

# Project Proposal

## ”Head Alignment using Deep Closest Point”

### 1 Abstract

With the advent of recent technologies, multi-view RGB-D recordings have become the prevalent way of data acquisition in operating rooms (OR). Previous works established the benefits of using 3D information to detect and anonymize faces of 2D images in such multi-view settings [1]. However, real world 3D data often suffers from noisy and incomplete point clouds, which yield erroneous alignments using probabilistic point set alignment methods like coherent point drift (CPD) [5] and its variants. In this project, we address this issue by creating and fine-tuning a deep learning based point set registration method to achieve more robust rigid transformations on noisy and incomplete point clouds from real world data. Our contributions are as follows:

- We compile a new synthetic dataset consisting of different head alignment problems
- We adapt the DCP method to replace the rigid registration step of the works of Bastian and Wang et al. [1]
- We show qualitative (maybe quantitative) results of the pre-trained and our fine-tuned model on data acquired from real world applications

### 2 Requirements

#### 2.1 Overview

The work of Bastian and Wang et al. [1] has introduced a method to detect and anonymize people’s faces in the OR by leveraging 3D data. In their pipeline, they perform a rigid registration between the mesh of a head and a cropped point cloud by using the off-the-shelf method FilterReg [3] and iterative closest point (ICP) [2] to align the mesh to the point cloud. However, inadequate mesh estimations, sparse and incomplete point clouds make it unreliable in certain situations to use probabilistic methods. Our goal is to develop a neural network based rigid point set registration method to align a given head mesh to the point cloud by using the architecture of DCP [6]. Fig. 1 depicts the upper body of a parametric human mesh model with the point cloud data.

#### 2.2 Dataset

The data used in Bastian and Wang et al. [1] is still undisclosed, hence we perform our experiments on different data. The data we use was acquired from the chair of Computer Aided Medical Procedures (CAMP) at TUM (Fig. 2) by capturing a simulated OR environment from six calibrated ceiling-mounted Microsoft Azure Kinect cameras. Each frame consists of six RGB images and a fused point cloud of the scene.

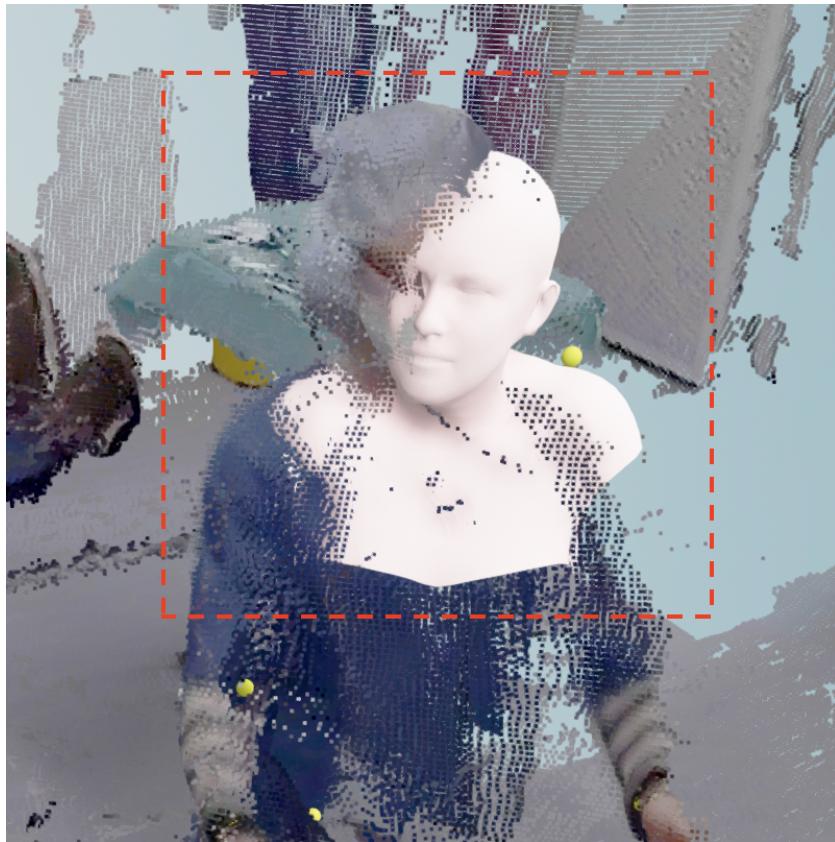


Figure 1: The upper body of an estimated parametric human mesh model with the point cloud. For a rigid registration it would be necessary to extract the head of the mesh.

### 2.3 Methods

**Training** The authors of DCP [6] already published their code base on GitHub<sup>1</sup>. We will first use their pre-trained model and compare their method qualitatively against existing methods like FilterReg [3] and ICP [2]. Subsequently, we extend the ModelNet 40 [7] dataset to adapt it to our use case. Apart from augmenting the ModelNet 40 dataset by mirroring real life data (noise and sparseness), we add synthetic data that resembles our use case. More specifically, we generate a synthetic dataset by sampling different human poses using the parametric human mesh model SMPL [4] and converting it to a point cloud. We copy the head, apply a random transformation to it and perform several augmentations on the original SMPL model point cloud.

**Evaluation** We will evaluate our approach qualitatively. Evaluating our approach quantitatively might become difficult, as there is no ground truth data available for our dataset. Moreover, generating ground truth is not trivial, as transformations can be

---

<sup>1</sup><https://github.com/WangYueFt/dcp>

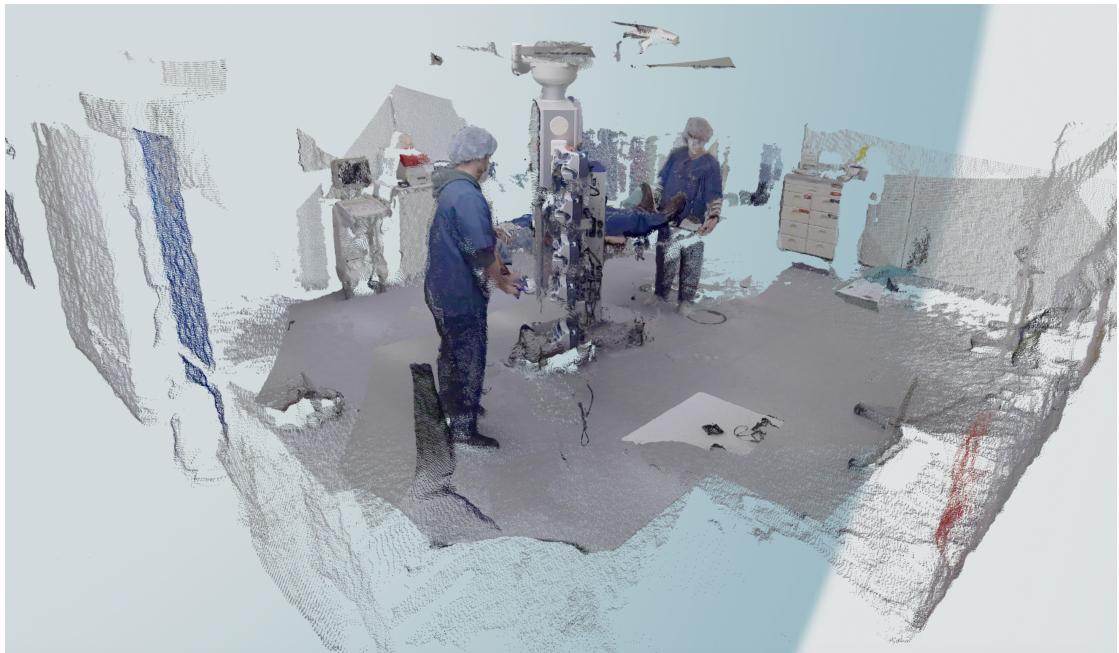


Figure 2: Example of a point cloud from our dataset we use for evaluation.

ambiguous if the point cloud is incomplete. We could, however, evaluate our approach quantitatively using the method introduced in Bastian and Wang et al. [1] by annotating face bounding boxes in the 2D images and rendering the head meshes back into the 2D images.

### 3 Team

Tony Wang 03726379  
Yushan Zheng 03763652  
Yutong Hou 03765252  
Johannes Volk 03731455

### References

- [1] Lennart Bastian, Tony Danjun Wang, Tobias Czempiel, Benjamin Busam, and Nassir Navab. Disguisor: Holistic face anonymization for the operating room. Manuscript submitted for publication, 2022.
- [2] P.J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
- [3] Wei Gao and Russ Tedrake. Filterreg: Robust and efficient probabilistic point-set registration using gaussian filter and twist parameterization. *CoRR*, abs/1811.10136, 2018.

- [4] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015.
- [5] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, 2010.
- [6] Yue Wang and Justin M. Solomon. Deep closest point: Learning representations for point cloud registration. *CoRR*, abs/1905.03304, 2019.
- [7] Zhirong Wu, Shuran Song, Aditya Khosla, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets for 2.5d object recognition and next-best-view prediction. *CoRR*, abs/1406.5670, 2014.