2010 Second International Workshop on Education Technology and Computer Science

Extraction of Human Body Skeleton Based on Silhouette Images

Jianhao Ding[1,2], Yigang Wang[2], Lingyun Yu[3]

[1]State Key Laboratory of CAD & CG, Zhejiang University, Hangzhou, China

[2]Institute of Computer graphics and Image Processing, Hangzhou Dianzi University, Hangzhou, China

[3]Department of Mathematics and Computing Science, University of Groningen, The Netherlands

***Abstract*—Skeleton extraction is essential for general shape representation. A typical skeletonization algorithm should obtain the ability to preserve original object's topological and hierarchical properties. However, most of current methods are high memory cost, computationally intensive, and also require complex data structures. In this paper, we propose an efficient and accurate skeletonization method for the skeleton feature points extracted from human body based on silhouette images. First, the gradient of distance transform is used to detect critical points inside the foreground. Then, we converge and simplify critical points in order to generate the most important and elegant skeleton feature points. Finally, we present an algorithm which connects the skeleton feature points and estimates the position of skeleton joints.**

Keywords- Feature Detection; Skeletonization; Joints estimation;

I. INTRODUCTION

Human motion analysis is notoriously difficult because human bodies are highly articulated and people tend to wear complex textures-clothes which could confuse the important features needed to distinguish poses. Human motion analysis concerns with some key techniques including action recognition[1-3] and motion tracking[4-6]. The feature extraction of human body from image plays an important role in performing some specific tasks. The skeleton extraction is essential for general shape representation and will affect system performance and algorithms' complexity. Especially, the features concerning with both position and motion of joints can help us to determine human body's pose and motion. The wide range of applications shows the comprehensive usefulness of reducing patterns to skeleton-joints representations, which can be attributed to the need of processing a reduced amount of data. The necessary properties of a skeletonization algorithm need to be accurate and robust to noise and can generate a connected skeleton in order to preserve its topological and hierarchical properties. However, most of the methods are computationally intensive and require a complex data structure[7-9].

The structure of this paper is given as below. Previous work in the area of extracting skeleton is introduced in section II. In section III and IV we describe two essential works including skeleton points detection and skeletonization respectively, which are the main contributions in our paper. Experimental results and discussions are given in section V. At last we draw our conclusions.

II. RELATED WORK

Many approaches of skeletonization have been proposed throughout the past decades. Most of skeletonization algorithms can be simply classified into two essential types.

One type is referred to as thinning algorithms. The thinning methods[10,11] used iterative algorithm to delete successive layers of pixels on the boundary of the pattern until only a skeleton remains, while keeping the topological structure of pixels. The deletion or retention of a pixel would depend on the configuration of pixels in a local neighborhood. As the skeleton generated by thinning can not ensure its accuracy and smoothness, it might also need further processing[12]. Tao Ju[13] computed skeletons of volumetric models by alternating thinning and a skeleton pruning routine.

Alternatively, the distance transformation is to convert a binary image into another image which provides each object pixel a value corresponding to the minimum distance from the background. Generally the algorithm based on distance transformation can obtain accurate skeleton construction but can't guarantee the connectivity of skeleton. Choi[14] proposed a skeletonization method based on a signed sequential Euclidean distance map. They had to compute a set of point pairs along the object boundary, which are the most closed contour points to the pixel. As a result, they can fast extract linear skeleton but can't guarantee the connectivity when the branch of object is very small. Ding[15] presented a novel method utilizing the distance transform, which the skeleton was obtained by growing with the restriction of one pixel width from the skeleton seed point, and the connectivity was ensured though growing processing. In order to extract skeleton, Melada[16] combined the thinning with Euclidean distance transformation. However it might result noisy branch and also destroyed the connectivity. Liu[17] proposed a fast algorithm to track the contour using Euclidean distance maps from outer to inner. But the skeleton joints were not estimated in their paper.

III. SKELETON POINTS DETECTION

The objective of this step is to obtain the position of the joints in the silhouette image finally, which is similar to Fan[18]

Sponsored by Planned Science and Technology Project of Zhejiang Province for contract 2008C24014.

works. But our algorithm is easier to get the skeleton and more robust for localizing skeleton joints. We utilize the gradient of distance transform to detect critical points from silhouette images. Additionlly, the joints position was not estimated in their paper.

A. *The gradient of distance transform*

According to the definition of skeleton and distance transform, the value of distance transform on skeleton points should be greater than their neighbors[19]. The extraction of skeleton is to find the ridge lines of objects. To obtain the ridge lines, we may need to explore the local features such as the gradient of distance transform image.

Suppose that the image distance transform is $DT(x, y)$, $DT(x_0, y_0)$ denotes the distance transform value in the pixel $p(x_0, y_0)$. The gradient of distance transform is denoted as $\nabla DT = (\frac{\partial DT}{\partial x}, \frac{\partial DT}{\partial y})$, so the norm of gradient vector is:

$$|\nabla DT| = \sqrt{(\frac{\partial DT}{\partial x})^2 + (\frac{\partial DT}{\partial y})^2}$$

Apparently ridge lines should locate in those smaller $|\nabla DT|$ points. However the number of points is too large to generate the linear skeleton directly.

B. *Extracting Local Maximum Points and Critical Points*

Definition 1 Local Maximum Points(LMP): $DT(p)$ is the distance transform at p, Q is the set of 8-connected neighbors of p. If $DT(p) \geq DT(q)$ for $\forall q \in Q$, we define the point p is a local maximum point of distance transform.

In terms of this definition, the LMP are the latent skeleton points. Fig. 1(a) shows the LMP denoted with yellow color. It is obvious that the LMP have equal distance transform values in some connected regions according to the definition 1 and the transitivity of connected relation.

Definition 2 Critical Points(CP): CP are the LMP whose $|\nabla DT|$ is minimum value point among its connected neighbors.

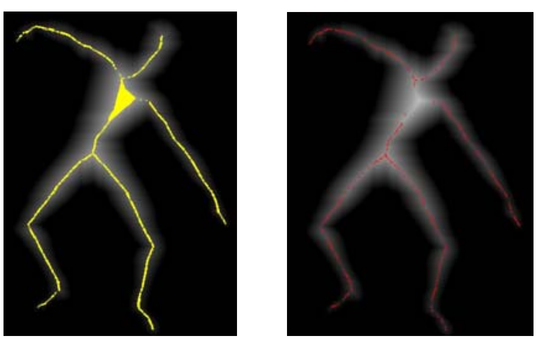


Figure 1. (a) Local maximum of DT (b) Critical points

We can easily obtain the critical points with one-pixel wide in the distance transform image in terms of this definition. Fig. 1(b) shows the position of critical points denoted in red color. Apparently, the critical points locate the position where these pixels have smaller $|\nabla DT|$. Therefore, we can suppose that the critical points are a subset of image skeleton. After connecting these critical points in some specific constraint condition, the algorithm can generate linear skeleton. However, comparing with skeleton joints, the number of critical points is too large. It will be very hard to connect them correctly when we face to so large set of scatter points. In this case, we would need to extract the feature points from the critical points further.

C. *Detecting Skeleton Feature Points*

Definition 3 Skeleton Feature Points(SFP): Considering a critical point as the center of a circle, we can determine the tangent circle C of silhouette edges for this point. We suppose the cps, which $cps \in C$, as meaning of the critical points set within the range of the circle C. Then we will just discard all points belong to cps and only remain the center point of C, which will be referred as the skeleton feature point.

According to the definition 3, we traverse the rest critical points with the same operation until no overlay regions of tangent circles for critical points left. The skeletonization process will benefit from different aspects offered by this simplification mechanism, such as reduced critical points.

IV. SKELETONIZATION

As the skeleton feature points we got in the section 3 are dispersive and irregularity, it is difficult to match or track these scatter points directly. Therefore, we need connecting the skeleton feature points and estimating joints.

A. *Connecting the skeleton feature points*

The objective of this step is to connect the feature points together in some specific routine.

We define a constraint condition to avoid any point to be incorporated when searching the successive feature points to be skeletonization. The constraint condition is very reasonable that any lines connecting two points must locate on the foreground. It means any pixel satisfied $DT(p) > 0$, where p is a pixel on lines.

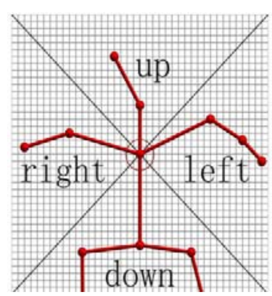


Figure 2. The definition of search direction from root point

We consider the current point as parent point and search the closest points satisfying the constraint condition as child points. Initially the point with maximum distance transform which generally located on body's chest is the root point. Its four children points are from four directions consisting of left direction, right direction, up direction and down direction respectively, as the fig. 2 shows. The search direction is determined by the slope information of the line connecting with the root point.

In each skeleton growing step, the algorithm will find the closest skeleton feature point which satisfied the constraint condition to the current point. If this point joins the skeleton, it will be deleted from the skeleton feature points set so as to avoid being searched repetitively. If no points being found any more, we refer the last point as an end point or leaf point. After reaching the leaf point, the algorithm will go back to the parent point and search from another child point. The algorithm repeats the process till the feature skeleton points list is empty to generate a tree of skeleton feature points finally.

B. *Estimating joints*

Here is a detailed description of the skeleton joints estimation. The skeleton of human body is a set of rigid parts which are joined by joints. The research of the anthropotomy[20] has indicated that, for different bodies, the proportions of body parts are approximately the same. For this reason, we can take a standard human skeleton model as an example, which is shown in fig. 3, and estimate the position of joints.

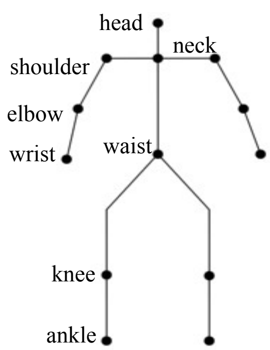


Figure 3. The standard human skeleton model

The skeleton feature points are tied together in a tree structure in the previous section. We traverse the tree of skeleton feature points to find the leaf points and the crossing points firstly, which have more than one child. From then on, the path distance from some given point to leaf point can be easily computed along to the connected skeleton feature points. The upper leaf point is considered as head joint. According to the shoulder direction and width proportion to the body height, we can obtain the shoulder joints easily from the root points. In the same way, we can determine the elbow joints from shoulder joints according to the length proportion of the upper arm to the total arm. Before computing two knee joints, we must determine the waist joint. Obviously the waist joint is the first point with two children in the down direction from the root point. Starting from the waist joint, the two knee joints are computed through the anthropotomy proportion. Finally we get all joints position of the standard dummy skeleton we defined and denote the joints position using yellow circles, as the fig. 4(e) shows.

V. RESULTS AND DISCUSSIONS

We have performed a number of experiments on single human body silhouette images with different styles and different poses. Fig. 4 shows our experiment results, including raw silhouette images, local maximum points, critical points, skeleton feature points and skeleton joints. Experiments show that the proposed method is satisfying. TABLE I lists the relevant number of the pixel points in the process of skeletonization, including local maximum points, critical points, skeleton feature points and skeleton joints, which images resolution are 640×480.

TABLE I. THE QUANTITY OF DIFFERENT FEATURES

	LMP	CP	SFP	SJ
image1	825	159	58	13
image2	735	116	39	13
image3	936	214	56	13

But the results are unsatisfied and even encounter some mistakes when the joints are hidden or kept out. That's because our approach uses the uniform scales of the human body parts to determine the position of the inner joints such as elbow and knee joints, there might cause some errors between the scales of the different bodies. In the subsequent work, we will solve hidden problem and estimate the 3D joints position[21,22] utilizing probabilistic model and multiple views methods.

Human motion is a complex pattern, this method must resolve the problems firstly, extracting the precision of the human silhouette from the complex background. Alternatively, how to determine the positions of the joints that are hidden or kept out is another formidable problem.

VI. CONCLUSIONS

In this paper, we have presented a method for articulated-pose identification from the silhouette images of human body. A dummy skeleton is extracted like the reality skeleton of the human from the silhouette images by using the gradient of distance transform, then the position of joints are determined by using the knowledge of the anthropotomy. Experiments show that the method can estimate the positions of joints based on the right human silhouette. The method has no restrictive presupposition compared with others.

REFERENCES

[1] Z. Zhang, Y. Hu, S. Chan, and L.T.Chia, "Motion context: A new representation for human action recognition", In ECCV, 2008(4), pp. 817-829.
[2] L. Wang, D. Suter, "Learning and matching of dynamic shape manifolds for human action recognition", IEEE Transactions on Image Processing, 2007, 16(6), pp. 1646-1661.
[3] Efros, A. Berg, G. Mori, and J. Malik, "Recognizing action at a distance", In ICCV, 2003, pp. 726-733.
[4] D. Vlasic, I. Baran, W.Matusik, and J. Popovi´c, "Articulated mesh animation from multi-view silhouettes", ACM Trans.Graph, 2008, 27(3), pp. 1-9.
[5] G. Cheung, S. Baker, and T. Kanade, "Shape-from-silhouette across time part ii: Applications to human modeling and markerless motion tracking", Int. J. of Computer Vision, 2005, 63(3), pp. 225-245.
[6] T. Drummond, R. Cipolla, "Real-time tracking of highly articulated

structures in the presence of noisy measurements", In Int. Conf. on Computer Vision, 2001, pp. 315-320.

[7] P. Dimitrov, C. Phillips, K. Siddiqi, "Robust and efficient skeletal graphs", In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hitton Head Island,USA: IEEE: 2000, pp.417-423.

[8] I. Bitter, A.E. Kaufman, M. Sato, "Penalized distance volumetric skeleton algorithm", IEEE Trans. on Visualization and Computer Graphics, 2001, 7(3), pp. 195-206.

[9] Y. Ge, J. M. Fitzpatrick, "On the generation of skeletons from discrete euclidean distance maps", IEEE Trans. on Pattern Analysis and Machine Intelligence, 1996, 18(11), pp. 1055-1066.

[10] C. Ma, S. Wan, "Parallel Thinning Algorithms on 3D (18, 6) Binary Images", Computer Vision and Image Understanding, 2000, 80(3), pp. 364-378.

[11] W. Xie, R.P. Thompson, R. Perucchio, "A topology-preserving parallel 3D thinning algorithm for extracting the curve skeleton", Pattern Recognition, 2003, 36(7), pp. 1529-1544.

[12] Che Wu-Jun, Yang Xun-Nian, Wang Guo-Zhao, "A dynamic approach to skeletonization", Journal of Software, 2003, 14(4), pp.818-823. (车武军, 杨勋年, 汪国昭, "动态骨架算法", 软件学报, 2003, 14(4), 818-823).

[13] T. Ju, M. Baker, W. Chiu, "Computing a family of skeletons of volumetric models for shape description", Computer-Aided Design, 2007, 39(5), pp. 352-360.

[14] W. Choi, K. Lam, W. Siu, "Extraction of the Euclidean skeleton based on a connectivity criterion. Pattern Recognition", 2003, 36(3), pp. 721-729.

[15] Ding Yi, Liu Wen-Yu, Zheng Yu-Hua, "Hierarchical connected skeletonization algorithm based on distance transform", Journal Infrared and Millimeter Waves, 2005, 24(4), pp. 281-285. （丁颐, 刘文予, 郑宇化, "基于距离变换的连通骨架算法", 红外与毫米波学报, 2005, 24(4), 281-285).

[16] Y. Mekada, J. Toriwaki, "Anchor point thinning using a skeleton based on the Euclidean distance transformation", Proceedings of the 16th International Conference on Pattern Recognition, 2002, pp.923-926.

[17] Liu Xiang-Bin, Zou Bei-Ji, Sun Jia-Guang, "Fast Euclidean Distance Transform Based on Contour Tracking", Chinese Journal of Computers, 2006, 29(2), pp. 317-323. (刘相滨, 邹北骥, 孙家广, "基于边界跟踪的快速欧氏距离变换算法", 计算机学报, 2006, 29(2), 317-323).

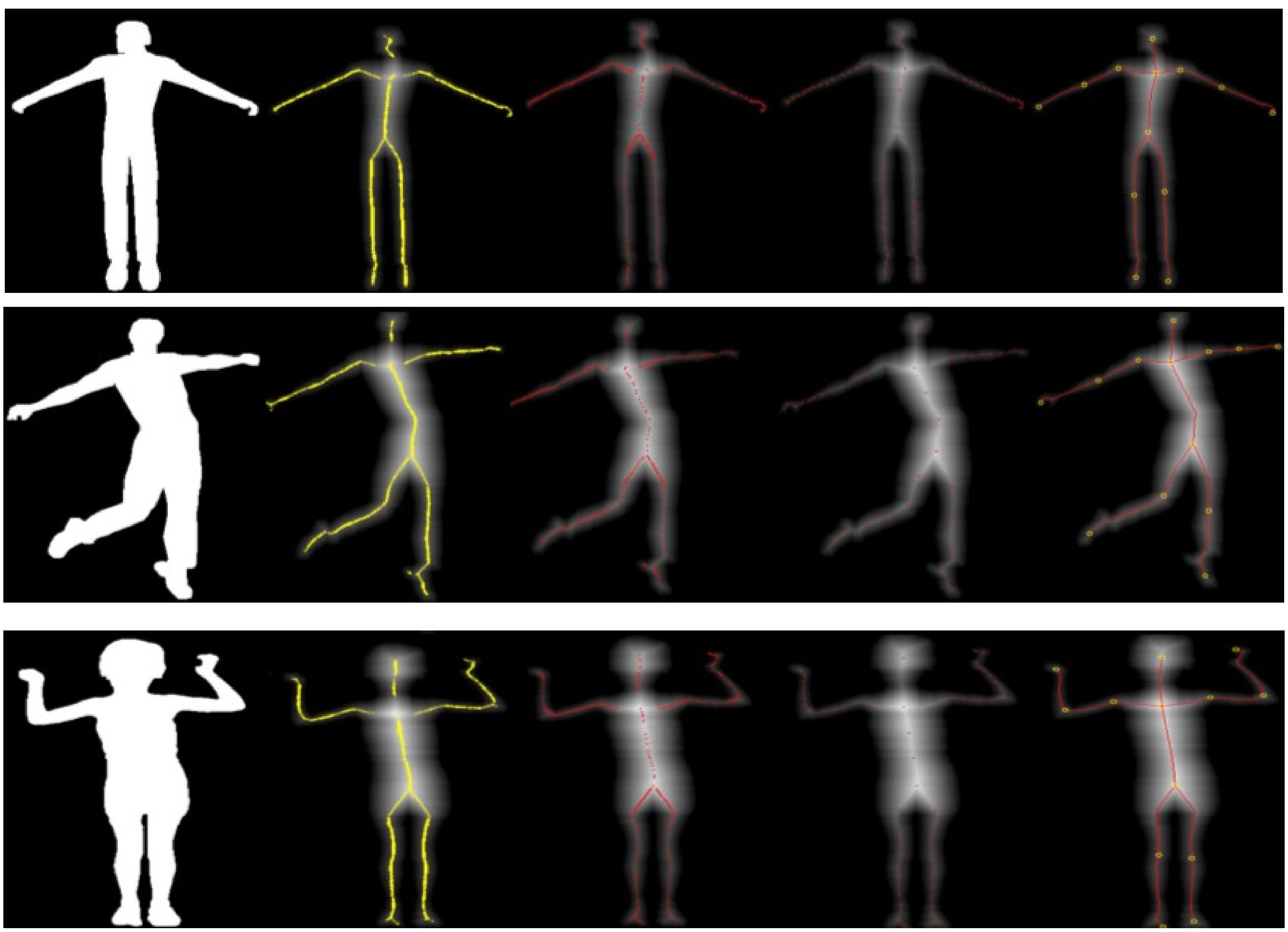
[18] Bin Fan, Zengfu Wang, "Pose Estimation Of Human Body Based On Silhouette Images", Proc. of ICIA'2004.

[19] Liu Jun-Tao, Liu Wen-Yu, Wu Cai-Hua,Yuan Liang, "A New Method of Extracting Objects' Curve-skeleton", Acta Automatica Sinica, 2008, 34(6), pp. 617-622. (刘俊涛, 刘文予, 吴彩华, 原亮, "一种提取物体线形骨架的新方法", 自动化学报, 34(6), 617-622).

[20] The compilation group of the anthropotomy, "Anthropotomy", Beijing Higher Education Publish House, 1997.

[21] A. Agarwal, B. Triggs, "Recovering 3D human pose from monocular images", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(1), pp. 44-58.

[22] F. Guo, G. Qian, "Learning and inference of 3D human poses from Gaussian mixture modeled silhouettes", In Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06), vol.2, pp. 43-47.



(a) raw silhouette (b) local maximum points (c) critical points (d) skeleton feature points (e) skeleton joints

Figure 4. Experimental results of human motion with different style and different pose