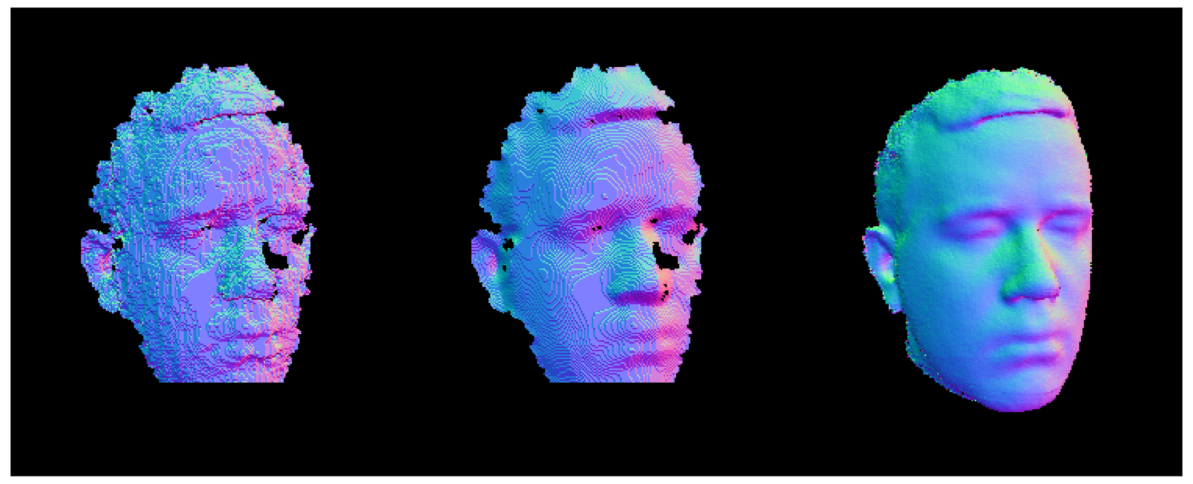


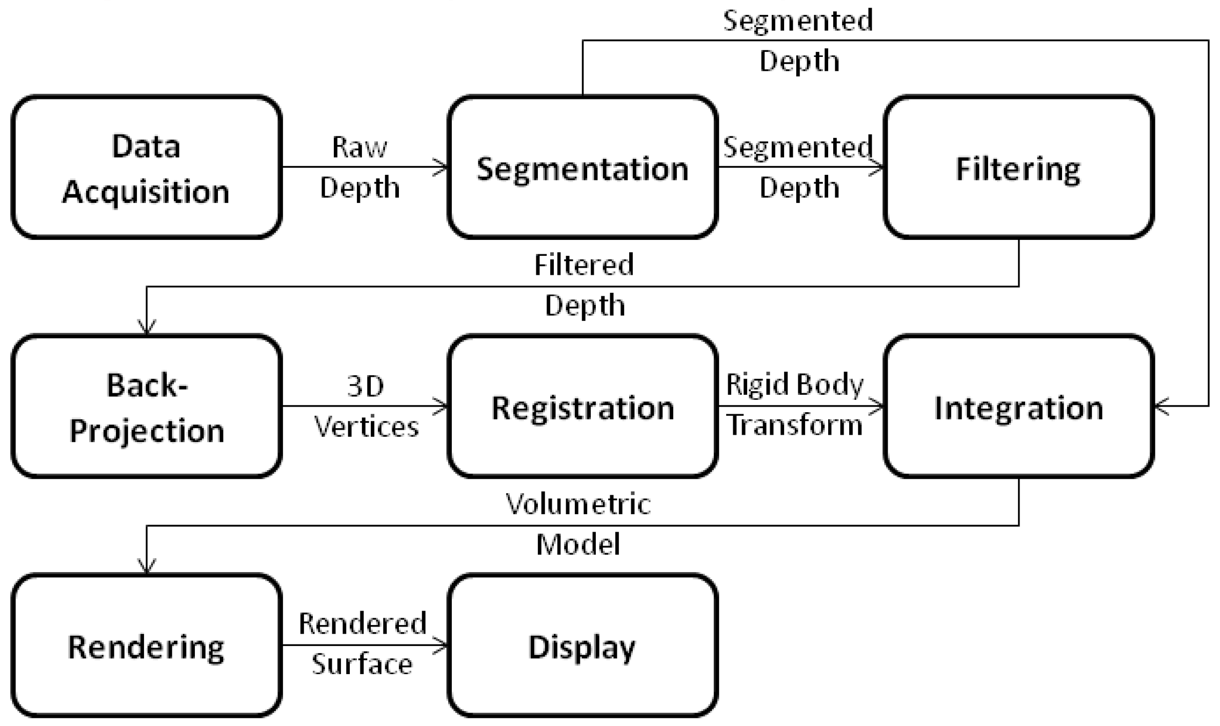
Real-Time 3D Face Modeling with a Commodity Depth Camera

Gregory P. Meyer and Minh N. Do
{gmeyer3, minhdo}@illinois.edu

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Overview

Our system generates a high-quality 3D face model by segmenting, registering, and integrating depth images captured by a commodity depth camera.



Contributions

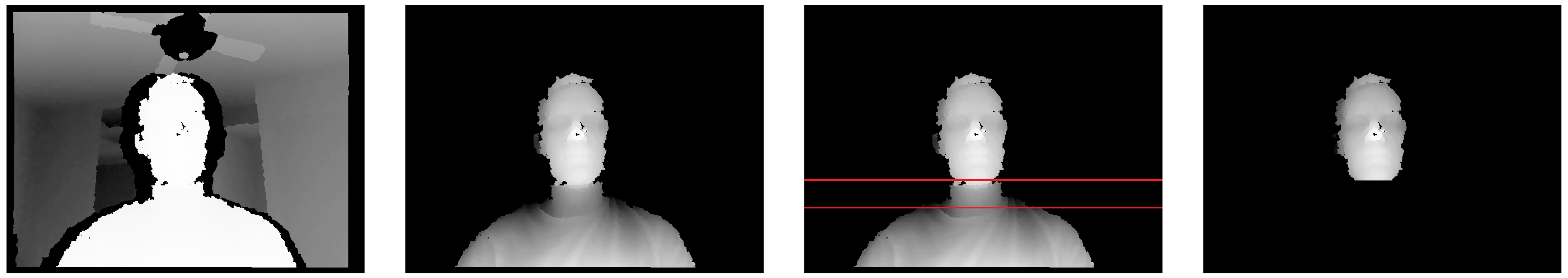
Our method is built upon the KinectFusion system [1] with modifications to enable segmented depth images to be robustly aligned. Our main contributions include the following:

1. Efficient segmentation of a person's face from a depth image.
2. Additional filtering to eliminate unreliable depth measurements typically due to the person's hair.
3. Pre-aligning the segmented depth images to improve the point correspondences found using projective data association for the iterative closest point (ICP) algorithm [2].

Segmentation

A depth image contains depth measurements for the entire scene, however, we only want to consider depth pixels on the person's face. Our method segments the depth image into foreground and background regions, where the foreground contains the user. Connected component analysis is used to determine the foreground region.

To determine the location of the person's head within the foreground region, we look for a horizontal scan line that separates the foreground into head and torso regions. The best separating line is the one that maximizes the difference between the average width of the head region and the average width of the torso region.



Input Depth | Foreground Depth | Head & Torso Regions | Segmented Depth

Filtering

Depth measurements captured by the Kinect are often noisy, especially in regions that contain hair. Noise can affect our ability to correctly align the depth images.

To improve registration, we eliminate depth measurements with a high local variance. In addition, we use a bilateral filter to reduce noise while preserving depth discontinuities.

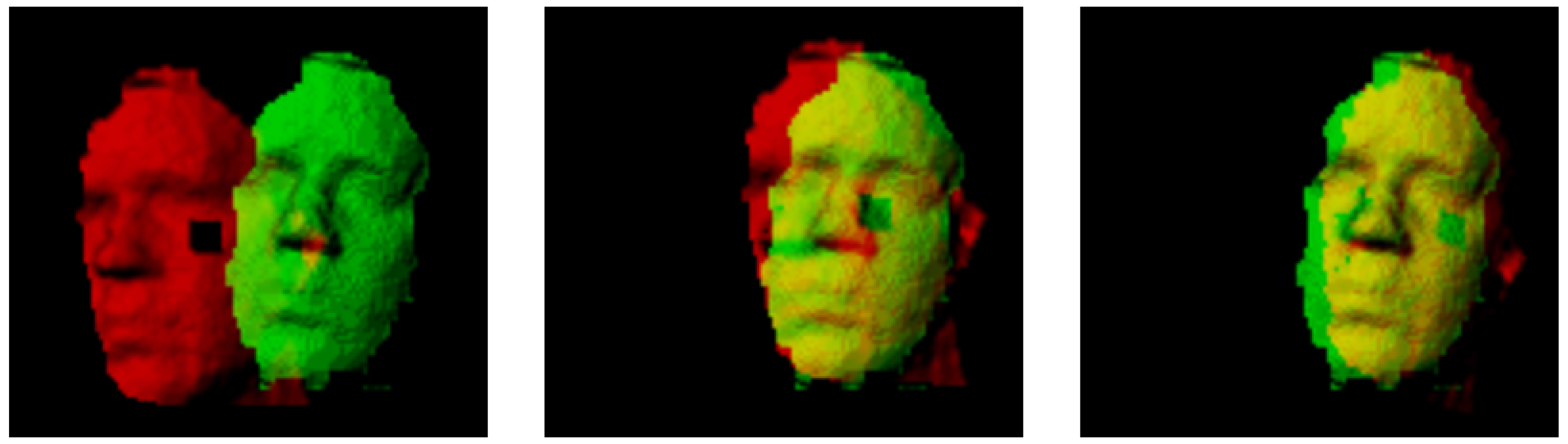
Input Image | Filtered Image

Registration

To construct the 3D face model, we need to align several depth images captured from different viewpoints. For each view, we must determine the rigid body transformation that takes the depth measurements from the camera's coordinate system to a global coordinate system.

ICP is a common method of computing a transformation that aligns 3D point clouds. [3] suggests a high speed variant of ICP that uses projective data association. Projective data association enables point correspondences to be computed quickly by searching along camera rays [2].

Finding corresponding points along camera rays becomes difficult when the depth images are segmented. To increase the number of corresponding points and improve registration, the views are pre-aligned based on their center of mass.

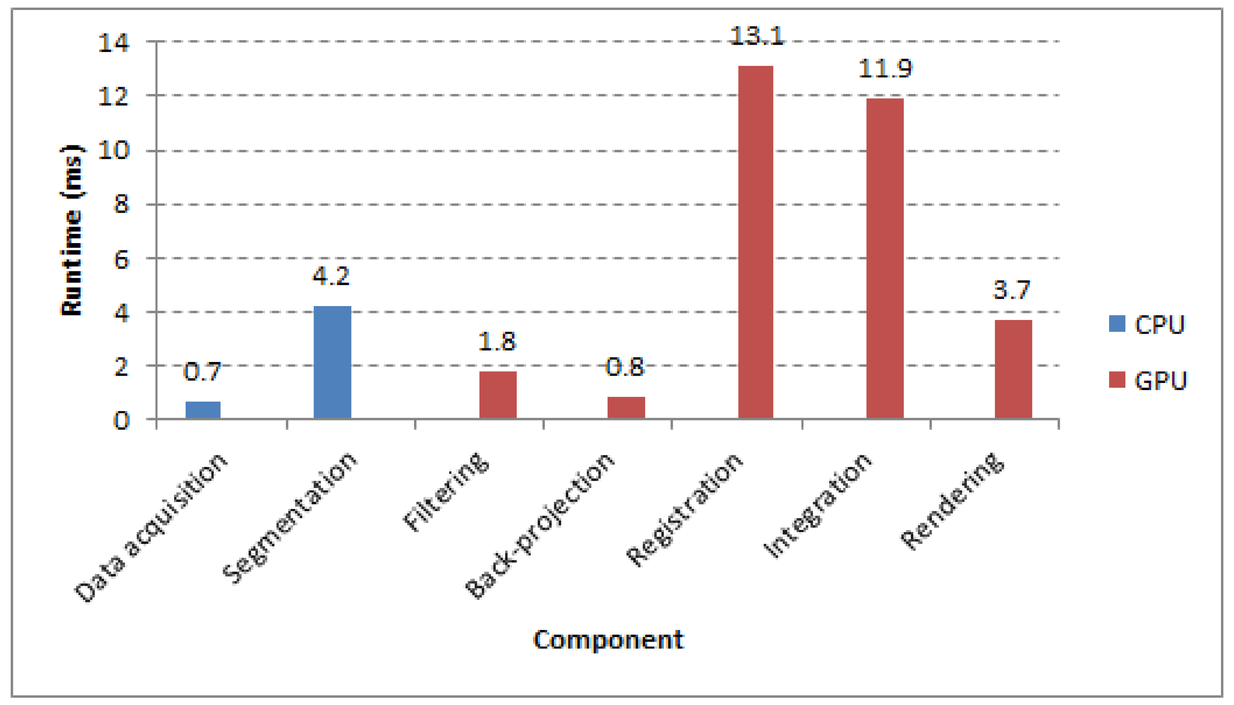


Unaligned Views | Pre-aligned Views | Aligned Views

Results

Input Image | Filtered Image | Output Model

The above model was generated using our system. Using an Intel Core i7 CPU and a NVIDIA Geforce GTX 570 GPU, our system is capable of processing more than 25 frames per second.



References

[1] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pp. 127–136, 2011.

[2] G. Blais and M. Levine, "Registering multiview range data to create 3d computer objects," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 17, no. 8, pp. 820–824, 1995.

[3] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pp. 145–152, 2001.

Source Code

Our system was implemented using C++ and CUDA, and the source code is available at
`http://www.ifp.illinois.edu/~gmeyer3/`