Motion Compensation and Evaluation of 3D Reconstruction of Head

Unnat Antani
Mentors: Can Kocabalkanli, Ozgur Guler, Reza Seifabadi
Johns Hopkins University

Introduction

- Worked on implementing algorithms for motion compensation.
- Worked on implementing state of the art Deep Learning based method for scene reconstruction.
- Worked on implementing and testing solution on custom real world dataset using a phone camera.
- Provided a robust tool for physicians to use, obviating need of measuring using tools like a measuring tape.

Problem

- It is difficult to measure and quantifies deformities on baby head.
- No specialized tools available for physicians
- If such deformities are left undiagnosed, they can cause complications later on and it is also more difficult to correct the deformity.
- Current implementation in the pipeline includes 2D vision based solution.

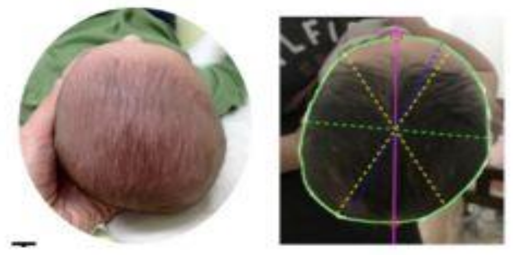

Fig 1

- However, a 3D scan would be able to capture much more details than a 2D vision based solution would.
- We implement and explore the results of two methods:
  - Multi-view Registration
  - Instant Neural Graphics

Multi-view Registration

- This method is based on aligning different pieces of geometry in global space
- It relies on representing the point clouds as pose-graphs
- The problem is then converted to feature matching on graphs
- The graphs store the position of the geometries.

Data Acquisition

- A baby head phantom is used as replica. Scans are acquired by a 3D scanner as shown below:

Fig 2

Fig 3

Implementation

- The scans are taken incrementally in a sequence based on the number of scans to take.
- The scans are then uploaded to AWS server, from which the necessary files are acquired for further processing

Multi-view Registration

Results

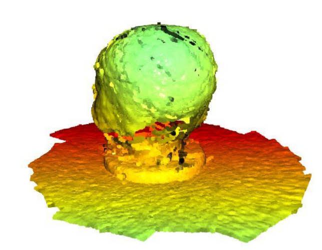

Fig 4

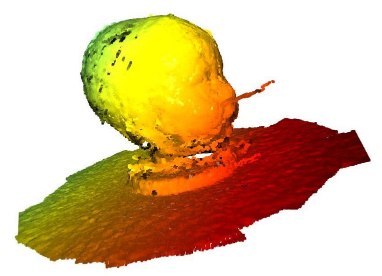

Fig 5

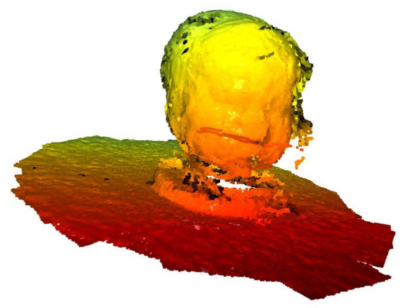

Fig 6

- The reconstruction occurs with lots of artefacts, however, generating acceptable results.

Instant Neural Graphics

- The basis of this approach is to represent a scene with **neural network** as **primitive**, through **overtraining** and **hash table encoding** at various resolutions.
- 2 Multilayer Perceptrons(MLP) are used in this case.
- The **density MLP** maps the hash encoded position to 16 output values, **coefficients of harmonic basis**.
- The input of **color MLP** is the output of density MLP, which determines the **projection based on input**. The final output is an **RGB triplet** for each position. (Fig 7.)

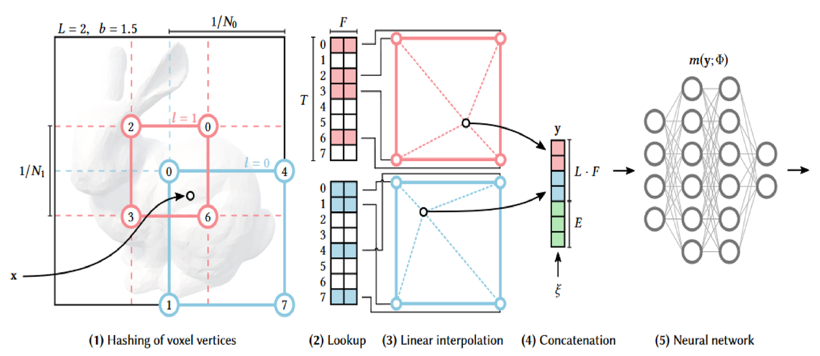

Fig 7[6]

Data Acquisition and Processing

- The Images are taken through phone and processed to input into a scene reconstruction software.
- During imaging, the **phantom is moved** by approx. **3 cm** to replicate noise through **motion**, apart from natural camera motion during capture.
- COLMAP uses all the images and the camera intrinsic and using feature matching, reconstructs the camera poses, registering all the images and generating all the necessary transformations(Fig 8.)
- These transformation along with the images are the input to the NeRF neural network.
- The output of the 3D scene reconstructed are shown in Fig 9. and Fig 10.

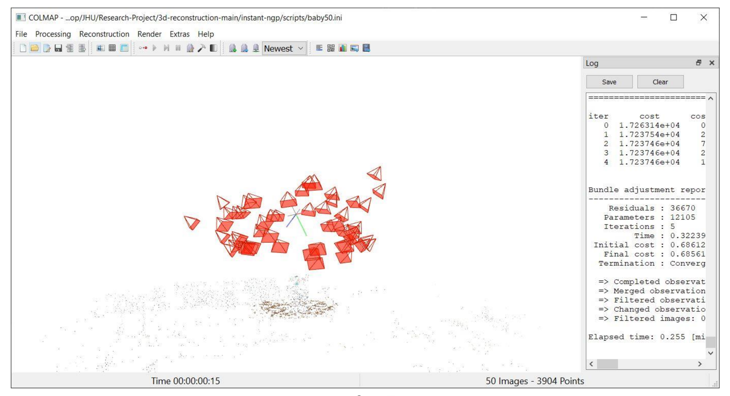

Fig 8

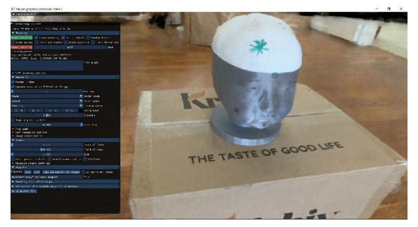

Fig 9

Fig 10

Results

- The 3D model reconstructed, after manual cleaning in Blender[8] is shown in Fig 11.
- The final registration against ground truth is shown in Fig 12.
- To calculate the CVA, another software was used which extracts the values based on landmarks. The results are shown in Fig 13. The average CVA obtained for this method is 1.7 mm, meeting requirement of < 2mm.

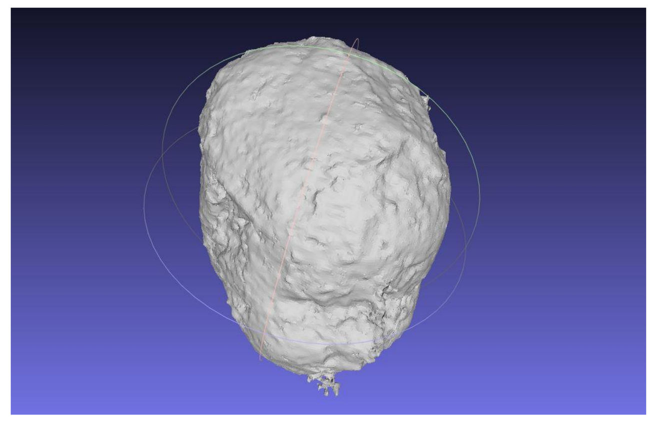

Fig 11

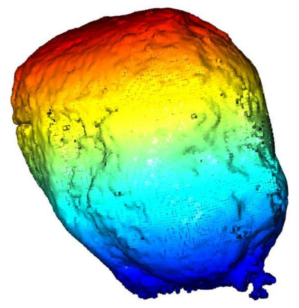

Fig 12

Measurements	Calculated
Head Circumference	205.1 mm
Largest Circumference	20.0 mm
CI	82.50
CVAI	-2.65
CVA	-1.83

Fig 13

Conclusion

Multi-view Registration

- Captures a lot of details through point clouds.
- Multi-view registration requires use of **3D scanners.**
- Requires **more number of scans for better resolution.**
- Scans can only be taken in **sequence.**
- Takes approximately a **minute** for entire pipeline to be implemented.
- It may or may not be able to form **closed loop**, based on number of scans taken. Also generates **artefacts**.

Instant Neural Graphics

- Relies only on **smartphone cameras**.
- No sequential imaging required.
- Has a very **simple** and **small** Neural Network Architecture.
- Learns and generates scenes in **milliseconds**.
- **3D reconstruction** gives results within **acceptable margins.**
- System must have a **GPU** for implementation.
- It also requires use of another open source software, i.e., **COLMAP**, for camera pose reconstruction.
- The model has to be **cleaned manually** to remove unwanted points before comparing with ground truth.

Future Work

- Refining and fine-tuning the Neural Networks
- Exploring and improving method of hash encoding
- Improving 3D reconstruction algorithm
- Automating image acquisition from phone through an app.

References

[1] 3D Point cloud Registration and Stitching [Link]
[2] Multiway Registration [Link]
[3 Robust Reconstruction of Indoor Scenes [Link]
[4] Instant-ngp [Link]
[5] NeRF Project Page [Link]
[6] Instant Neural Graphics Primitives with a Multiresolution Hash Encoding [Link]
[7] COLMAP [Link]
[8] Blender [Link]

Acknowledgments

This project was made possible by the incredibly valuable insights by the mentors; Can Kocabalkanli, Ozgur Guler and Reza Seifabadi. I would also like to thank Babak V-Ghaffari, who joined the project later on and worked alongside me to give more insights and test the approach and the data. I would also like to thank Professor Russell Taylor for pushing to be more detail oriented by giving feedback on the presentations and the reports.