Human Skeleton Extraction of Depth Images Using the Polygon Evolution

Huan Du[1], Jian Wang[1,2,*], Xue-xia Zhong[3,1], Ying He[1] and Lin Mei[1]

[1] Cyber Physical System R&D Center,
The Third Research Institute of Ministry of Public Security, Shanghai 201204, P. R. China
huan_du@163.com, wjconan@ieee.org x, 489331003@qq.com, 13524514531@126.com

[2] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, P. R. China
wjconan@ieee.org

[3] School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China
zhongxuexia2013@163.com

Abstract. This paper proposes a novel skeleton extraction approach in the depth image based on the polygon evolution. The external contour of person is firstly extracted from the depth image and evolved to a external polygon using a polygon evolution method. Subsequently, the depth histogram is used to extract internal self-occlusion body parts, and contours of these parts are evolved to internal polygons. In external and internal polygons, skeleton points are extracted under different criterias respectively. Finally, all skeleton points are linked to a complete skeleton. Experimental results on a variety of postures demonstrate the robustness and reasonability of our skeleton extraction approach.

Keywords: skeleton extraction, depth image, polygon evolution, external polygon, internal polygon.

1 Introduction

As a natural means of activity allowing complex information to be conveyed, human motion has been one of the most popular research hotspots, and its applications have cover a variety of fields, such as human-machine interaction, security surveillance, content-based retrieval, sports training, virtual reality, etc. In order to analyze motion in an image serial or a video, human motion usually need to be separated into combination of a series of movements of body parts.

For tracking human full-body pose in real-time, a person must wear cumbersome markers or special suits with a large number of sensors in order to provide position signal of skeletons and joints to camera-based motion capture systems. In a past decade, unmarked human pose estimation methods[1, 2] and improved approaches

* Corresponding author.

V. G. Duffy (Ed.): DHM 2014, LNCS 8529, pp. 15–23, 2014.

with multiple cameras[3, 4] have attracted more attention as a research focus in computer vision. However, tracking complex human movements under a general environment is still a great challenge due to the sensitivity of the image to illumination variation and body occlusions.

Having a substantial immunity to lighting conditions and variations in visual appearance, novel depth cameras develop rapidly based on recent technological advances in very recent years. Such camera allows acquiring dense scans of a scene for constructing depth images, which gives a easy way to obtain three-dimensional model in real-time. Regarded as a prominent representative of depth camera, Microsoft Kinect incorporates several advanced sensing hardware, i.e. a depth sensor, a color camera, and a four-microphone array, for providing various perception capabilities on full-body 3D motion capture, facial recognition, and voice recognition[5]. With help of the depth camera, many researchers have proposed different algorithms to address pose estimation and human motion capture from depth images[6]. For a given body part, e.g. head, its six degrees of freedom (DOF) of motion could be recover from a sequence of depth images[7]. Combining local optimization with global retrieval techniques, a data-driven hybrid strategy speeds up pose estimation procedure for real-time tracking full-body motions[8].

In order to obtain a well-connected skeleton, a connectivity criterion was proposed to generates a connected Euclidean skeleton[9]. Based on a set of point pairs along the object boundary, basic idea of the criterion is to determine whether a given pixel is a skeleton point independently. Based on a graph-based representation of the depth data used to measure geodesic distances between body parts that are invariant to body movement, a skeleton body model can be fitted by detecting anatomical landmarks in the 3D data and fitting for obtaining human full-body pose estimation[10]. By executing graph contraction and surface clustering iteratively under given constraints, a curve skeleton of a 3D shape may be extracted with the correct topological structure[11]. Without a large number of marked-based motion capture data for training, human skeletons are extracted from depth images based on the symmetry of skeletons to object boundaries which are identified with different types[12]. In order to gain the skeleton with a simplest possible structure that provides a best possible reconstruction of a given shape, skeleton pruning as a trade-off between skeleton simplicity and shape reconstruction error is cast[13]. Another skeleton pruning method uses contour partitioning to obtain skeletons which do not have spurious branches[14]. Based on a learned boundary edge function, the kinematic skeleton is extracted by computing a set of motion boundaries which correspond to all possible articulations of the 3D object[15].

As the essential for general shape representation, a skeletonization algorithm must be able to extracted accurate skeleton, be robust to noise, occlusion, position translation and rotation transformation. For precise motion analysis, a connected skeleton has to preserve topological and hierarchical properties of human body[9]. Unfortunately, most state-of-the-art methods cannot overcome these problem with low computational complexity.

Using polygon evolution, this paper proposes a novel skeleton extraction method to extract external skeleton and internal self-occlusion skeleton. The criterion of skeleton

extraction in polygons can remove redundant branches in skeleton model, and preserve original anatomical topology. The remain of this paper is organized as follows. Section 2 describes the proposed skeleton extraction approach in detail. Experimental results are presented in Section 3, and conclusions are given in Section 4.

2 The Proposed Approach

The procedure of the proposed approach is shown in Fig. 1. It contains four stages: (1) external polygon generation, (2) internal polygon generation, (3) skeleton extraction, and (4) skeleton linking. In the following sections, we introduce these stages respectively.

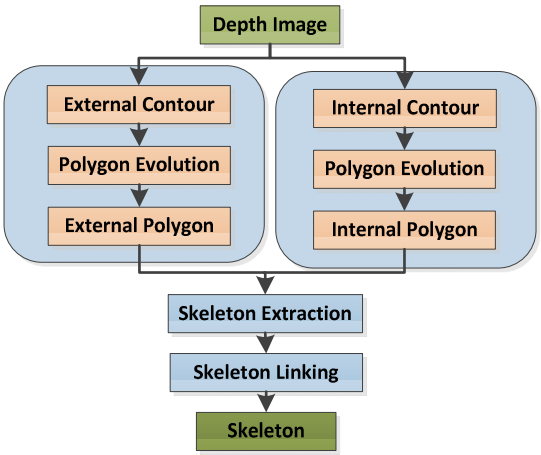

Fig. 1. The procedure of the proposed approach

2.1 External Polygon Generation

In the depth image, each player is randomly assigned an index number by Kinect[16]. We use this index number to obtain the mask depth image I_d of each player (see Fig. 2 (b)) and extract the contour of this mask as the external contour of the player C_e (see Fig. 2 (c)). Due to the external contour C_e is closed, we evolve it to a external polygon P_e (see Fig. 2 (d)) using the Douglas-Peucker(DP) algorithm in[17].

Specifically, partial contours of player's limbs sometimes may be inside the mask. In order to ensure the accuracy of extracted skeleton, we should add these contours to the external polygon. Firstly, we use Canny edge detector[18] to extract edges in the mask depth image. Then using the DP algorithm to evolve these edges. We reserve edges which are evolved to approximate vertical lines and add these lines L_e to the external polygon P_e. The final external polygon P_e is shown in Fig. 2 (f).

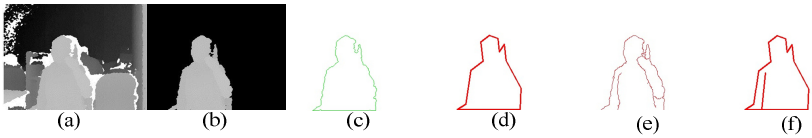

Fig. 2. The procedure of extracting the external polygon. (a) the original depth image; (b) the mask depth image; (c) the external contour; (d) the polygon of the external contour; (e) the Canny edge detector; (f) the external polygon.

2.2 Internal Polygon Generation

Because of the complicacy of human postures, there may appear self-occlusion situation that player's arms are inside the mask. In this case, we can not extract skeleton in these parts just depending on the external polygon. To solve this problem, we need extract these internal contours separately.

Depth data represent distances from the camera to the nearest object, we use Pyramid segmentation algorithm[19] to divide the depth image I_d into uniform depth blocks. Then statistic depth values of blocks and establish the depth histogram H_d for the depth image (see Fig. 3 (b)). Generally, in the depth histogram H_d, the value of the background $H_d(0)$ is maximum and the value of the torsel $H_d(K)$ is the second maximum. Self-occlusion parts are in the front of the torsel. According to these priori knowledge, we define self-occlusion body parts are:

$$p \in \aleph,\ if\ \ Depth\left(p\right) > K \tag{1}$$

where $\aleph$ denotes self-occlusion body parts, K is the depth value of the torsel in the depth histogram, p denotes one pixel in the depth image. The extracted self-occlusion body part is shown in Fig. 3 (c).

Similarity as the external contours, we extract the contour of the self-occlusion body part (see Fig. 3 (d)) and evolve it to an internal polygon P_i using the DP algorithm. The internal polygon P_i is shown in Fig. 3 (e).

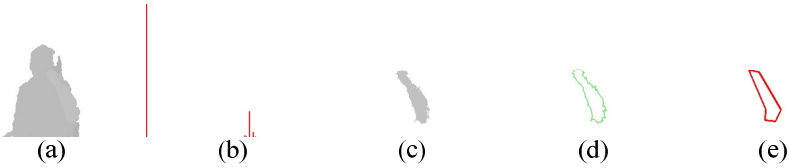

Fig. 3. The procedure of extracting the internal polygon. (a) the segmented depth image; (b) the depth histogram; (c) the self-occlusion body part; (d) the internal contour; (e) the internal polygon.

2.3 Skeleton Extraction

The central axis of contours is usually known as skeleton. To simplify the extraction process, we extract skeleton in evolutive polygons rather than in contours.

To guarantee the property that the skeleton is symmetrical to the polygon, we use the criterion proposed in[12] to compute the skeleton:

$$\begin{cases} D^2(q_1, p) - D^2(q_2, p) \le \max\left(\|x_1 - x_2\|, \|y_1 - y_2\|\right) \\ D(p, q_0) \le D(q_1, q_2) \end{cases} \tag{2}$$

where p is a given point inside the polygon, q_1 and q_2 are two closest edge points to p which are on two different edges of the polygon (q_1 and q_2 are called generating points of p in the rest of paper), q_0 is the midpoint of line $\overline{q_1 q_2}$, $D(\bullet)$ denotes the Euclidean distance, (x_1, y_1) and (x_2, y_2) are the coordinates of q_1 and q_2 respectively.

Usually, if p satisfies the Eq.2, we regard p as a skeleton point. However, this criterion just can ensure the symmetry of the skeleton with respect to polygons, and can not suppress noise and remove spurious skeleton branches. In this case, we increase constraint strategies to external and internal polygons respectively.

In the external polygon P_e, if two generating points of p are all on edges of the polygon, the point which satisfies both Eq.2 and Eq.3 is a skeleton point:

$$\begin{cases} D\left(\overline{q_1}, \overline{q_2}\right) > T_1 \\ L_1 \bigcap L_2 = \varnothing \end{cases}, \; if \;\; q_1 \in P_e, q_2 \in P_e \tag{3}$$

where L_1 and L_2 are edges which q_1 and q_2 are on respectively. If q_1 is a vertex of the external polygon P_e, $\overline{q_1}$ denotes the vertex q_1. If q_1 is not a vertex of the external polygon P_e, $\overline{q_1}$ denotes the edge L_1. $\overline{q_2}$ is similar to $\overline{q_1}$. T_1 is a parameter to suppress spurious skeleton points. Under the constraint of Eq.3, excess skeleton points which may be generated by adjacent edges can be suppressed in the external polygon.

If a generating point of p is on the line L_e, the point which satisfies both Eq.2 and Eq.4 is a skeleton point:

$$p \in Square\,[L_1, L_2], \; if \;\; q_1 \in L_e \parallel q_2 \in L_e \tag{4}$$

where $Square[L_1, L_2]$ denotes a convex quadrilateral with edges L_1 and L_2.

The illustration of extracted skeleton points in the external polygon is shown in Fig. 4 (a).

In the internal polygon P_i, the point which satisfies both Eq.2 and Eq.5 is a skeleton point:

$$\begin{cases} D\left(\overline{q_1}, \overline{q_2}\right) > T_2 \\ \alpha < angle\left(L_1, L_2\right) < \beta, \quad if \;\; L_1 \bigcap L_2 \neq \varnothing \end{cases} \tag{5}$$

where $angle(\bullet)$ denotes the angle between two lines. T_2, α and β are parameters to suppress spurious skeleton points. Under the constraint of Eq.5, we can guarantee generating points of selected skeleton points are at the suitable distance, not too close nor too far.

The illustration of extracted skeleton points in the internal polygon is shown in Fig. 4 (b).

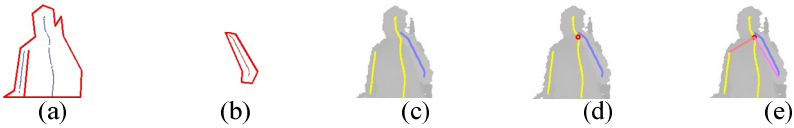

Fig. 4. The illustration of skeleton extraction and linking. (a) the external skeleton; (b) the internal skeleton; (c) external and internal skeletons in the depth image; (d) the shoulder center in the depth image; (e) the complete skeleton.

2.4 Skeleton Linking

In skeleton points, some points are close to each other. We firstly perform point linking to group them into skeleton lines. As shown in Fig. 4 (c), yellow lines are external skeletons generating by the external polygon, blue lines are internal skeletons generating by the internal polygon.

To generate a complete skeleton, we need to connect upper limbs and the torsel. The shoulder center $P_{s\text{-}c}$ (see the red point in Fig.4 (d)) is the most accurate point in the skeleton tracing by Kinect. According to the relative position with $P_{s\text{-}c}$, we judge each skeleton line is left upper limb, right upper limb or their overlap and link them with the shoulder center $P_{s\text{-}c}$. The complete skeleton is shown in Fig. 4 (e).

3 Experiments and Comparison

We implement our approach using C++. It takes about 1.8 seconds to process one single frame without code optimization on regular desktop with AMD core 4 3.8GHz CPU and 4GB RAM.

In order to evaluate the skeleton extraction performance of our approach, we extract skeletons from depth images of humans with different postures. Specially, we divide postures into three categories of face-on standing postures, lateral standing postures and sitting postures. And then compare our approach with Kinect[5] and Shen's approach[13].The implementation code of Shen's approach can be downloaded from http://wei-shen.weebly.com/publications.html.

Fig. 5 and Fig. 6 show skeletons of face-on and lateral standing postures respectively. Fig. 7 shows skeletons of sitting postures. As a skeleton pruning method, Shen's approach can obtain reasonable skeletons based on external contours for face-on standing postures as shown in Fig. 5 (c). However, for lateral standing postures and sitting postures, there are large structural changes in skeletons due to the particularity of postures in Fig. 6 (c) and Fig. 7 (c). In general, Shen's approach can generate simplified skeleton based on external contours but can not gain self-occlusion internal skeleton. Generally, for outsize body movement, Kinect's skeleton tracking can

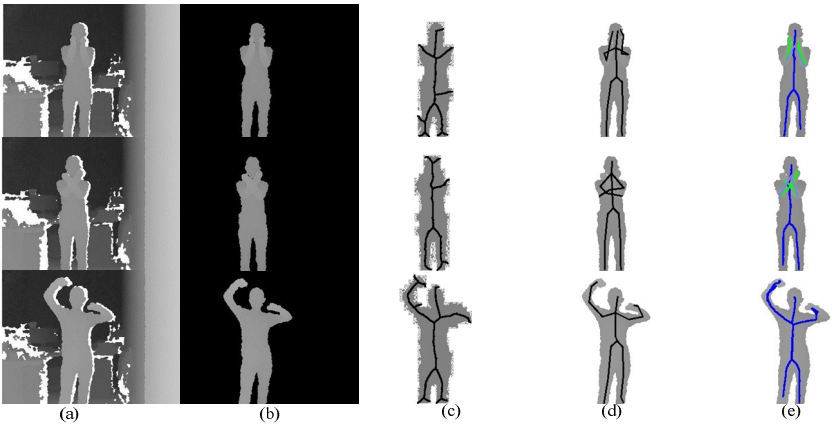

Fig. 5. Face-on standing postures. (a) original depth images; (b) mask depth images; (c) Shen's approach ; (d) Kinect's skeleton; (e) our approach.

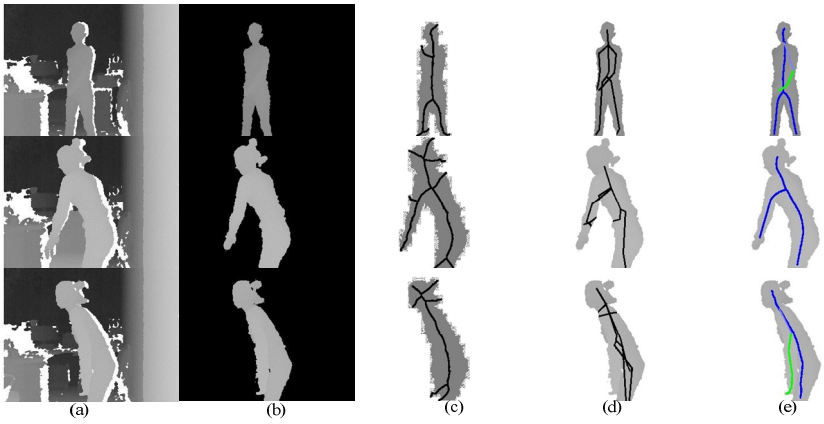

Fig. 6. Lateral standing postures. (a) original depth images; (b) mask depth images; (c) Shen's approach ; (d) Kinect's skeleton; (e) our approach.

obtain accurate skeleton results such as the 1st and 3rd in Fig. 5 (d). In the case of occlusion, Kinect sometimes estimates accurate skeleton positions of occluded body parts such as 1st in Fig. 6 (d). But in most cases, estimated skeleton results are not particularly accurate (see the 2nd in Fig. 5 (d), the 2nd and 3rd in Fig. 6 (d)). For sitting postures, results of Kinect's skeleton tracking are inaccurate due to there is no sit mode (see Fig. 7 (d)). Compared with Shen's approach and Kinect, our approach not merely can obtain simplicity external skeleton without excess spurious branches but also can gain reasonable internal skeleton no matter in face-on standing postures (see Fig. 5 (e)), lateral standing postures (see Fig. 6 (e)) or sitting postures (see Fig. 7 (e)).

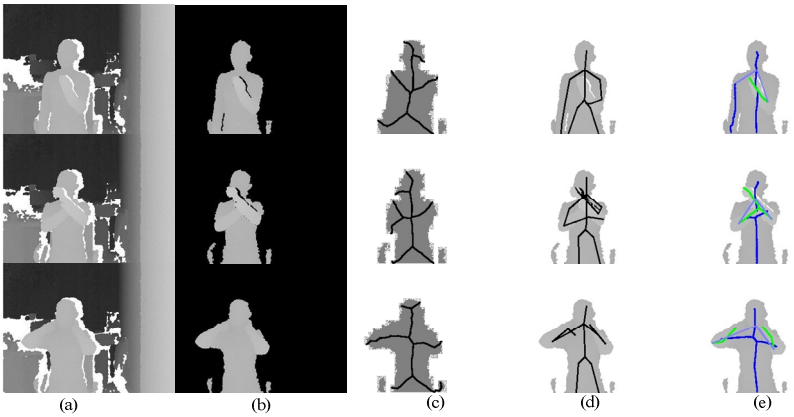

Fig. 7. Sitting postures. (a) original depth images; (b) mask depth images; (c) Shen's approach ; (d) Kinect's skeleton; (e) our approach.

4 Conclusions

In this paper, we present a novel human skeleton extraction approach in the depth image based on the polygon evolution. A external polygon is evolved from the external contour which is extracted form the depth image. Then internal polygons are also evolved by internal self-occlusion contours which are extracted through the depth histogram. Besides, skeleton points are extracted under different criterias in external and internal polygons respectively. Experimental results on a variety of postures demonstrate the robustness and reasonability of our skeleton extraction approach. In our future work, we will try to extract skeleton from depth images of human with different appendants like packsacks, hand bags and so on.

Acknowledgement. Our research was sponsored by following projects:

- National High-tech R&D Program of China ("863 Program") (No. 2013AA014603);
- National Science and Technology Support Projects of China (No. 2012BAH07B01);
- Program of Science and Technology Commission of Shanghai Municipality (No. 12510701900, No. 13ZR1410400, No. 12DZ0512100);
- 2012 IoT Program of Ministry of Industry and Information Technology of China;
- Program of Third Research Institute of the Ministry of Public Security (No.C13348).

References

1. Urtasun, R., Darrell, T.: Sparse probabilistic regression for activity-independent human pose inference. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2008)
2. Jaeggli, T., Koller-Meier, E., Gool, L.V.: Learning generative models for multi-activity body pose estimation. International Journal of Computer Vision 83(2), 121–134 (2009)
3. Kehl, R., Gool, L.: Markerless tracking of complex human motions from multiple views. Computer Vision and Image Understanding 104(2), 190–209 (2006)
4. Bandouch, J., Engstler, F., Beetz, M.: Accurate human motion capture using an ergonomics-based anthropometric human model. In: Proceedings of V Conference on Articulated Motion and Deformable Objects Andratx, Mallorca, Spain (2008)
5. Zhang, Z.: Microsoft Kinect sensor and its effect. IEEE Multimedia 19(2), 4–10 (2012)
6. Kohli, P., Shotton, J.: Key developments in human pose estimation for Kinect. In: Consumer Depth Cameras for Computer Vision, pp. 63–70. Springer, London (2013)
7. Kondori, F.A., Yousefi, S., Li, H., et al.: 3D head pose estimation using the Kinect. In: Proceedings of International Conference on Wireless Communications and Signal Processing, Nanjing, China, pp. 1–4 (2011)
8. Baak, A., Müller, M., Bharaj, G., et al.: A data-driven approach for real-time full body pose reconstruction from a depth camera. In: Proceedings of IEEE International Conference on Computer Vision, Barcelona, Spain, pp. 1092–1099. IEEE (2011)
9. Choi, W.-P., Lam, K.-M., Siu, W.-C.: Extraction of the Euclidean skeleton based on a connectivity criterion. Pattern Recognition 36, 721–729 (2003)
10. Schwarz, L.A., Mkhitaryan, A., Mateus, D., et al.: Human skeleton tracking from depth data using geodesic distances and optical flow. Image and Vision Computing 30, 217–226 (2012)
11. Jiang, W., Xu, K., Cheng, Z.-Q., et al.: Curve skeleton extraction by coupled graph contraction and surface clustering. Graphical Models 75(3), 137–148 (2013)
12. Shen, W., Xiao, S., Jiang, N., et al.: Unsupervised human skeleton extraction from Kinect depth images. In: Proceedings of the 4th International Conference on Internet Multimedia Computing and Service, Wuhan, China, pp. 66–69 (2012)
13. Shen, W., Bai, X., Yang, X., et al.: Skeleton pruning as trade-off between skeleton simplicity and reconstruction error. Science China Information Sciences 56(4), 1–14 (2013)
14. Bai, X., Latecki, L.J., Liu, W.Y.: Skeleton Pruning by Contour Partitioning with Discrete Curve Evolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(3), 449–462 (2007)
15. Benhabiles, H., Lavoue, G., Vandeborre, J., et al.: Kinematic skeleton extraction based on motion boundaries for 3D dynamic meshes. In: Proceedings of Eurographics Workshop on 3D Object Retrieval, Cagliari, Italy, pp. 71–76 (2012)
16. Leyvand, T., Meekhof, C., Wei, Y.-C., et al.: Kinect Identity: Technology and Experience. Computer 44(4), 94–96 (2011)
17. Douglas, D.H., Peucker, T.K.: Algorithms for the reduction of the number of points required to represent a line or its caricature. The Canadian Cartographer 10(2), 112–122 (1973)
18. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(6), 679–698 (1986)
19. Antonisses, H.J.: Image segmentation in pyramids. Computer Graphics and Image Processing 19, 367–383 (1982)