Automatic skeleton generation using hierarchical mesh segmentation

Ryan Anthony J. de Belen,* Rowel O. Atienza†
Electrical and Electronics Engineering Institute, University of the Philippines - Diliman

Abstract

In this paper, a pipeline is presented to automatically create high fidelity and interactive humanoid toys in virtual and augmented reality. First, the mesh is segmented into several parts using normal characteristic value (NCV) and global point signatures (GPS) as candidate points for segmentation. Then, joint locations are generated based on the segmentation results. Furthermore, the skinning weights are generated by solving the Laplace diffusion equation. Experimental results show that our pipeline is robust enough to extract skeletal structures from graphic artists' models as well as from scanned models. In addition, our pipeline is deformation-invariant as it can generate the same skeletal structure of a model having different poses. Finally, our system output can be directly used to setup skeleton-based animations as well as real-time virtual and augmented reality applications within minutes.

Keywords: rigging, skeleton extraction, character skinning, geometric processing, segmentation, virtual and augmented reality

Concepts: •**Computing methodologies** → **Animation;**

1 Introduction

In general, creating platforms where users can interact with 3D characters requires four stages: 3D model reconstruction, skeleton extraction, skinning, and animation. Character model reconstruction can be achieved in several ways: by artists using graphics software (e.g. Autodesk Maya), by hand-held laser scanners and depth sensors [Bogo et al. 2014], or photogrammetry [Egels and Kasser 2003]. To allow for user interaction, the 3D model must then be rigged and skinned to make a skeletal model [Igarashi et al. 2007]. Finally, transfer of motion capture data to a similar skeletal structure must be used to animate the 3D characters. The laborious process and the need for a fair proficiency with the software applications make simple character animation difficult than it could be.

Different pipelines have already been made that focus on skeleton extraction and character skinning. However, they may be difficult to implement, require large dataset, or produce varying skeletal structures which makes it difficult to set-up similar skeleton-based animations. In this paper, we present an automatic pipeline that robustly estimates skeletal structure of a humanoid 3D character and demonstrates a smooth skinning technique for skeletal-based animations given only the 3D mesh model. The ultimate goal is to create a system where people can input any 3D character model and produce highly interactive models in the digital world like in virtual or augmented reality (e.g. "Toys-to-Life").

*e-mail:rjdebelen@up.edu.ph
†e-mail:rowel@eee.upd.edu.ph

VR Meets PR 2016,, December 05-08, 2016, , Macao
ISBN: 978-1-4503-4548-4/16/12$15.00
DOI: http://dx.doi.org/10.1145/2992138.2992150

2 Automatic Skeleton Generation

Given a mesh information, a skeleton must be first embedded to it. In this section, we present an improvement of an existing technique and an introduction of a new approach in skeleton generation.

2.1 Normal Characteristic Value

Cheng et al. [2011] compute the normal characteristic values (NCV) of a mesh to find candidate segmentation points. The NCV computation is shown in Equation 1.

$$NCV(i) = \sum_{j \in N(i)} \frac{(1 - normal(v_i) \cdot normal(v_j))}{2m} \tag{1}$$

where $N(i)$ is the set of the 1-ring neighbor vertices of vertex i, m is the number of neighbors of vertex i, $normal()$ is the vertex's normal vector, and $\cdot$ is the dot product operation.

NCV computation is performed for all the mesh vertices and each vertex is a segment point candidate. The segment point candidates can be refined by choosing vertices with NCV larger than ϵ, the average of all NCV. Afterwards, k-means clustering is performed using the refined segment point candidates to find the candidate joint locations. The number of clusters was picked in accordance to the typical number of joints used in motion capture data. As a result, the system generates an anatomical skeletal structure that can be easily used in any animation software. Finally, the centroid of a cluster is considered to be a joint location (see Figures 1a and 1b).

2.2 Global Point Signatures

We use global point signature (GPS) embedding as a deformation-invariant surface representation where mesh segmentation is performed. Given a point p on the mesh surface, Rustamov [2007] defines GPS as the infinite-dimensional vector shown in Equation 2.

$$GPS(\mathbf{p}) = \left(\frac{1}{\sqrt{\lambda_1}} \phi_1(\mathbf{p}), \frac{1}{\sqrt{\lambda_2}} \phi_2(\mathbf{p}), \frac{1}{\sqrt{\lambda_3}} \phi_3(\mathbf{p}), ... \right) \tag{2}$$

where ϕ_i and λ_i are the values of the eigenfunction and eigenvalue of the discrete Laplace-Beltrami operator at point p, respectively. It is important to note that the eigenvalues are arranged in ascending order and the eigenfunctions should be appropriately normalized. Since the GPS domain has $1/\sqrt{\lambda}$ dependence, the distances will be dominated by smaller eigenvalues. Constructing a discrete Laplace-Beltrami operator is a highly non-trivial task. Due to better convergence properties, the discrete Laplace-Beltrami operator of Xu [2004] was used for the computations.

Hierarchical mesh segmentation is performed using k-means clustering in the GPS domain to infer the skeletal structure. For the first stage of the segmentation, the body is segmented into its major appendages; the head, body, arms and legs. Afterwards, the results are further segmented into different parts. The arm is segmented into three parts (elbow, forearm, and hand), the body is segmented into four parts (left and right shoulders and left and right hips), and the legs are segmented into three parts (knee, thigh, and foot). Finally, by getting the centroids of each segment, the skeletal structure of the character is generated (see Figures 1c and 1d).

3 Skinning

The character mesh and the extracted skeleton are disconnected until skin attachment specifies how the rotation and translation of the bones apply deformation to the character mesh. The skin attachment is computed by assigning bone weights based on the proximity of the embedded bones smoothed by a diffusion equilibrium equation over that character's surface. Similar to [Baran and Popović 2007], an analogy to heat equilibrium is used to find the weights.

4 Results and Discussion

NCV computation depends solely on how the normals of the mesh are oriented with respect to each other. Using it on meshes that have many concave and convex parts results into having more candidate segmentation points and may result to poor skeleton extraction performance. In comparison to the NCV computation, GPS embedding proved to be more robust and powerful. Not only that the output corresponds correctly to where they should be, it does not depend on the complexity of the mesh. In addition, k-means clustering in the GPS domain results in a deformation-invariant segmentation of 3D models. Finally, using our algorithm on scanned models proved to be possible. Without any pre-processing, our algorithm was still able to extract the skeletal structure of the scanned model.

Since the GPS is an infinite-dimensional vector, it was necessary to determine the number of eigenvalues and eigenvectors to be used for the intended segmentation purpose. By computing the root mean square error (RMSE) of the system output for each model as the GPS domain increases, we were able to determine the optimum GPS domain where we can perform the hierarchical mesh segmentation. Experimental results show that lower dimensional spaces in the GPS domain still exhibit the same segmentation power as the higher dimensional spaces. Hence, the minimum number of eigenvalues is chosen to allow for the algorithm to be more efficient in terms of processing time.

For the majority of the models, the system run time was dominated by the computation of the discrete Laplace-Beltrami eigenvalues and eigenvectors. K-means clustering was used for mesh segmentation which is fast and simple yet still sufficiently robust compared to more costly solutions such as mean shift clustering or agglomerative clustering. Experimental results show that we can easily incorporate high fidelity and interactive characters within minutes.

5 Conclusion

This paper presented a pipeline for automatic skeleton generation and character skinning. Given only the mesh information, the following works are mainly done: Firstly, an improvement to an existing technique in skeleton extraction was done to allow for motion capture data to be utilized quickly and effortlessly. Secondly, a surface representation used in shape classification was explored for mesh segmentation and skeleton generation. Experimental results showed that it is more robust compared to the previous existing technique when applied to a complex mesh input. In addition, our pipeline presented a pose-invariant mesh segmentation result. As a result, input of the same model but with different poses lead to the same skeletal structure. Finally, our system output can be easily incorporated and used to setup skeleton-based animations in various 3D modeling and animation software, games, as well as virtual and augmented reality applications. The pipeline is semi-supervised (for cluster initialization) and easy to implement.

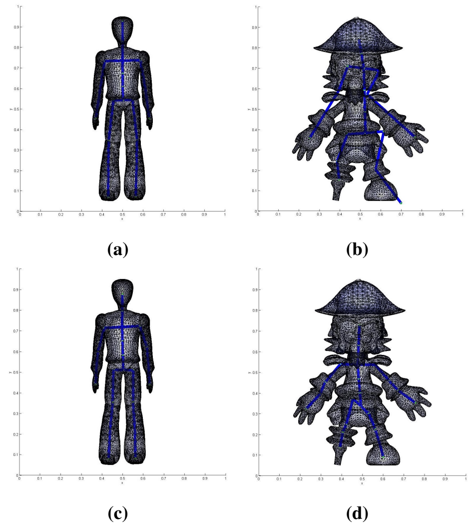

Figure 1: *(top) automated skeleton generation using NCV computation and (bottom) automated skeleton generation using GPS embedding.*

Acknowledgements

The authors would like to acknowledge the following: Synthesizing and Sharing Animation Building Blocks for Rapid Creation of 3D Motion Scenes (ANEEME) Project, Department of Science and Technology (DOST), and Philippine Council for Industry, Energy, Emerging Technology Research and Technology (PCIEERD). Finally, we want to thank Joel Casimiro, our graphic artist, for his contribution in setting up the virtual environment.

References

BARAN, I., AND POPOVIĆ, J. 2007. Automatic rigging and animation of 3d characters. In *ACM Transactions on Graphics (TOG)*, vol. 26, ACM, 72.

BOGO, F., ROMERO, J., LOPER, M., AND BLACK, M. J. 2014. Faust: Dataset and evaluation for 3d mesh registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3794–3801.

CHENG, W., CHENG, R., XIAOYONG, L., AND SHULING, D. 2011. Automatic skeleton generation and character skinning. In *VR Innovation (ISVRI), 2011 IEEE International Symposium on*, IEEE, 299–304.

EGELS, Y., AND KASSER, M. 2003. *Digital photogrammetry*. CRC Press.

IGARASHI, T., MATSUOKA, S., AND TANAKA, H. 2007. Teddy: a sketching interface for 3d freeform design. In *Acm siggraph 2007 courses*, ACM, 21.

RUSTAMOV, R. M. 2007. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proceedings of the fifth Eurographics symposium on Geometry processing*, Eurographics Association, 225–233.

XU, G. 2004. Convergence of discrete laplace-beltrami operators over surfaces. *Computers & Mathematics with Applications 48*, 3, 347–360.