3D Photography: Final Project Report

Full Body 3D Scanning

Sam Calabrese Abhishek Gandhi Changyin Zhou
{smc2171, asg2160, cz2166}@columbia.edu

Figure 1: *Our model is a dancer. We capture his full-body 3D model by combining image-based methods and range scanner, and then do an animation of dancing.*

Abstract

In this project, we are going to build a high-resolution full-body 3D model of a live person by combining image-based methods and laser scanner methods. A Leica 3D range scanner is used to obtain four accurate range data of the body from four different perspectives. We hire a professional model and adopt many measures to minimize the movement during the long laser-scanning. The scan data is then sequently processed by Cyclone, MeshLab, Scanalyze, VRIP, PlyCrunch and 3Ds Max to obtain our final mesh. We take three images of the face from frontal and left/right side views, and use a program called Facegen to create an accurate model of the face and an estimated model of the head. We further use 3Ds Max to merge the two parts together, apply skin shader, and create animations. Compared with most existing methods including many image-based 3D modeling and several commercial products using laser scanners, our method exhibit higher precision and resolution.

Keywords: Laser Scanner, Image-Based 3D Modeling, Human Full body

1 Introduction

1.1 Background

In the past decades, enormous 3D sensing algorithms and systems have been developed [Blais 2004]. There are mainly three basic range measurement principles behind these algorithms or systems, triangulation, time-of-flight and laser phase difference [Beraldin et al. 2000]. Laser scanners using time-of-flight or phase difference can provide high measurement precision at long range [Allen 2007]. But they have to measure distances pixel by pixel, so that can only be applied to some still and rigid objects.

Laser scanner based on triangulation can run a little faster since it can emit a slice of laser instead of just a beam. However, it works in a much smaller range, experiences more severe occlusion problems, and still is only applicable to still rigid objects.

Compared with most laser scanners, image-based methods using triangulation principles are much faster and able to provide real-time 3D sensing. These methods include depth from motion [Aloimonos and Spetsakis 1989], shape from shading [Zhang et al. 1999], depth from defocus/focus [Nayar et al. 1996][Watanabe and Nayar 1998][Schechner and Kiryati 2000][Zhou and Lin 2007], and structure from stereo [Dhond and Aggarwal 1989]. They often require the object surface to be textured, non-textured, or lambertian. These requirements often make them impractical in many cases. In addition, image-based methods usually cannot give a precision depth estimation since they do patch-based analysis. Structured light could help to improve their performance significantly and has attracted lots of researchers in these years. However, most of existing systems either only provide low-precision, low-resolution results [Nayar et al. 1996][Du et al. 2007] or require the object to be rigid [Rusinkiewicz et al. 2002][Blais et al. 2004].

1.2 Motivation

Obtaining a precise 3D model of themselves is a dream of many people. [We got a bunch of responses after posting a simple advertisement on Craiglist for a 3D model volunteer.] Once we have the 3D model, there will be lots of interesting applications, including clothes design, animation and augmented realistic.

Precisely modeling a live person is a challenging problem. Simply speaking, laser scan methods are too slow to capture 3D model of live human, and laser is not applicable to the face considering the safety of eyes. Image-based method cannot provide a satisfactory model in the sense of precision and resolution.

1.3 Our Solution

In this project, we are going to model a live human by taking advantage of both image-based methods and laser scanner methods. We plan on using a program called Facegen, which takes 3 images of the face and using the inherent symmetry of the human face creates an accurate model of the face and an estimated model of the head. For the body we will use the Leica 3D range scanner to get accurate 3D model data. Then, we use 3Ds Max to attach the head to the body. Since texture is not incredibly important, as our model will not be fully clothed for more accurate data, we will import the model into a 3Ds Max and provide skin and clothing and the texture for the face is automatically provided from Facegen. With the full body 3D model, we create an animation as illustrated in Figure 1.

Our project presents some interesting technical challenges. For example, one major technical challenge is getting the person who is being scanned to stay in the same position relative to themselves, for example their arms shouldn't move up or down. In our experiment, we hire a professional model, use two well-positioned tripods to rest the arms of model, mark the foot position, and use a camera to track and correct the movement of model. Also the number of scans needs to be reduced because of the person, so that the 3d data will not be as good as a statue where you have the benefit of doing the scans over multiple days where none of the dimensions or orientations will change. Finally texture mapping, clothing, and animating the model is a difficult challenge.

1.4 Comparison

While 3D scanning of human models is used commercially for clothes design, movies and video games, many use either known pattern projection or image based models that although are easier on the person, do not produce perfect results. Or others use dangerous lasers that if you are not careful will hurt the person. Our method combines both of these to produce our model, which will be both a better model and more comfortable for our subject.

2 Experiment Setup and Data Capture

2.1 Body

Not like scanning statues, we only have one chance to obtain all the scan data, since the model is live and cannot return to the original pose exactly after several days. It makes our experiment design critical. The key problem is how to help the model keep stable during the long scan, and return to the original pose for the next scans after rest.

Firstly, we need a professional model who can stand more stable for a long time. We ask many friends around us, contact the people of art school, post an advertisement on the website Craiglist, and finally choose one from tens of responses. We choose to do experiment at night in the Room 6LW3, where has enough space and allows us to avoid possible disturbing.

Two tripods are adjusted to proper distance and heights to rest the arms and reduce the occlusion. The position of feet is marked at the beginning of the scan. Four scan views are planned around the model from frontal-left, frontal-right, back-left to back-right. The distance from the scanner to the model ranges from 2.5 to 3 meters, and the scanner is set as high as the shoulders. Ten targets are placed around the subjects and properly named for the registration usage.

In addition, a D-SLR camera Canon 20D is set up to monitor the model movement. We keep comparing the images during one scan or between different scans to make sure there is no obvious movement.

Scan precision is set to 2mm, so that it shall take less than 15 minutes to complete one scan. Though we only have four scans, the whole scan lasts for four hours because it takes a long time to move the Leica Scanner and set it up in a new position. That's what we didn't expect.

We show the four scan data in Figure 2. Vertices numbers of these initial scans are huge, ranging from 90K to 150K. The high resolution data pose a challenge to our following data processing.

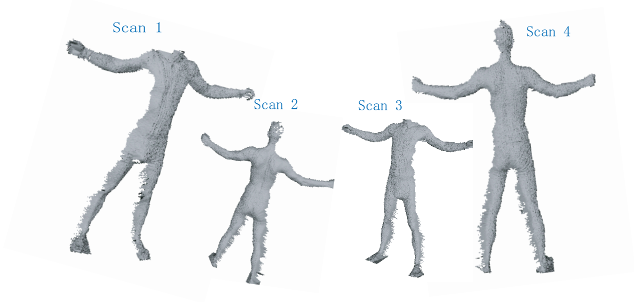


Figure 2: *Four high-resolution surfaces were obtained by using Lecia Range Scanner*

2.2 Head

According to the guide of the program Facegen, we took three pictures of the model as we can see in Figure 3. Facegen makes usage of the inherent symmetry of the human face and provides a rich set of general face models to fit in. By taking three proper pictures and tune the parameters carefully, we finally got the 3D model shown below. This model might not be exactly same as the groundtruth, but looks acceptable.

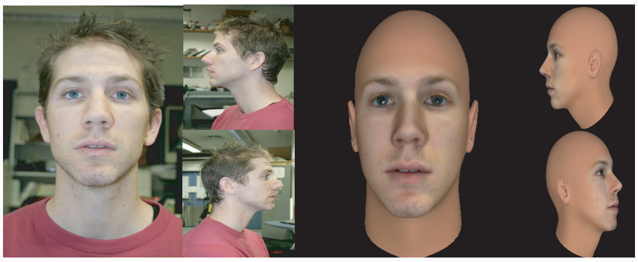


Figure 3: *Use Facegen to estimate the head and face model based on three images*

3 Data Processing

With the scan data, we tried to do registration using Cyclone, the software of Lecia scanner. However, Cyclone can only provide a quite rigid registration. The slight movement of the model during the scan (we can discuss it in more details in Section 5), causes serious unmatch problems. The registered model using Cyclone is shown in Figure 4. We can see obvious unmatch at the elbows and legs.

So first we exported the data to a PTX file from Cyclone and converted it to a PLY file using the software Meshlab. Unfortunately the ply files from Meshlab were not able to be used. When we exported the PTX files from Cyclone, Cyclone put all of the scans into the same coordinate system and put the camera overhead. So when we tried to use VRip to resolve the surface from the point cloud, it would not get proper weights to assign to vertices. VRip uses the dot product between the view direction and the normal for the surface to get a weight. This weight represents the confidence that the algorithm has in this point. But since the view direction in these

Figure 4: *3D model created by Cyclone. The slight movement during the scanning causes serious reconstruction problems*

PLY files did not correspond to the original orientation or camera position, we received erroneous results.

Therefore, to solve this problem each surface was pulled back into Meshlab and rotated back to the original orientation by hand. Also, it was noticed that the model had moved slightly during one of the scans and his right arm was significantly forward, in order to get a good surface we deleted all of the points on the arm in that scan. Then these new PLY files were re-registered using Scanalyze, a program for aligning range image data distributed by Stanford, who also distribute VRip. Using Scanalyze, we registered the data and exported it for VRip, a function within the program which saved the transformation matrices to get each scan into a single coordinate system. This method did not destroy the original view direction and allowed VRip to extract a surface from the points and fill all the smaller holes using space carving.

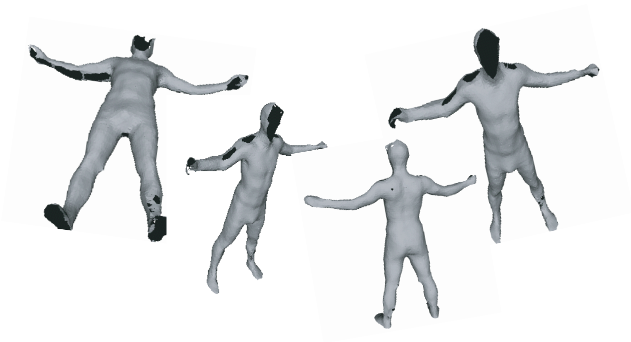


Figure 5: *3D model created by using Vrip and refined by Meshlab (before hole-filling)*

The resulting meshes were too large to actually work with, containing over 3 and a half million polygons. Therefore, we needed to decimate the mesh using PlyCrunch, a simplification program included with VRip. This was run three separate times to simplify the mesh to under 10000 polygons. The resulting mesh was full of holes because the triangles had been deleted but the points were accurate and part of the surface created within VRip. So this decimated mesh was brought back into Meshlab and resurfaced from the points and smoothed. See Figure 5.

Finally the mesh was brought into 3dsMax, where using an incredibly simple hole filling method was used to fill any remaining holes. This method just identifies the holes and connects the vertices from one edge to their closest neighbor on the other edge. Following hole filling, the mesh was smoothed again, the scanned points on the models head were removed and the head from Facegen was attached to the mesh. See Figure 6. Then using 3ds Max's biped object and stock motion capture data, the mesh was rigged and animated. Then using Mental Ray, a skin shader was applied to mesh which implements Sub-surface scattering for realism. Several snapshots of our animation are shown in Figure 1.

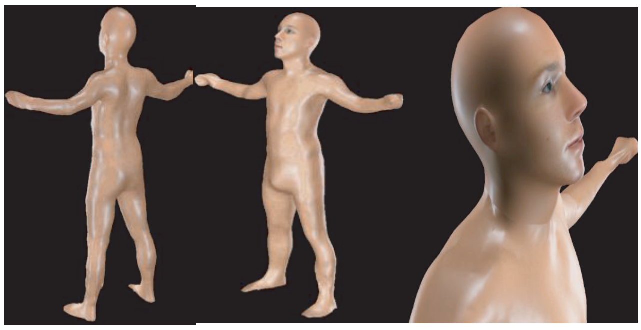


Figure 6: *3D model after hole filling, Skin texture mapping, and attaching the head by using 3D Max*

4 Comparison

As is afore-mentioned above, we segregated the scanning of the head and the body neck-down. Thus, we ended up using both, a safer, laser-free image based method, and the more accurate laser scan. In order to analyze the quality of our outputs, we performed a qualitative comparison with the outputs obtained by a couple of professional companies, specialized in full body human scan. One of them, AvatarMe, relies on image based methods, and the other one, Headus, performs a laser scan. We put in more effort and time in order to make up for the cost limitations faces by us compared to commercial organizations. The criteria used for the comparison are as follows:

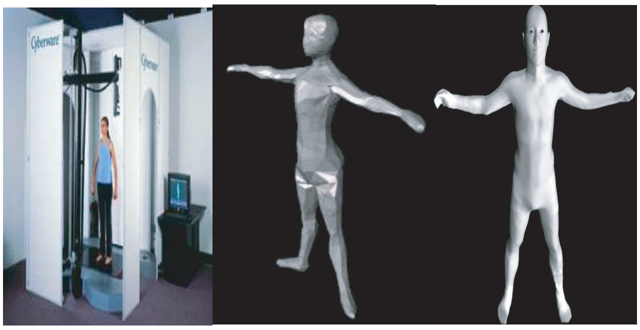
1. Resolution of the output
2. Accuracy of the scanned data
3. Error prevention during scanning
4. Cost/Ease of hardware setup

Based on the previously mentioned criteria, the following is a qualitative comparison of our output, with that of two professional companies, specializing in full body human scans. (See Figure 7):

1. Resolution of the output Commercial organizations have significantly higher resources, such as time, money and people at their expense, to support high level of quality. Lacking the above, we made way for quality via quantity of data captured. For example, as opposed to the 3 million polygons scanned by the commercial organizations on average, we scanned for 3.8 million polygons. Also, we used a point density of 2mm as opposed to 4mm as specified by the company Headus.

2. Accuracy of scanned data Post scanning, the scans captured, go through registration, meshing, hole-filling, etc. with the use of some softwares. We did not stick to just one good software and actually used a combination of many to get the best results. The softwares we used were, Cylone(Scan and Register), Scanalyze(Prepare for VRip), VRip(Surfaceing), PlyCrunch(Mesh decimation, cleaning and decimation), MeshLab(Remesh) and 3dsMax (Fill Holes, further cleaning, Attach Head, Animate, etc...).

3. Error prevention during scanning & Cost/Ease of the setting up the hardware: Both the above mentioned criteria go hand in hand and have been clubbed together. The professionals invest their resources in platforms such as the WBX platform, to ensure that the movement and inaccuracies within a single scan and between multiple scans is minimal. Also, we didn't have a predefined dataset to match our outputs against. Given our limits, and even otherwise, our final output was extremely satisfying and of a finest quality.



(a) Settings of AvatarMe (b) Results of AvatarMe (c) Our Results

Figure 7: *Compare with the results of AvatarMe. Our result exhibits higher resolution*

5 Discussion

Our result is not bad, compared with some commercial systems, especially when you notice that our method doesn't use any prior body model and only relies on the scan data. Actually, we could have done better if several mistakes were avoided during our experiment. We didn't put the camera at the best position, though setting a camera to track the movement is a good idea. So that during one scan, the camera is occluded by the scanner. The movement during this scan causes some troubles for the late processing and significantly reduces our model precision.

Hand is another problem. Fixing the hand position is difficult. It might be a better choice to scan the hands separately and then attach to the body afterward.

Though it is nearly a pure engineering project, we can still see several research topics. For example, how to best plan the scan views for the least occlusion and best 3D reconstruction? Another interesting research topic is non-rigid registration, considering the body movement/distortion to some extend is predictable.

Acknowledgements

Special thanks to our model, Daniel, for his professional, passion and great cooperation. Thanks to Prof. Allen, Karan, Matei and Paul for your kindly help and support.

References

ALLEN, P., 2007. 3d photography 2007 fall. Class notes on Active 3D Sensing.

ALOIMONOS, Y., AND SPETSAKIS, M. 1989. A unified theory of structure from motion.

BERALDIN, J., BLAIS, F., COURNOYER, L., GODIN, G., AND RIOUX, M. 2000. Active 3D Sensing. *Modelli E Metodi per lo studio e la conservazione dell'architettura storica*, 22–46.

BLAIS, F., PICARD, M., AND GODIN, G. 2004. Accurate 3d acquisition of freely moving objects. In *3DPVT04*, 422–429.

BLAIS, F. 2004. Review of 20 years of range sensor development. *Journal of Electronic Imaging 13*, 231.

DHOND, U., AND AGGARWAL, J. 1989. Structure from stereo-a review. *Systems, Man and Cybernetics, IEEE Transactions on 19*, 6, 1489–1510.

DU, H., ZOU, D., AND CHEN, Y. Q. 2007. Relative epipolar motion of tracked features for correspondence in binocular stereo. In *IEEE International Conference on Computer Vision (ICCV)*.

NAYAR, S., WATANABE, M., AND NOGUCHI, M. 1996. Real-time focus range sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*, 12, 1186–1198.

RUSINKIEWICZ, S., HALL-HOLT, O., AND LEVOY, M. 2002. Real-time 3D model acquisition. *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 438–446.

SCHECHNER, Y., AND KIRYATI, N. 2000. Depth from Defocus vs. Stereo: How Different Really Are They? *International Journal of Computer Vision 39*, 2, 141–162.

WATANABE, M., AND NAYAR, S. 1998. Rational Filters for Passive Depth from Defocus. *International Journal of Computer Vision 27*, 3, 203–225.

ZHANG, R., TSAI, P., CRYER, J., AND SHAH, M. 1999. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 8, 690–706.

ZHOU, C., AND LIN, S. 2007. Slit Aperture Pairs for Depth from Defocus and Out-of-focus Deblurring. *ICCV 2007 (submission)*.

Tools and software:
Facegen: http://www.facegen.com
Meshlab: http://meshlab.sourceforge.net
Vrip, Plycrunch: http://graphics.stanford.edu/software/vrip/
Scanalyze: http://graphics.stanford.edu/software/scanalyze/
3Ds Max: http://www.autodesk.com/3dsmax