3D Reconstruction Of the Human Torso From Four Orthogonal Model Views Captured Using a Stationary RGB-D Camera

A Thesis Presented to the Academic Faculty

by

Shehab Attia

Research Advisor: Dr. Rudolph L. Gleason Jr.

In Partial Fulfillment of the Requirements for the Degree Bachelors of Science in the School of Biomedical Engineering

Georgia Institute of Technology

April 2019

Acknowledgements

Special thanks to Dr. Gleason for all the advice and mentoring he gave me during my time in his lab. He always pushed me to solve challenging and novel problems, which made me grow in unimaginable ways. Thanks to Dr. Weiler and Dr. Dixon for getting me involved in the project and for their advice and review of my work. Lastly, I'm grateful to all my family and friends who gave me their undivided attention when my code refused to do what it was meant to and I had to talk to them out loud to make sense of what I was doing to debug my code.

Abstract

Digital 3D models of the human body can be used to measure anthropometric features to diagnose diseases. Circumference measurements of body parts can only be obtained from 360-degree spin 3D models. These type of models are created using a combination of scanning hardware and stitching software. Current stitching techniques use an Iterative Closest Point algorithm on manually or automatically generated registration landmarks, provided that there is a rotation angle difference of 60 degrees or less between adjacent frames for sufficiently matching features. This paper proposes a method to stitch four orthogonal (90-degree) views captured using one Kinect camera by iteratively translating and rotating adjacent frames to minimize the Root Mean Square Error between the contours of the torso in 2D planes. Future work will involve mending holes in the stitched models and solving the shear transformation to improve results on non-rigid bodies.

Table of Contents

List of Abbreviations

2D, 3D	Two-dimensional, Three-dimensional
ICP	Iterative Closest Point
IR	Infrared
RGB-D Camera	Camera capable of capturing color and depth.
SCAPE	Shape Completion and Animation of People
GPU	Graphics Processing Unit
RMSE	Root Mean Square Error

Introduction

Digital 3D models of the human body have been shown to be accurate in obtaining anthropometric features, such as lengths, diameters and circumferences of body parts, compared to traditional anthropometric devices, such as tape measures, callipers and pelvimeters.[1, 2] These anthropometric features can be used to diagnose diseases such as Cephalopelvic Disproportion.[2] While the front-view scan can be used to measure 2D metrics such as height, a complete 360-degree spin model is needed to measure circumferences around the body parts, such as the torso, legs or arms. Both scanning hardware, such as a depth camera or laser scanner, and stitching software are required to create a digital 360-degree spin 3D model of a person.[1, 3]

A full-body laser scanner is a device that a human can stand inside. The device emits laser rays around the circumference of the body to provide very accurate 3D models with 2mm accuracy.[3] Full-body laser scanners can cost between $20,000-$50,000 and take up a large amount of space, equivalent at least to the volume of the model being scanned.[4]

Depth cameras are less expensive than laser scanners, costing anywhere between $150-$10,000, and only need 2-3m of space to acquire a scan.[1] The reduction in cost comes at the expense of resolution. Many low-end depth cameras have a resolution of only 640x480 pixels, meaning that the data points captured per frame are very sparse, leading to loss of information of the model being captured.[1] Moreover, captured models can be noisy due to light occlusion or IR reflection.[4] Pre-processing algorithms that separate the body from the background, remove noise from the image, fill holes and rotate the scan to the optimal orientation can be used to improve the quality of scans captured by depth cameras.[4] To obtain

3D models from a depth camera, pre-processing, stitching and meshing steps are needed to combine captured frames.

Common stitching techniques used today require manual human input to identify landmarks to guide the stitching or an automated algorithm to find registration landmarks. Moreover, consecutive frames must have a rotation angle difference of 60 degrees or less, normal to a common plane, in order to have matching features between frames. The most common algorithm used to stitch frames is the Iterative Closest Point (ICP) algorithm, which solves for a transformation matrix that minimizes the distance between registered landmarks on adjacent frames.[3, 5] Other stitching techniques try to fit the acquired scans to statistical human models, such as the Shape Completion and Animation of People (SCAPE) model. The SCAPE model is computationally expensive and training data must be available to use this stitching technique.[6]

Cui et al. and Fernando et al. demonstrate two different ways to a obtain real-time 3D model using a single Microsoft Kinect, while Alexiadis et al. demonstrate real-time 3D model acquisition using six Microsoft Kinect cameras [4,7]. Each proposed method aims to capture 3D scans from different angles so that there are enough overlapping features to stitch multiple images using the Iterative Closest Point (ICP) algorithm.[4,7,9] Cui et al. obtain 10 frames at 30 degree rotation intervals, then use ICP to stitch them together into one 3D mesh. Alexiadis et al. obtain six images from each Kinect camera at 60 degree rotation intervals and fuse them together using ICP. Both methods yield good results. However the biggest issue is that the person must remain in a fixed position during acquisition so that the computer software can easily find matching features in the multiple images scanned.[4.9]

Cui et al. also report that a lot of processing is needed on the images obtained from the Kinect to remove noise and unwanted features.[4] They explain that noisy scans are obtained due to the low resolution of the depth camera and due to reflective materials on the subject, such as hair and clothes.[4] The Kinect camera's infrared sensors produce a lot of noise and holes in the image data when the infrared beams scatter upon hitting hair or black clothing, which is a current limitation of the Kinect camera. Unwanted features are generated by accessories the person is wearing or by objects in the background during the scan, such as lamps or chairs or the floor or the walls, if the subject is standing indoors. All the noise and holes in the data must be cleaned and filled in before stitching can happen. Also background features must be separated from the body in order to generate a 3D model of the body devoid of unwanted elements.[1,4]

Reeda et al. suggest that 3D scans from hand-held devices can be stitched by a method other than ICP.[6] They use a Machine Learning techniques to train the computer to recognize body shapes and generate statistical body models that they can use to fit any new scans onto.[6] An example of a statistical body model is the Shape Completion and Animation of People (SCAPE) model. This type of modelling is very advanced and computationally complex; however, its results are very promising. SCAPE models have pre-identified landmarks such as joints of the wrists, arms, thigh and legs. Therefore, they can be used to animate pose and gestures. This allows for very elaborate animations to be produced. It is not possible to generate

SCAPE models without the training data. Hence, the computation for these models is quite extensive.

Due to the limitations of ICP when there are not enough features between adjacent frames, there are no known ways to automatically stitch scans with orthogonal views that have not been manually landmarked.[3] This paper aims to use a low-cost depth camera to automatically stitch together a torso of a human model using only four perpendicular 3D scans, and without any previous manual landmarking or automated/trained feature detection. This proposed method relies on the bilateral symmetry and curvature of the torso of the human body and uses a Root-Mean-Square Error (RMSE) approximation to find the correct fit of the scans onto each other. The purpose of this method is to eliminate the need for manual user input to identify stitching landmarks and to provide a low-cost and computationally-efficient method of obtaining a 3D model of the human body.

Methods and Materials

A Kinect V2 RBG-D camera was used to capture four orthogonal views of a person, as shown in **Figure 1**. Although a human body is a non-rigid body, the torso's shape stays relatively the same throughout the different views, hence only the torso currently works with this application. The curvature of the edge boundaries of the torso is mostly symmetrical and consistent across views, as shown in **Figure 2**. This proposed method is computationally inexpensive when compared to the SCAPE or Kinect Fusion techniques and should work on any 3D Point Cloud that has uniform and symmetric contour across the four orthogonal views.

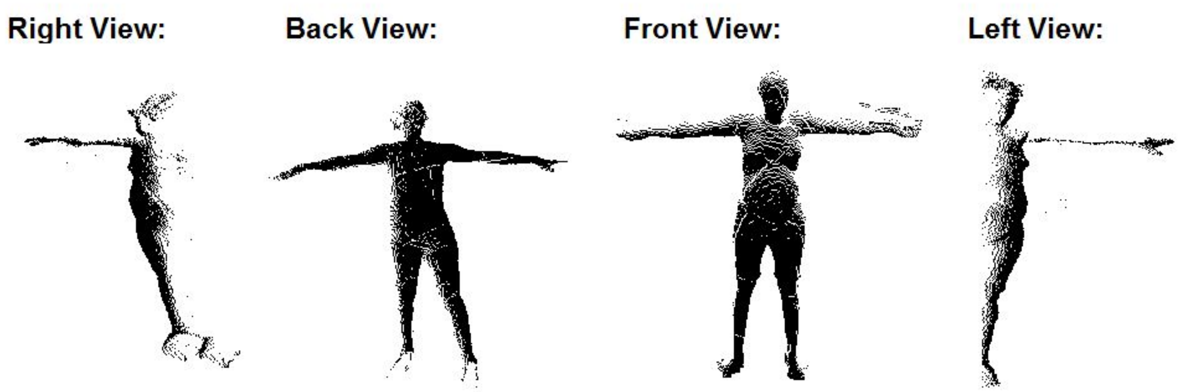

Figure 1. Four orthogonal views (right, back, front, left) captured from a stationary Kinect depth camera.

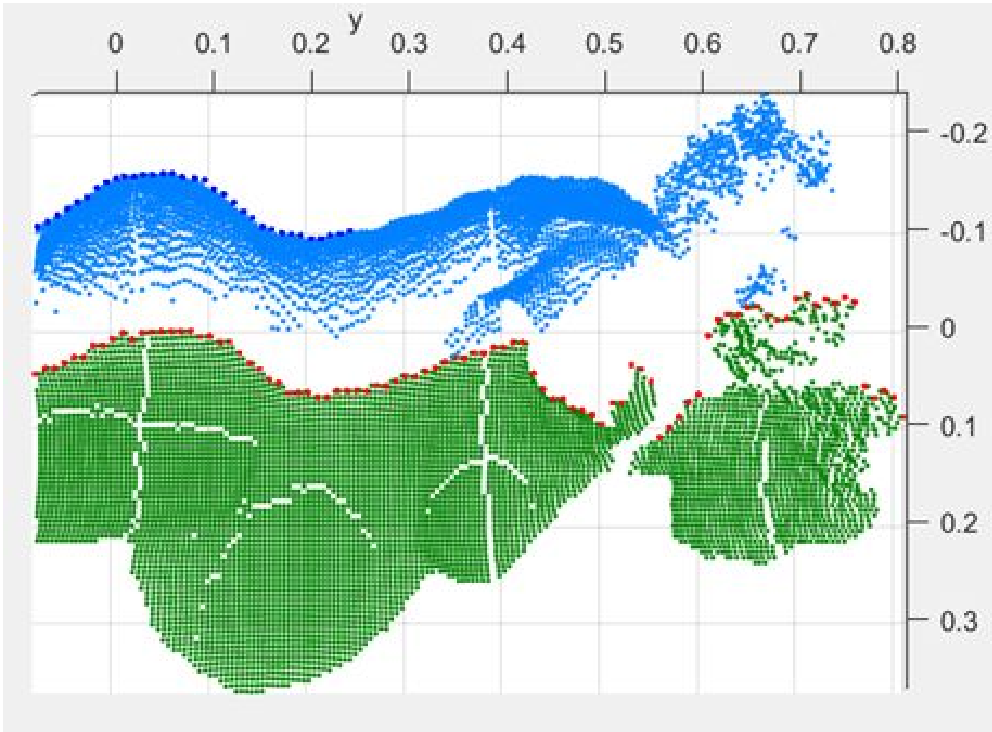

Figure 2. Back view (blue) and Right view (green) shown in the Y-Z plane, with blue and red dots showing the contour of the torso.

A pre-processing script removes noise and background information from the scans, using a density-based K-Nearest Neighbor filter algorithm. Afterwards, an automated landmarking script is used to identify the legs and arms, and those are removed from the scan, leaving the torso and head for stitching.

The main feature used for stitching is the contour of the torso. The contour is extracted by clustering data points using a voxel grid filter, then finding the minimum and maximum points in small slices along the y-plane. The result is a low-density 2D silhouette of the torso, as shown by the blue and red dots in **Figure 2**. Sixteen contours are extracted in total, four for each view, normal to the x and z planes. These contours are used to bring each view closer to the adjacent view.

Predefined step sizes of 0.01m and 2 degree rotation are used to iteratively transform one of the four contours of one view over a fixed contour of another view in the same plane. After each transformation takes place, the Root Mean Square Error (RMSE) is calculated using Euclidean distance and stored in an array.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=0}^{n}(\|V_{1,i}-V_{2,i}\|)^2}{n}}$$

Where V is a vertex in the point cloud and n is the total number of vertices

After all of the transformations are completed at the desired step sizes, the minimum RMSE index is found in the array using a search function, and the transformation rotation and translation matching that RMSE value is used to transform the point cloud before continuing the process with the next set of contours. A simplified diagram of the process is shown in **Figure 3**.

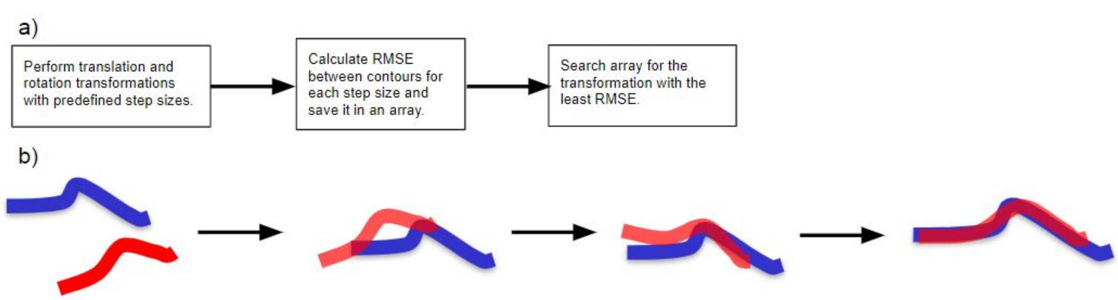

Figure 3. a) A block diagram showing the process to find the transformation needed to minimize the distance between two contours. b) A simplified example of the transformations taking place to minimize the distance between two contours.

Results and Future Work

The matching contours method provides a computationally inexpensive method to reconstruct a 360-degree spin 3D model of the human torso from four orthogonal views. This proposed method overcomes a great obstacle in standard ICP stitching methods, which require at least eight captured views yielding a rotation angle of less that 60 degrees between frames. Moreover, no training data or a GPU is needed to perform this proposed method. When performed on completely rigid models, such as prisms or boxes, the algorithm has almost 100% accuracy. However, when performed on a non-rigid model such as the human torso, there are a lot of areas where the model has incomplete overlap between the adjacent frames, as shown by the white seams in **Figure 4**. These seams are created where the views are not completely symmetrical due to slouching or the person not being completely parallel to the camera during scanning.

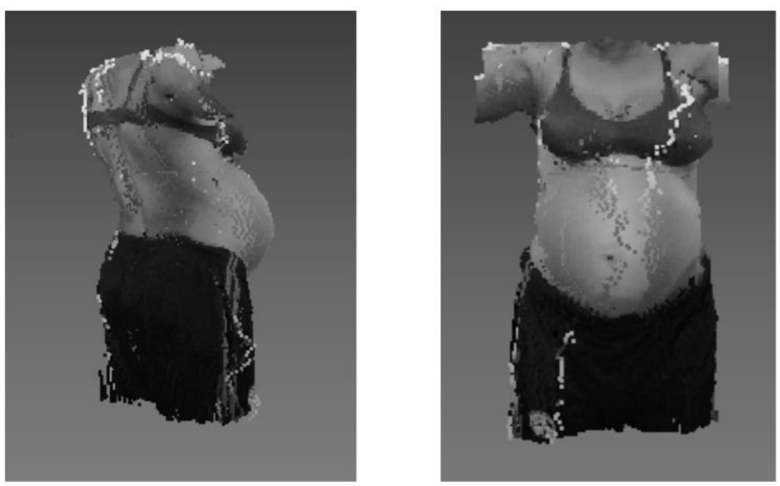

Figure 4. Shows the result from stitching four orthogonal views of the human torso.

Future work would involve developing the algorithm to perform shear transformations in small voxel steps to transform pieces of the non-rigid body onto itself, to overcome the obstacle of slouching or changing posture. Moreover, more work needs to be done to validate the accuracy of this stitching model when compared to a different 360-degree hand-held scanning or stitching method. Lastly, because there are a lot of holes in the model due to the lack of depth information from the orthogonal views, a marching cubes algorithm could be used to average neighboring vertices and fill-in holes to generate a smoother model.

References

1. Niu, J., Zhang, Y., & Wu, J. (2009, 25-27 May 2009). Challenges in the application of three-dimensional anthropometry. Paper presented at the 2009 4th IEEE Conference on Industrial Electronics and Applications.
2. Gleason Jr, Rudolph L., et al. "A safe, low-cost, easy-to-use 3D camera platform to assess risk of obstructed labor due to cephalopelvic disproportion." PloS one 13.9 (2018): e0203865.
3. Lovato, C., Castellani, U., Fantoni, S., Milanese, C., Zancanaro, C., & Giachetti, A. (2009). Computer assisted estimation of anthropometric parameters from whole body scanner data. Paper presented at the 3D Physiological Human Workshop.
4. Cui, Y., Chang, W., Nöll, T., & Stricker, D. (2013). KinectAvatar: Fully Automatic Body Capture Using a Single Kinect. In J.-I. Park & J. Kim (Eds.), Computer Vision - ACCV 2012 Workshops: ACCV 2012 International Workshops, Daejeon, Korea, November 5-6, 2012, Revised Selected Papers, Part II (pp. 133-147). Berlin, Heidelberg: Springer Berlin Heidelberg.
5. Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., . . . Fitzgibbon, A. (2011). KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. Paper presented at the Proceedings of the 24th annual ACM symposium on User interface software and technology, Santa Barbara, California, USA.
6. Reeda, M. P., Parka, B.-K., Kima, K. H., & Jonesa, M. L. (2015). Statistical prediction of body landmark locations on surface scans. Paper presented at the Proceedings 19th Triennial Congress of the IEA.

7. Fernando, J., Contreras, C. R., & Meneses, J. E. (2014). Real-time 3D reconstruction of the human torso using a Kinect sensor. Paper presented at the Latin America Optics and Photonics Conference.
8. AICON 3D Systems breuckmann smartSCAN review - 3D scanner. Retrieved February 15, 2017, from http://www.aniwaa.com/product/3d-scanners/aicon-3d-systems-breuckmann-smartscan/
9. Alexiadis, D. S., Zarpalas, D., & Daras, P. (2013). Real-time, realistic full-body 3D reconstruction and texture mapping from multiple Kinects. Paper presented at the IVMSP Workshop, 2013 IEEE 11th.