeleton model to a 3-D synthetic human model. Quantitative and qualitative evaluation results on synthetic and real data show that complex human poses can be recovered more reliably with lower errors compared to other conventional techniques for 3-D pose recovery.

Keywords 3-D human pose recovery · Body part tracking · Coherent point drift · Depth image · Point set registration

Communicated by B. Huet.

✉ Sungyoung Lee
sylee@oslab.khu.ac.kr

✉ Tae-Seong Kim
tskim@khu.ac.kr

Dong-Luong Dinh
luongdd@ntu.edu.vn

[1] Department of Information Technology, Nha Trang University, 2 Nguyen Dinh Chieu, Nha Trang, Vietnam

[2] Department of Computer Engineering, Kyung Hee University, 1 Seocheon-dong, Giheung-gu, Yongin-Si, Gyeonggi-do, Republic of Korea

[3] Department of Biomedical Engineering, Kyung Hee University, 1 Seocheon-dong, Giheung-gu, Yongin-Si, Gyeonggi-do, Republic of Korea

1 Introduction

In recent years, with the introduction of depth imaging devices, 3-D human pose recovery from depth silhouettes has become an active research topic in computer vision. Recovering 3-D human poses for complex positions in which body parts such as arms or legs cross and self-occlude is still challenging. These research challenges are driven by everyday applications such as entertainment games, surveillance, sports science, health care technology, human–computer interactions, motion tracking, and human activity recognition [1–4]. Based on recent studies of 3-D human pose recovery from human depth silhouettes [1, 5], techniques can be grouped into two categories: recovering poses frame by frame without using temporal information (without tracking) and recovering poses with tracking using temporal information.

In the first type of approach, a 3-D human pose is recovered from a depth silhouette frame by frame. No temporal information is used in pose recovery, making each frame independently. In this methodology, a technique is required to detect or recognize human body parts. Typically, there are two approaches to detect the body parts by detecting body landmarks from geodesic maps or to recognize body parts via supervised classification. In the geodesic map-based methodology [6, 7], the depth silhouette is represented in a graph-based data structure of 3-D points

(i.e., the geodesic map). Geodesic distances among all 3-D points are computed upon a graph-based representation. Since these geodesic distances are maintained during human movement, anatomical landmarks such as the head, hands, and legs can be detected in the human depth silhouette. For instance, primary landmarks are identified by detecting points with maximal geodesic distances from the body center mass. The primary landmarks of each human depth silhouette are used to recover a corresponding 3-D human pose. One advantage of this approach is its low computational costs, since it uses a simple, graph-based depth data representation. However, its limitations are that some pairs of detected body parts such as hands or feet cannot be distinguished as left and right since they have similar geodesic distances. Also for human poses in which body parts touch or overlap, new connected edges might appear in geodesic map representations, preventing correct detection of the positions of primary landmarks. As a result, primary landmark detection can be unstable [8]. To overcome the drawbacks of this methodology, in the supervised learning-based method [9, 10], body parts in human depth silhouette are recognized based on a pixel-wise supervised classification using trained classifiers to recover a 3-D human pose. This method could recognize up to 31 body parts [9]. In addition, 3-D human pose recovery works in real-time by fast classification via random decision trees (i.e., random forests). However, potential limitations of this method remain. First, the method requires a human pose database of depth silhouettes with corresponding body part-labeled maps for training the body parts classifier. Therefore, misrecognition can occur if the training database does not include enough human poses. Second, each pixel is recognized independently from its neighbor pixels and body parts are labeled based on recognized pixels and their positional distribution in the human depth silhouette. For these reasons, 3-D human pose recovery for some complex poses is prone to errors and failure. Mislabeling of the body parts is derived from misrecognition of pixels. Again, the two approaches of the first type attempt to recover a 3-D human pose from a depth silhouette on a frame-by-frame basis without temporal tracking.

In the second type of approach, the use of spatiotemporal information by tracking poses from a sequence of depth images helps recovery of complex human poses. All forms of generic prior knowledge such as temporal variation of poses in motion and spatial structures of the body poses or parts are critical for assisting human pose recovery using tracking and temporal information. Recently, human body pose tracking in motion via point correspondences using point set registration has been proposed. Iterative closest point (ICP) is a typical point set registration algorithm commonly used for tracking human poses. In studies [11–13], the ICP algorithm was used to find point correspondences between a depth silhouette and a 3-D human model by tracking a human body pose in motion. Potential limitations of this methodology remain, since the ICP requires that the initial position of the given point sets is adequately close and it often fails for complex human poses. Therefore, ICP can return the local optima of body poses for some complex poses.

In this paper, we propose a new methodology for robust 3-D human pose recovery with complex poses via nonrigid point set registration and tracking human body parts using depth data. The key contributions of our proposed approach for the recovery of complex 3-D human body poses are summarized in the following steps: (i) we initialized the first set of points by recognizing corresponding body parts, resulting in a body part-labeled map via a pixel-wise supervised classification as introduced in the supervised learning-based method, (ii) we found point correspondences between two point sets of human depth silhouettes using the coherence point drift (CPD) algorithm for nonrigid point set registration as the first application for tracking human body poses and body part labels. The core advantage of CPD is its capability to preserve the human pose structure when optimal transformations are used with the sets of 3-D points by forcing the sets to move coherently as groups [14], (iii) we tracked human body parts and their labels in a given human depth silhouette, first using the initialized human depth silhouettes from the first step and then tracking them with our proposed relabeling methodology, and (iv) after point registration and tracking, the body parts in the human depth silhouette were detected and recognized. A human skeleton model was created based on the joint information of the identified body parts. Finally a 3-D human body pose was recovered by mapping the orientation of each body part of the skeleton to a 3-D human model.

The remainder of this paper is structured as follows. Section 2 briefly reviews related work on point set registration for recovering a 3-D human pose using depth data. In Sect. 3, we describe our proposed 3-D human pose recovery methodology. Section 4 presents experimental setups, obtained experimental results and comparison with previous works. Concluding remarks are given in Sect. 5.

2 Related work on point set registration

Recent methods do not use depth silhouettes directly but represent depth information in a set of 3-D points as a 3-D surface mesh, tracking human body motion by fitting each body part using point set registration. We give a brief overview of the related works in recovering a 3-D human pose from a series of human depth silhouettes and their assessment.

In point set registration techniques, the fundamental task for shape matching, image registration, deformable motion tracking, and content-based image retrieval can be formulated as a point-matching problem. The point set registration has two main transformations: rigid or nonrigid. Rigid transformation allows only for translation, rotation, and scaling; nonrigid transformation includes anisotropic scaling and skewing methods, and has been applied in real-world applications. Commonly used algorithms for point set registration [16] are ICP and CPD. ICP is a well-known algorithm for fitting or rigid registration between two point sets. Although ICP has been successfully applied to many registration problems, it has several constraints. For example, the position of two given point sets must be adequately close and the two point sets must fully overlap. Also, this method is subject to a local minimum problem. However, the CPD algorithm is based on the probability method [14]. This algorithm searches and assigns point correspondences between two point sets to ensure global optimality of a solution. It preserves the topological structure of registered point sets through Laplacian coordinates [17] using a velocity function for a template point set, namely the centroid of the Gaussian mixture models (GMM), forcing the GMM centroid to move coherently as a group. The CPD algorithm iteratively calculates unknown parameters in the GMM by expectation maximization (EM) [18–20]. As a result, CPD is a more robust and accurate algorithm than ICP [14].

Several studies have tracked human body motion by obtaining point correspondences using point set registration [11–13, 21–23]. To recover 3-D human poses, point correspondences between a given depth silhouette and a 3-D human body model are found using the ICP algorithm. In these studies, a 3-D human body model is commonly represented as an articulated object where body part information can be a general approximation (e.g., cylinders, ellipsoids, and super quadrics) [12, 22] or an estimation of an actual subject's outer surface [13, 21, 23]. ICP was used on each part of an articulated model with depth data to find point correspondences to assist fitting or tracking 3-D human body motion.

Another recent approach used point set registration techniques (ICP, CPD, and maximum a posteriori) to find point correspondences between a given depth silhouette and the most similar 3-D human pose, detected from a template dataset as a local optimization method [24–26]. Point correspondences were used to optimize differences. By combining the most similar pose detections and pose correction techniques, this approach robustly recovered 3-D human poses including complex poses. This approach is considered effective for solving the problem of self-occlusion using a precaptured motion database. However, this approach requires a precaptured motion database and corresponding skeleton body configuration. The approach does not work if the similar pose is not available in the database.

To evaluate 3-D human pose estimation systems, it needs to have ground-truth dataset. Some studies [24, 27–29] utilized a marker-based motion capture system and derive the joint positions as the ground truth. Some other studies [30–32] performed rather qualitative evaluations by visual inspections, if quantitative ground-truth data are not available. One main limitation of the former evaluation is the lack of realistic ground truth dataset which assumed and considered as ground truth dataset to compare with recovered human pose results.

In this study, we have developed a novel method of recovering complex human poses using temporal tracking of human poses and body parts from a series of human body depth silhouettes. We used the advantages of CPD to find point correspondences between two successive sets of 3-D points of human depth silhouettes and to track human body parts. The joint information derived from the detected body parts was used to recover 3-D human body poses. For the quantitative evaluation of our system, we created synthetic data of 3-D poses from which the ground-truth references were derived and compared to ours and two other conventional techniques. Then, we performed qualitative evaluations on real data.

3 Proposed 3-D human pose recovery methodology

Figure 1 describes the steps of our processes. First, our system used a human T-pose depth silhouette, yielding a body part-labeled map by pixel-supervised classification via random forests [15] for initialization. After initialization, each human depth silhouette that was represented as the set of 3-D points was matched to the previous silhouette via point set registration to obtain point correspondences. To track body parts and labels, we first used the initialized body part-labeled map for the human T-pose depth silhouette and then the body part map labeled by our proposed relabeling methodology for the successive frames. From body parts of the human depth silhouette identified by tracking, joint proposals were estimated and a corresponding human skeleton model was created. Finally, 3-D human poses were recovered by mapping the joint information to a 3-D synthetic human model.

3.1 Depth silhouette representation

Depth silhouette is the obtained result from the depth image removed the foreground. To convert a human depth silhouette into a set of 3-D points, let X, Y, and Z be the

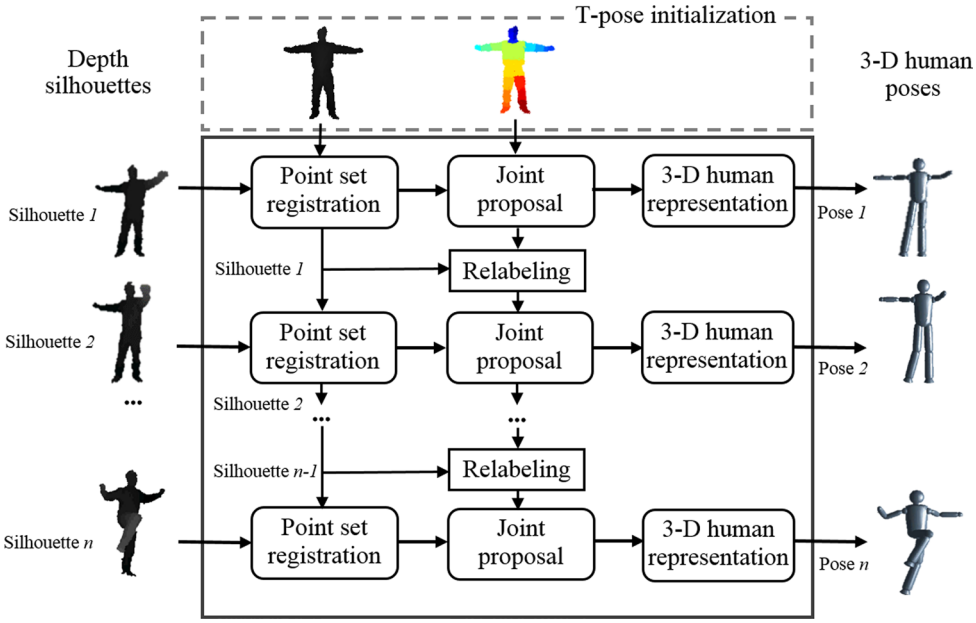


Fig. 1 The framework of our proposed 3-D human pose recovery system from depth data

(a) **(b)** **(c)**

Fig. 2 Depth silhouette representation and downsampling. **a** Depth silhouette, **b** set of 3-D points, **c** uniformly downsampled depth points

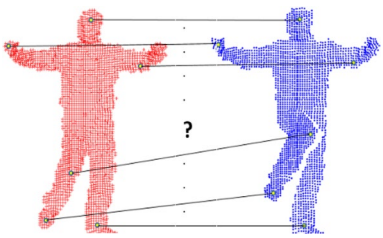


Fig. 3 A point set registration problem. Given two 3-D points sets from two successive human depth silhouettes (*left* human depth silhouette of a previous frame; *right* human depth silhouette of a current frame). The problem is to find point correspondences between the two point sets

coordinates of the 3-D points of the x, y, and z dimensions, respectively. The corresponding relationship between the coordinates of the points and the pixels in a depth silhouette is expressed as Eq. (1). The sample results of converting a depth silhouette to a set of 3-D points are given in Fig. 2.

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

where the distance f is the focal length, D is the distance or depth values, and c and v are the column index and row index of the pixels in a depth image.

3.2 Point set registration

Given two 3-D point sets for two successive human depth silhouettes (let the set of 3-D points S_D be the human depth silhouette from a previous frame and the set of 3-D points S_C be the human depth silhouette from a current frame), the problem of point set registration is finding all point correspondences between the sets. In Fig. 3, we illustrate point correspondences of two successive silhouettes by drawing lines connecting the pairs. We used nonrigid point set registration transformations of CPD, as previously proposed [14], to solve this problem.

To reduce the computational burden from the large number of 3-D points in each human pose, we used a uniform downsampling before point set registration. In our experiments, we selected about 400 points per each silhouette. The results of uniform downsampling are in Fig. 2c. The two point sets were considered to have found the point correspondences Algorithm 1. Demonstrative results of point set registration for two successive human depth silhouettes are given in Fig. 4.

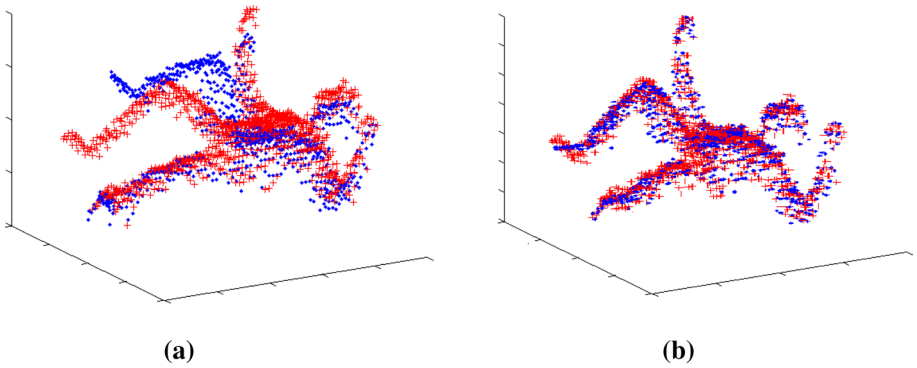


Fig. 4 Nonrigid point set registration between two successive human depth silhouettes from Fig. 3. **a** Before point set registration and **b** after point set registration

Algorithm 1: **CPD for Nonrigid Point Set Registration**

- 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 the Gaussian smoothing filter size, λ is the smoothness regularization weight, б is the standard deviation, θ is the set of the transformation parameters, G is the Gaussian kernel matrix of S_C, P_r is the posterior probability, and W is the matrix of the coefficients.

3.3 Body part tracking via point set registration

3.3.1 Initialization

The goal of the initialization step was to identify body parts and to assign labels on them in a human depth silhouette. Initialization produced a human depth silhouette and corresponding human body part-labeled map of a T-pose. We asked a subject to make a T-pose, obtained a depth silhouette, and created the body part-labeled map. To label body parts on the human depth silhouette of the T-pose, we used a pixel-wise supervised classification via trained random forests [15]: the training needed only a small synthetic T-pose database for labeling body parts. The human depth silhouette and its labeled map with fifteen labeled parts of a T-pose are shown in Fig. 5. This work also helped to eliminate shape and size differences of the human depth silhouettes in the initialization step to get the best result of nonrigid point set registration. In addition, we estimated the height of T-pose as well as the length ratio of body parts to create a corresponding skeleton model.

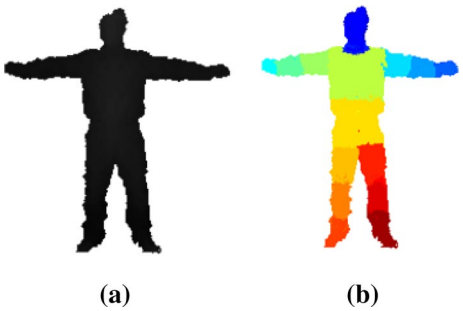


Fig. 5 Initialization using a T-pose. **a** Human depth silhouette and **b** its corresponding human body part-labeled map

3.3.2 Relabeling

After initialization, we tracked the body part labels of successive depth silhouettes and handled drift points using the body part-labeled map from the previous frame S_D as the template to track labels on the point set S_C for an incoming depth silhouette. To this end, we devised a relabeling technique for all points of set S_D which was used as a template set for labeling body parts as presented in Algorithm 2. This work helped our system to reduce the effects of some mislabel points in each body part which were proceed from point set registration errors to find point correspondences. This problem impacted on body parts tracking results. The demonstrative result of our proposed relabeling methodology on the set S_D is presented in Fig. 6.

constraints of body parts in the length ratio D_j. To determine the labels of all points in S_D, LTS started by computing Euclidean distances for all connected point pairs in S_D together. These distances were considered weight values for the connected edges in the weighted graph **G** with vertices v_i corresponding to points p_i in S_D. The purpose of building the weighted graphic $\boldsymbol{G}$ was to effectively use the shortest path algorithms for finding the shortest geodesic distances between any point pairs of **G**. On the graph **G**, the geodesic distance matrix $\boldsymbol{M}$ of all **G** vertices v_i to the other vertices that corresponded the joint positions of J_j in **G** was constructed by finding the shortest path using Dijkstra's algorithm [33]. Finally, all points of S_D were assigned the label of the corresponding joint J_j using the highest likelihood of $P(d_{geo}(v_i, J_j)|D_jH)$.

Algorithm 2: **Labeling Template Set (LTS)**

- **Inputs**:
  - Given a set of points S_D
  - Fifteen joint positions of labeled body parts J_j
  - Length constraints of body parts defined in the length ratio D_j and estimated human height H
- **Output**: All labeled points in S_D
- **Step 1**: Represent the point set of S_D as a weighted graph $\mathcal{G}$ in which Euclidean distances of all point pairs in S_D are considered the weight values of edges in $\mathcal{G}$ and its vertices v_i corresponding to points p_i in S_D.
- **Step 2**: Construct the geodesic distance matrix $\mathcal{M}$ of $\mathcal{G}$ by computing the shortest geodesic distances d_{geo} of each vertex v_i to all joints J_j by using Dijkstra's algorithm.
- **Step 3**: Assign the label L for the vertices v_i of $\mathcal{G}$ into groups of the J_j by means of the highest likelihood as

$$L_{v_i} = \arg\max_{P_{j=1,2,\ldots 15}} \{P(d_{geo}(v_i, J_j) \mid D_j H)\},$$

$$P(d_{geo}(v_i, J_j) \mid D_j H) = \frac{1}{1 + e^{\alpha(d_{geo}(v_i, J_j) - D_j H)}}.$$

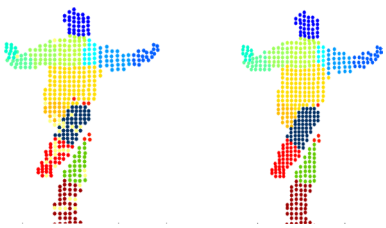


Fig. 6 A sample result of our proposed relabeling methodology: **a** before relabeling and **b** after relabeling

In Algorithm 2, we relabeled the set of points S_D and used it as a template set for registration with S_C based on the known joint positions of body parts J_j and the defined length

3.4 Joint proposal and 3-D human pose representation

Based on the results of point set registration and relabeling, the body parts of each human depth silhouette were detected and identified by tracking. From the labeled body parts, we estimated joint proposals to represent corresponding body parts. Joint positions for each body part were determined using the mean shift algorithm [34]. This algorithm uses the subsets of 3-D points having the same label and returns the centroids of body parts. From the proposed joint positions and defined length constraints of each body part, we created a human skeleton model. Finally, orientation information of each

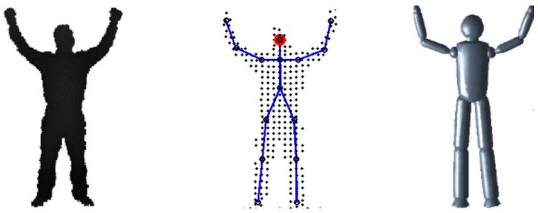


Fig. 7 Illustration of main results in our proposed human pose recovery system. **a** Human depth silhouette, **b** human skeleton model, and **c** 3-D human pose recovery

body part was determined from the skeleton model to map on a 3-D human model for representing a 3-D human pose recovery as described previously [15]. The joint proposal and 3-D human pose representation are shown in Fig. 7.

4 Experimental results

4.1 Experimental setups

To evaluate our system quantitatively, we need to have ground truth dataset for this work. In our work, we used Carnegie Mellon University (CMU)'s motion capture (MoCap) data [35] that contains sets of human joints from which joint angles were directly extracted as the ground-truth data and these sets of human joints were used to apply 3-D synthetic model and to create the dataset of depth silhouettes that was used as input of our system and the results compared with the ground-truth. The main steps are presented in Fig. 8. Herein, each human joint configuration of 3-D human pose in MoCap file (*.bvh), it was used to directly extract joint angles from the eight main joints including left and right elbows, shoulders, knees and hips which were considered in our experiments, to save them as ground truth data. Then, this joint configuration of 3-D human pose was also used to generate corresponding depth silhouettes for recovering 3-D human pose via our system. The detail of these steps is that the joint configuration of 3-D human pose was firstly mapped to 3-D synthetic model using a commercial 3-D graphic package [36] named Motion Builder. Then, depth silhouettes were created from 3-D synthetic model using 3-Ds Max packages [36]. These depth silhouettes were used as input of our human pose recovery system and the results of eight estimated joint angles via our system compared with the ground-truth data.

For the qualitative assessment of our proposed methodology with real data, we used human depth silhouettes captured by a depth camera. Each 3-D human pose was recovered from a single human depth silhouette using nonrigid point set registration and body parts tracking since ground truth poses from real data were not available. The results of the recovered 3-D human poses were evaluated by visual inspection of recovered poses and corresponding color images.

Our system was implemented in both Matlab and C++. We ran our experiments on a 3.5 GHz Pentium Core i5 and

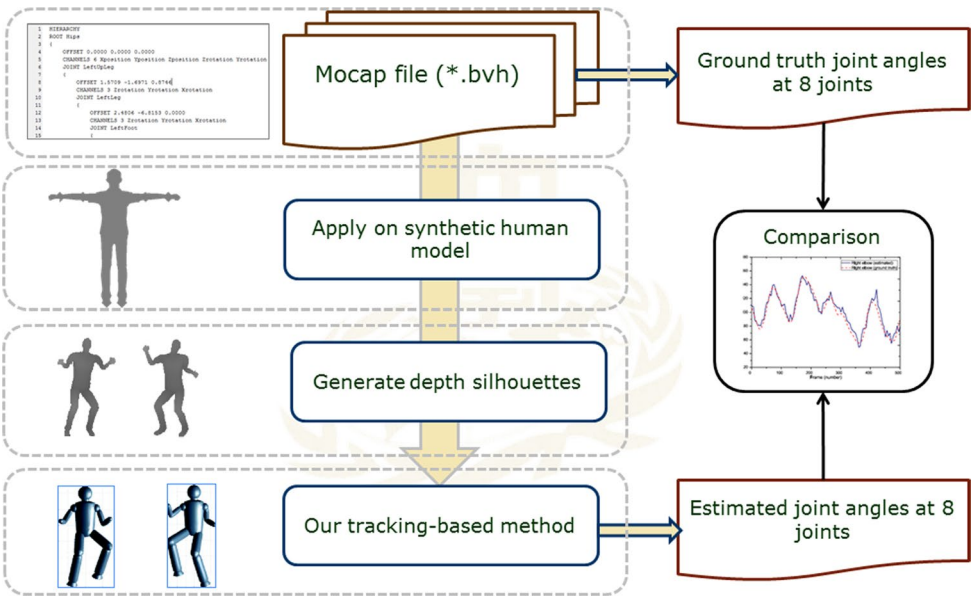
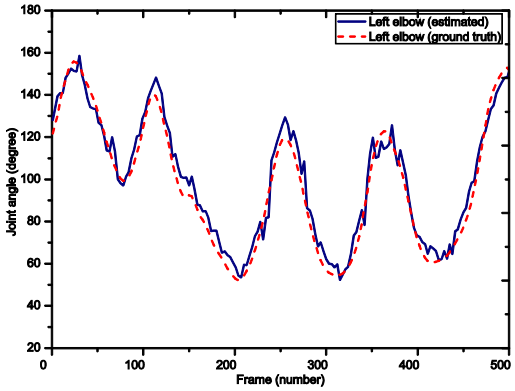
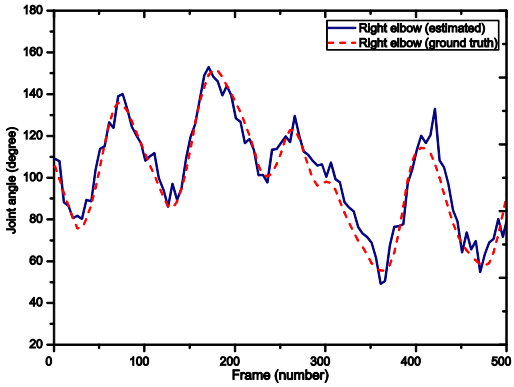


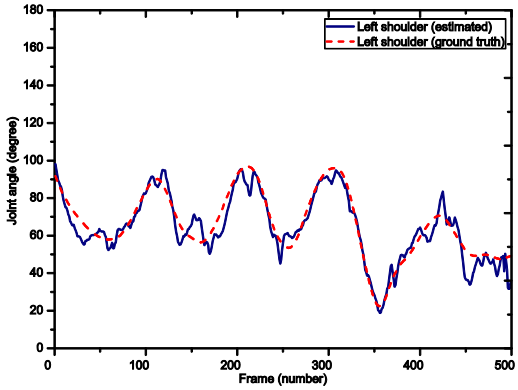
Fig. 8 The main steps of the quantitative evaluation on our system


Left elbow (estimated)
Left elbow (ground truth)
Joint angle (degree)
Frame (number)

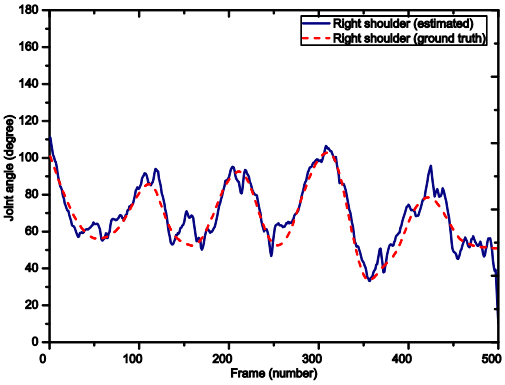
(a)


Right elbow (estimated)
Right elbow (ground truth)
Joint angle (degree)
Frame (number)

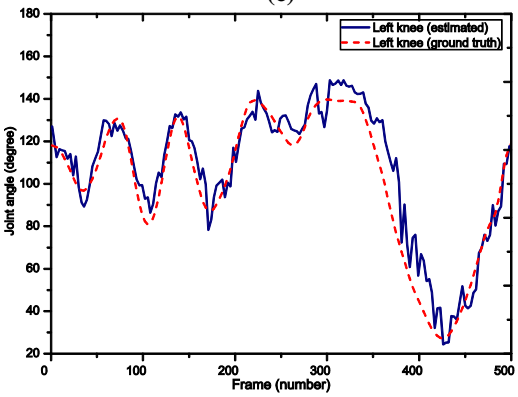
(b)


Left shoulder (estimated)
Left shoulder (ground truth)
Joint angle (degree)
Frame (number)

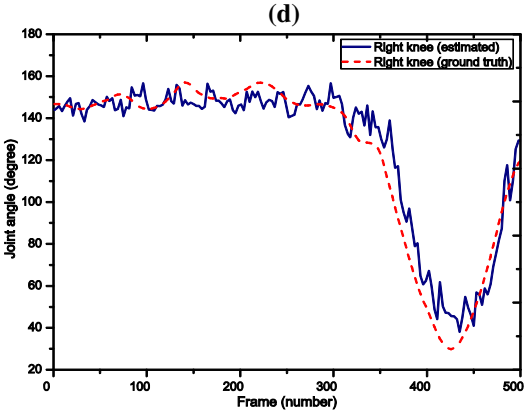
(c)


Right shoulder (estimated)
Right shoulder (ground truth)
Joint angle (degree)
Frame (number)

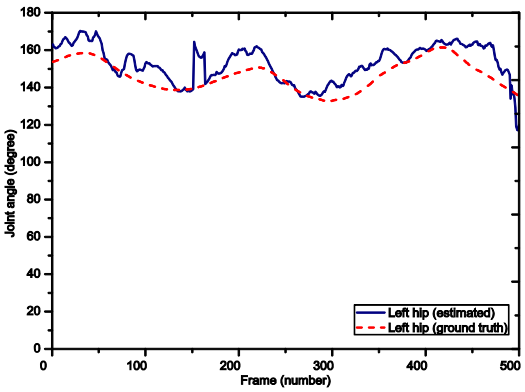
(d)


Left knee (estimated)
Left knee (ground truth)
Joint angle (degree)
Frame (number)

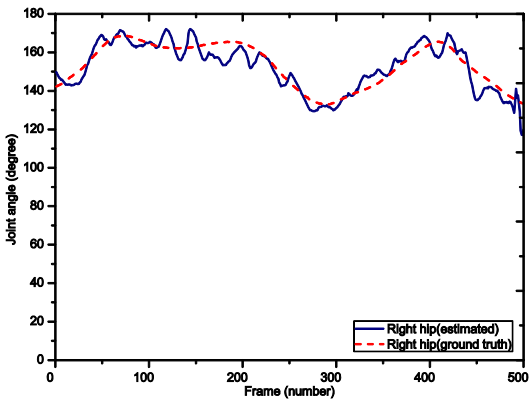
(e)


Right knee (estimated)
Right knee (ground truth)
Joint angle (degree)
Frame (number)

(f)


Left hip (estimated)
Left hip (ground truth)
Joint angle (degree)
Frame (number)

(g)


Right hip(estimated)
Right hip(ground truth)
Joint angle (degree)
Frame (number)

(h)

◀**Fig. 9** Comparison of ground truth and estimated joint angles in synthetic data. **a** Joint angle of left elbow, **b** joint angle of right elbow, **c** joint angle of left shoulder, **d** joint angle of right shoulder, **e** joint angle of left knee, **f** joint angle of right knee, **g** joint angle of left hip, and **h** joint angle of right hip. *Solid lines* indicate the estimated joint angles and the *dashed lines*, the ground truth joint angles

4 GB of RAM. Computation cost of our human pose recovery system is about 0.4~0.6 s per frame. Most of the used computation time is at the step of set point registration of CPD. Therefore, the computation time is dependent on the rate of downsampling as well as the size of human depth image.

4.2 Experiments with synthetic data

We performed a quantitative evaluation using a series of 500 depth silhouettes of various poses created from motion information of CMU MoCap DB [35]. Figure 9 shows the quantitative evaluation of the eight corresponding measured joints including left and right elbow, shoulders, knee and hip joints using our proposed method.

We computed the average reconstruction error between the estimated and ground-truth joint angles as:

$$\varepsilon_\theta = \frac{\sum_{i=1}^{n} |\theta^{\text{est}} - \theta^{\text{grd}}|}{n} \tag{2}$$

where n was the number of frames, i was the frame index, θ_i^{grd} was the ground-truth angle and θ_i^{est} was the estimated angle. Average errors at the eight joints were 5.54°, 5.72°, 5.32°, 5.64°, 8.02°, 8.27°, 6.40°, and 5.21° as presented in Table 1. The average reconstruction error of the eight joint angles was 6.26°.

For comparisons against the conventional methods, we have evaluated the performance of our proposed methodology by comparing against the conventional methods [9, 15]. For comparison against mean shift-based method [9], we reproduced the human pose recognition and used our synthetic DB in [15] to train RFs. We evaluated the mean shift method through quantitative assessments on the synthetic dataset. The quantitative assessment result with the same synthetic DB of our proposed method and the method [9] is presented in Table 1. The average reconstruction errors at eight joints of the left and right elbows, shoulder, knees and hips of the method [15] were 9.24°, 9.41°, 9.12°, 9.81°, 9.91°, 10.15°, 10.34°, and 9.67° compared to 5.54°, 5.72°, 5.32°, 5.64°, 8.02°, 8.27°, 6.40°, and 5.21° of our methods. For comparison against the principal direction analysis based method [15] which is based on body-part labeling as introduced in the supervised learning-based method. The mean reconstruction error of the PDA method at eight key joints was 7.10° compared to 6.26° of our method. The detail of the average error at each joint is presented in Table 1. The average reconstruction errors of our method were better than the PDA method at the right and left knee joints, while the average reconstruction errors were similar for both methods for the right and left arms. PDA is based on the labeling methodology and our proposed method is based on the point set registration-based methodology. However, our proposed method focuses on improving the results of more complex human poses, which is still a challenge for pose reconstruction by tracking body parts of each pose.

4.3 Experiments with real data

To evaluate real data, we asked a subject to perform arm and leg movements with simple and complex pose sequences. Two experiments were performed. Figure 10 shows sample results of the first experiment on depth images with the crossing of an arm or leg body parts in poses: first and fourth columns, color images; second and fifth columns, human depth silhouettes; third and sixth columns, recovered 3-D human poses using the proposed method.

We performed a second experiment with complex poses. Figure 11 demonstrates results of recovering 3-D poses with self-collisions or overlapping of arm or leg body parts: first and third rows, color images; second and fourth rows, recovered 3-D human poses from the proposed method.

For comparison of the conventional method with real data, we used the 3-D human pose recovery system using PDA as described [15]. We qualitatively evaluated PDA

Table 1 Comparison of average reconstruction error (degrees)

Evaluated angles	Left elbow	Right elbow	Left shoulder	Right shoulder	Left knee	Right knee	Left hip	Right hip	Mean
Shotton et al. [9]	9.24	9.41	9.12	9.81	9.91	10.15	10.34	9.67	**9.70**
Dinh et al. [15]	5.69	5.63	6.47	6.88	8.22	8.73	8.37	6.84	**7.10**
Our proposed method	5.54	5.72	5.32	5.64	8.02	8.27	6.40	5.21	**6.26**

Bold values indicate the average error values

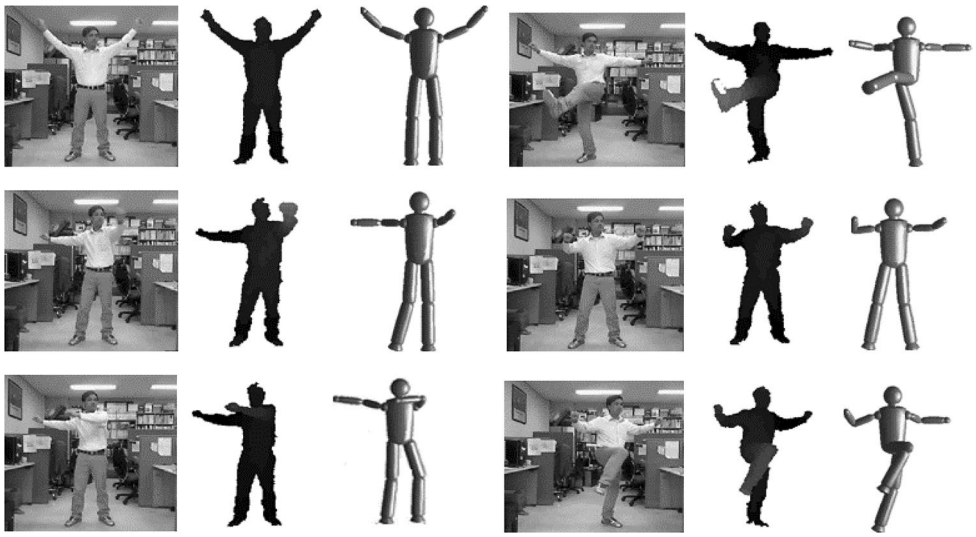


Fig. 10 Sample results of the proposed 3-D human pose recovery method from six depth silhouettes with crossing of arms or legs

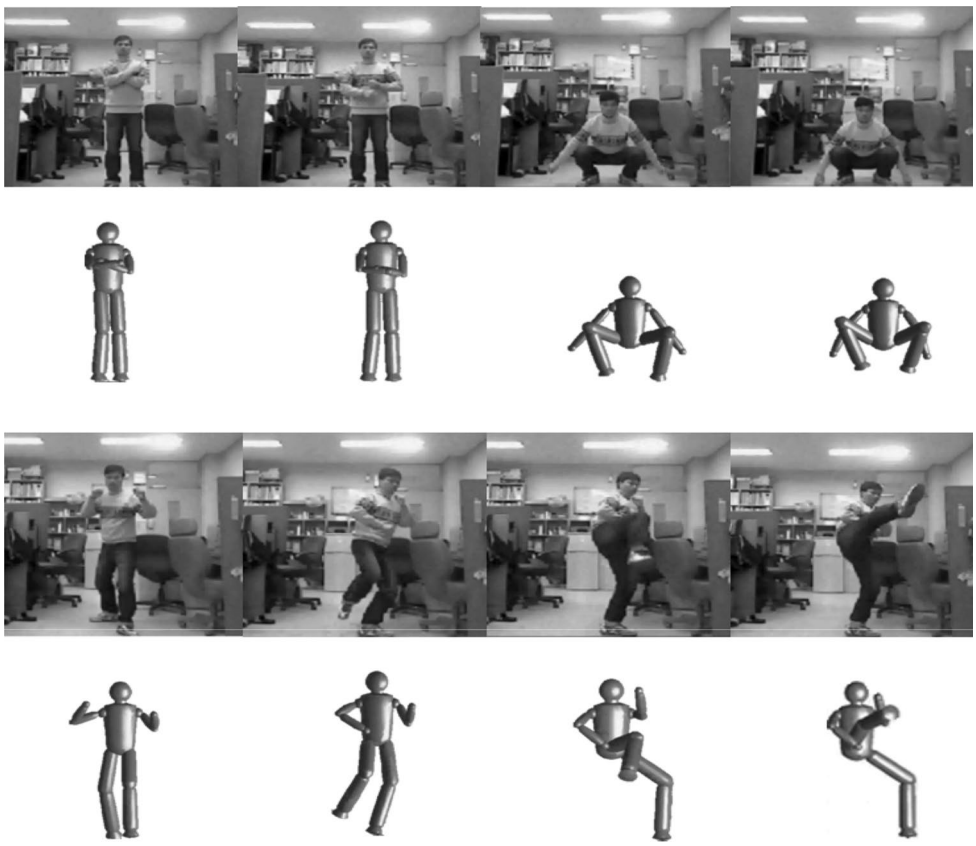


Fig. 11 Sample results of the proposed 3-D human pose recovery method from eight movements of arms and legs with self-collisions and occlusion

and our method using the same real data with results of the recovered 3-D human poses represented on the same 3-D human model. Figure 12 shows, second row, typical results of four recovered 3-D human poses in a crossing and self-occlusion movement sequence of arm or leg body parts using PDA; third row, as above with the proposed method. The first through fourth columns show the crossing movements of two legs and arms. By visual inspection of the results of the recovered 3-D human pose and the corresponding RGB images, the recovered 3-D human poses of our proposed method were more robust and accurate than the PDA method as shown in Fig. 12.

Fig. 12 Comparison of the proposed method vs. a PDA method [15] for four poses with occlusion of arms and legs. (*First row*) RGB images, (*second row*) results from PDA, and (*third row*) the results from the proposed method

5 Conclusion

We present a new approach to recover 3-D human poses based on human body parts tracked via nonrigid point set registration from a single depth silhouette. Our approach focused on solving 3-D human complex pose recovery. This is a challenging problem in 3-D human pose recovery, particularly with leg or arm crossing and self-occluding human poses. Quantitative assessments of our proposed method using synthetic data showed an average reconstruction error of 6.89° for the four main joint angles. Qualitative assessments using real data showed that our system performed reliably for complex pose sequences containing crossing and occluding movements of body parts. These results showed that our proposed method recovered complex human poses. In addition, our experimental results showed that CPD was an effective method for nonrigid point set registration to track human poses and body part labels. This algorithm allowed our system to find point correspondences between two point sets to ensure global optimality of the solution and preserve a human pose structure when optimal transformations were used for sets of 3-D points.

Acknowledgments This research was supported by the MSIP (Ministry of Science, ICT and 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)). This work was also supported by the Industrial Core Technology Development Program (10049079, Development of Mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

References

1. Poppe, R.: Vision-based human motion analysis: an overview. Comput. Vis. Image Underst. **108**(1), 4–18 (2007)
2. Uddin, M.Z., Thang, N.D., Kim, T.S., Kim, J.T.: Human activity recognition using body joint angle features and hidden markov model. ETRI J. **33**(4), 569–579 (2011)
3. Jalal, M., Uddin, M.Z., Kim, T.S.: Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. IEEE Trans. Consum. Electron. **58**(3), 863–871 (2012)
4. Yang, M.T., Chuang, M.W.: Fall risk assessment and early-warning for toddler behaviors at home. Sensors **13**(12), 16985–17005 (2013)
5. Chen, L., Wei, H., Ferryman, J.: A survey of human motion analysis using depth imagery. Pattern Recog. Lett. ISSN 0167–8655 (2013)
6. Kalogerakis, E., Hertzmann, A., Singh, K.: Learning 3D mesh segmentation and labeling. ACM Trans. Graph. (TOG) **29**(4), 102–114 (2010)
7. Plagemann, C., Ganapathi, V., Koller, D., Thrun, S.: Real-time identification and localization of body parts from depth images. In: proceedings of IEEE International Conference on Robotics and Automation (ICRA), pp. 3108–3113 (2010)
8. Schwarz, L.A., Mkhitaryan, A., Mateus, D., Navab, N.: Human skeleton tracking from depth data using geodesic distances and optical flow. Image Vis. Comput. **30**(3), 217–226 (2012)
9. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Moore, R.: Real-time human pose recognition in parts from single depth images. Commun. ACM **56**(1), 116–124 (2013)

10. Taylor, J., Shotton, J., Sharp, T., Fitzgibbon, A.: The Vitruvian manifold: inferring dense correspondences for one-shot human pose estimation. In: Proc. of IEEE Conference Computer Vision and Pattern Recognition (CVPR), pp. 103–110 (2012)
11. Kim, D., Kim, D.: A fast ICP algorithm for 3-D human body motion tracking. Signal Process. Lett. IEEE **17**(4), 402–405 (2010)
12. Mundermann, L., Corazza, S., Andriacchi, T.P.: Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated models. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–6 (2007)
13. Corazza, S., Mündermann, L., Gambaretto, E., Ferrigno, G., Andriacchi, T.P.: Markerless motion capture through visual hull, articulated ICP and subject specific model generation. Int. J. Comput. Vision **87**(1–2), 156–169 (2010)
14. Myronenko, A., Song, X.: Point set registration: coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2262–2275 (2010)
15. Dinh, D.L., Lim, M.J., Thang, N.D., Lee, S., Kim, T.S.: Real-time 3-D human pose recovery from a single depth image using principal direction analysis. Appl. Intell. **41**(2), 473–486 (2014)
16. Tam, G.K., Cheng, Z.Q., Lai, Y.K., Langbein, F.C., Liu, Y., Marshall, D., Rosin, P.L.: Registration of 3-D point clouds and meshes. A survey from rigid to nonrigid. IEEE Trans. Vis. Comput. Graph. **19**(7), 1199–1217 (2013)
17. Yuille, A.L., Grzywacz, N.M.: A mathematical analysis of the motion coherence theory. Int. J. Comput. Vis. **3**(2), 155–175 (1989)
18. Jian, B., Vemuri, B.C.: A robust algorithm for point set registration using mixture of Gaussians. IEEE Int. Conf Comput. Vis. **2**(20), 1246–1251 (2005)
19. Jian, B., Vemuri, B.C.: Robust point set registration using gaussian mixture models. IEEE Trans. Pattern Anal. Mach. Intell. **33**(8), 1633–1645 (2011)
20. Sang, Q., Zhang, J., Yu, Z.: Non-rigid point set registration: a bidirectional approach. ICASSP **2012**, 693–696 (2012)
21. Droeschel, D., Behnke, S.: 3-D body pose estimation using an adaptive person model for articulated ICP. In: Intelligent Robotics and Applications, pp. 157–167 (2011)
22. Siddiqui, M., Medioni, G.: Human pose estimation from a single view point, real-time range sensor. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1–8 (2010)
23. Brox, T., Rosenhahn, B., Gall, J., Cremers, D.: Combined region and motion-based 3-D tracking of rigid and articulated objects. IEEE Trans. Pattern Anal. Mach. Intell. **32**(3), 402–415 (2010)
24. Wei, X., Zhang, P., Chai, J.: Accurate realtime full-body motion capture using a single depth camera. ACM Transact. Graph. (TOG) **31**(6), 188 (2012)
25. Ye, M., Wang, X., Yang, R., Ren, L., Pollefeys, M.: Accurate 3D pose estimation from a single depth image. 2011 IEEE International Conference on Computer Vision (ICCV), pp. 731–738 (2011)
26. Baak, A., Müller, M., Bharaj, G., Seidel, H.P., Theobalt, C.: A data-driven approach for real-time full body pose reconstruction from a depth camera. In: Consumer Depth Cameras for Computer Vision, pp. 71–98 (2013)
27. Ganapathi, V., Plagemann, C., Koller, D., Thrun, S. Real time motion capture using a single time-of-flight camera. In: CVPR, pp. 3108–3113 (2010)
28. Ganapathi, V., Plagemann, C., Koller, D., Thrun, S.: Real-time human pose tracking from range data. In: ECCV, pp. 738–751 (2012)
29. Helten, T., Baak, A., Bharaj, G., Muller, M., Seidel, H.P., Theobalt, C.: Personalization and evaluation of a real-time depth-based full body tracker. In: 3DV, pp. 279–286 (2013)
30. Ye, M., Yang, R.: Real-time simultaneous pose and shape estimation for articulated objects using a single depth camera. In: CVPR, pp. 2353–2360 (2014)
31. Fossati, A., Dimitrijevic, M., Lepetit, V., Fua, P.: From canonical poses to 3-D motion capture using a single camera. Pattern Anal. Mach. Intell. IEEE Transact. **32**(7), 1165–1181 (2010)
32. Lee, M.W., Cohen, I.: A model-based approach for estimating human 3D poses in static images. Pattern Anal. Mach. Intell. IEEE Transact. **28**(6), 905–916 (2006)
33. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**(1), 269–271 (1959)
34. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. PAMI **24**(5), 1–5 (2002)
35. CMU motion capture database. http://mocap.cs.cmu.edu
36. Autodesk 3ds Max, 2012