



Faculty of Engineering and Information Technology

Human Body 3D Scanner (Virtual me)

Esteban Gabriel Andrade Zambrano

Student Number: 12824583

Project Number: AUT-21-04017

Major: Mechanical and Mechatronics Engineering

Supervisor: Dr. Teresa Vidal Calleja

A 12 Credit Point Project submitted in partial fulfilment of the requirement for the Degree of

Bachelor of Engineering

May 28, 2021

Statement of Originality

I, Esteban Gabriel Andrade Zambrano declare that I am the sole author of this report, that I have not used fragments of text from other sources without proper acknowledgment, that theories, results and designs of others that I have incorporated into my report have been appropriately referenced and all sources of assistance have been acknowledged.

Refer to Appendix for signed statement of originality

Abstract

Human Body 3D Scanner (Virtual me) (12cp)

Esteban Gabriel Andrade Zambrano - AUT-21-04017

Supervisor: Dr. Teresa Vidal Calleja

Major: Mechanical and Mechatronics Engineering.

Objective: Process recorded sensor data from a human body and create a scaled model of the scanned subject that can be used for clothing fitting.

The capability of scanning different models is critical for manu industries ranging from fashion to medical and manufacturing. Nevertheless many of the current implementations are extremely expensive. Hence it imposes a barrier for the technology. Similarly, many of the current applications do not create extremely accurate models and in various occasions thgse models would be discarded.

Therefore, this project will create both hardware and software implementation for a human 3D scanner with a moderate budget as well as accurate results in order to be able to produce accurate and scaled models of the scanned subject. The proposed method for data acquisition will be with a custom rig. The rig will be a set of poles with cameras that will capture images from the person that is inside the rig. Additionally, a Lidar will be used to capture a PointCloud as well as an Image mof the subject inside the rig.

With the acquired images from all the cameras as well as the Lidar, a photogrammetry process will be used in order to obtain an initial mesh. This mesh is the result of the photogrammetry reconstruction process. Moreover, the initial mesh will be processes with a series of developed algorithms in a pipeline sequence. The pipeline will process the mesh with operations such as plane removal, outliers removal and clustering.

The Scaling will use the PointCloud from the Lidar as a reference and adjust the Scale and Pose of the processed mesh. Once the Mesh has been scaled appropriately with the reference PointCloud, a Poisson algorithm will be used in order to reconstruct and obtain the final scanned model.

The results in this report illustrate the process on how with a constrained data set, it is possible to obtain accurate scaled scanned model that can be used in many applications.

Acknowledgements

I would like to acknowledge and thank my supervisor Dr. Teresa Vidal Calleja for her help and support throughout this project. I would also like to acknowledge and thank Dr. Cedric Le Gentil for all his guidance, support and input throughout the project development. Furthermore, I would like to acknowledge Mr Asher Katz for his contributions in the development of the rig and the image acquisition process component of the project.

On the other hand, I would also like to acknowledge Dr. Mark Liu as well as the UTS Robotics Institute for assistance and resources provided for the completion of this project.

Similarly, I would like to acknowledge the Open Source Community, as assets developed by this community helped in the development and completion of this project.

Finally, I would like to thank my family for their massive and ongoing support throughout this project and my studies. In equally importance I would like to thank God, who with his countless blessings have gave me the opportunity to develop and complete this project.

Contents

List of Figures	vi
List of Tables	viii
1 Introduction	1
1.1 Reseach Question	1
1.2 Project Contextualization	1
1.3 Problem Definition	1
1.4 Background	2
1.5 Applications	3
1.6 Overview	4
1.6.1 Methodology	4
1.6.2 Structure	4
2 Literature Review	6
2.1 Photogrammetry	7
2.2 Meshroom	8
2.2.1 Camera Initialization	10
2.2.2 Feature Extraction	10
2.2.3 Image Matching	11
2.2.4 Feature Matching	12
2.2.5 Structure From Motion	13
2.2.6 Prepare Dense Scene	15
2.2.7 Depth Map	15
2.2.8 Depth Map Filter	16
2.2.9 Meshing	16
2.2.10 Mesh Filtering	17
2.2.11 Texturing	17
2.3 PCL	18

2.3.1	Statistical Outlier Removal Filter	19
2.3.2	Pass Through Filter	20
2.3.3	Voxel Grid Filtering	21
2.3.4	Plane Model Segmentation	22
2.3.5	KdTree	23
2.3.6	Euclidean Cluster Extraction	24
2.3.7	Principal Component Analysis	26
2.3.8	Moving Least Squares	26
2.3.9	Poisson Surface Reconstruction	29
2.4	Open3D	29
3	Methodology	31
3.1	Data Collection	31
3.1.1	Layout & Configuration	33
3.1.2	Cameras Calibration	36
3.2	Photogrammetry Reconstruction	38
3.3	Data Processing Framework	52
4	Results	53
5	Conclusion	54
6	Recommendations	55
Bibliography		56
A Appendix		61

List of Figures

2.1	Photogrammetry Principle Illustration	7
2.2	Meshroom Graphical User Interface	8
2.3	Removing outliers using a StatisticalOutlierRemoval filter	20
2.4	Filtering a PointCloud using a PassThrough filter	21
2.5	Illustration of a voxel grid containing color values	21
2.6	Illustration of Plane model segmentation	23
2.7	Dataset Before & After RANSAC algorithm (Strutz, 2016)	23
2.8	Example of a 2-dimensional k-d tree	24
2.9	Example of point cloud cluster	25
2.10	Moving Least Square Smoothing	27
2.11	Curvatures of MLS before & after	27
2.12	Uniform Density Upsample Method After & original	28
2.13	Poisson Reconstruction Example. On the left is the input pointcloud dataset, whereas on the right is output of the Poisson Surface Reconstruction Algorithm	29
2.14	Open3D Poisson Reconstruction Example	30
3.1	Tall Pole Cameras Location With Housing	32
3.2	Lidar Mounted in Pole	32
3.3	Cameras Lay-out connection	33
3.4	Tall Pole Cameras Configuration.	34
3.5	Open3D Poisson Reconstruction Example	34
3.6	RIG scanning With Poles Locations	35
3.7	Example from Captured Dataset	37
3.8	Successful Meshroom Pipeline (2021 version)	38
3.9	CameraInit Node Settings	39
3.10	FeatureExtraction Node Settings	40
3.11	ImageMatchingNode Node Settings	41
3.12	FeatureMatching Node Settings	42

3.13 Structure From Motion Output	43
3.14 Structure From Motion Settings	44
3.15 PrepareDenseScene Settings	45
3.16 DepthMap Node Settings	46
3.17 DepthMapFilter Node Settings	47
3.18 Meshing Node Output	48
3.19 Entire Meshing Node Settings	49
3.20 MeshFiltering Node Settings & Output	50
3.21 Texturing Node Settings & Output	51

List of Tables

1.1	3D Scanning Applications	3
2.1	Meshroom CameraInit Node	10
2.2	Meshroom FeatureExtraction Node Settings	11
2.3	Meshroom Image Matching Node	12
2.4	Meshroom Feature Matching Node	13
2.5	Meshroom StructureFromMotion Node	14
2.6	Meshroom Prepare Dense Scene Node	15
2.7	Meshroom Depth MapNode	16
2.8	Meshroom Depth Map Filter Node	16
2.9	Meshroom Meshing Node	17
2.10	Meshroom Mesh FilteringNode	17
2.11	Meshroom Texturing Node	18
3.1	Cameras Specification	32
3.2	Intel RealSense L515 Specification	33
3.3	Cameras Calibration Parameters	36
A.1	Risk Matrix	61
A.2	Comunication Plan	62

Introduction

1.1 Reseach Question

"Human Body 3D Scanner: The development of software for 3D data reconstruction of a Human body scanner with multiple sensors"

1.2 Project Contextualization

The project is based on creating a Human Body 3D scanner. It will have two specific streams that include the development of the mechatronic design of a 3D scanner for a human and the software development for 3D data reconstruction. This proposal is based on developing the software for 3D data modelling and reconstruction of the Scanned data.

Similarly, with the 3D reconstructed model of the human has the aim to be utilised to test different fashion clothing items. This has the intent to adjust the sizing of the clothes fittings based on the Scanned data. The clothing models will adjust automatically depending on the dimensions of the data of the scanned model. The project will have different stages that range from testing different sensors for data acquisition, testing different data stitching frameworks to the deployment of the software in the 3D scanner mechatronic device.

1.3 Problem Definition

Being able to scan different object and models is crucial for many industries. Many applications are in the used in the fashion industry, medical industry, manufacturing ,etc. However, many of the given implementations are extremely expensive , thus making the

technology inaccessible for many companies and users in general. There are many forms of implementations, as there are multiple technologies in the market that facilitate the process in which several devices and software techniques are used. Nevertheless, there is no current industry application that maximises the potential use of the Human body 3D models. Many of the challenges faced is that the software implementation for 3D reconstruction of the models is not particularly accurate, therefore creating imperfect models that on many occasions will need to be discarded.

Hence, this project component will contribute and develop the technology to produce software that will be able to produce accurate models from the gathered data from the sensors. These models will be utilised to try different fashion items and adapt the size fittings accordingly. With the competition of this project many stakeholders, industries and institutions could rely on accurate software that will allow to create a 3D model of a person or object.

1.4 Background

The human society has the world comprehension of the surrounding world through visual perception. This principle allows differentiating distinctive kinds of shapes, objects, colours, textures, and the spatial pose of the surroundings. Based on this information, it is possible to analyse the number of objects in a determined location, object type, object size, object pose in different coordinate frames. Thus, it impacts how as a society we interact with objects or scenes. As a result, it is essential to imitate this perception to acquire real-world data in different formats that include:

- RGB images
- Depth images
- 3D point clouds
- Multispectral images
- Laser readings

All these acquire data can be obtained from a wide variety of commercial or industrial sensors. With this data, it will be possible to use computer processing techniques to model the object or scene (Murcia et al., 2018).

1.5 Applications

In recent years, the use of 3D body scanners has gained importance in several industries. Within the fashion industry, it can aid clothes manufacturers to obtain accurate body measurement data of body dimensions. As mentioned by Sturm et al. (2013), this new technological approach has the potential to alter the future of the fashion and clothing manufacturing industry.

With the rise of innovation of 3D image reconstruction, the interest to gather precise measurements of humans has raised. Due to the fact, that in the clothing industry is extremely important to create better fittings for different shapes of human bodies. Furthermore, virtual try-on solutions have gained popularity in physical and online retail stores (Spahiu et al., 2014).

On the other side, 3D scanners have gained participation in the medical industry. These systems are described as "non-invasive and low cost", thus making it appealing for epidemiological surveys and clinical uses. (Treleaven and Wells, 2007) The geometrical measurements could be associated with shape, size, volume, and surface area of the body parts. It could aid to be a sustainable approach to screen children and patients with obesity, deformities, or specific anatomic defects. Therefore, it will ease the diagnose process and allow to treat and monitor medical conditions holistically and improve the life quality of patients with non-invasive tests. The table below illustrates the use of a 3D scanner in the medical field with the purpose to identify and monitor various medical conditions. From which the diagnose, treatment and monitor procedures will differ based on the acquired data.

Application	Epidemiology	Diagnosis	Treatment	Monitoring
<i>Measurement</i>				
Size	Anthropometric surveys	Growth defects	Scoliosis	Fitness and diet
Shape	Screening	Abdominal shape	Prosthetics	Obesity
Surface area		Lung volume	Drug dosage	Diabetes
Volume			Burns	
<i>Visualization</i>				
Head		Melanomas	Eating disorders	
Chest			Facial reconstruction	
Whole body			Cosmetic surgery	

Table 1.1: 3D Scanning Applications
(Treleaven and Wells, 2007).

1.6 Overview

As mentioned before, project consists of developing a 3D scanner that is allowed to scan person and be able to use that data in order to be able to virtually fit different clothing. This capstone project mainly focuses on the software and processing of the data for the 3D reconstruction.

The aim of this project is to develop algorithms and an implementation pipeline that will use the acquired image data as well as the reference pointcloud data in order to produce an accurate and appropriately scaled 3D reconstructed model of a person.

This reconstructed and scaled model of the scanned person, is intended to be utilised as the model for testing virtual clothe fitting and be able to correctly determine the size of the garments. This technology could have a massive industry in the clothing industry and online retailers as this could solve an issue with the returns policy due to incorrect sizing for different customers groups.

1.6.1 Methodology

To achieve the aim, a rig will be used to capture the data. The Rig will consist of a set of poles with multiple cameras. Furthermore a Lidar Intel RealSense L515 will be used to capture another image and the corresponding pointcloud that will be used for reference. Once the rig captures the data (28 images) as well as the lidar data(1 image, 1 pointcloud), the images will be processed with photogrammetry approach with Meshroom. The Processed data from Meshroom will produce a preprocessed reconstruction mesh.

After this the mesh will be processed in the developed algorithmic pipeline and it will be properly scaled with the reference pointcloud from the Lidar. Once the mesh has been processed and scaled, it will be reconstructed appropriately and it will generate a final model of the scanned person with the with only the necessary components and appropriate scale.

Once the process is completed the mesh will be saved to disk as a ply file that can be later use in other applications such as clothing fitting.

1.6.2 Structure

This Report will start with a literature review in order to analyse and fully understand the algorithms and techniques that are used in both components. It will help to understand the techniques used in the photogrammetry step and the developed algorithm pipeline.

Following this, details will be presented and illustrated as to how the process is carried out and the results that were obtained from different data sets. It will fully illustrated how to import the data to Meshroom, what parameters are required to add and modify in order to change the most robust reconstruction. There will be an indetail explanation of what each step executes.

Furthermore, there will be an detailed explanation of the algorithm pipeline for data processing. Here it will be explain what each component is expected to execute as well as how it was created and wrapped into a framework.

Literature Review

Being able to scan different objects and subjects has been a challenging task for researchers. Getting an accurate spatial location of the objects is crucial for this type of application. The use of 3D point clouds has facilitated this process as it allowed to obtain the following parameters:

- Depth
- Intensity
- Pulse width
- Light echo

This information can be obtained with different kind of sensors. There is a wide variety of off the shelf sensors that can provide 3D point clouds. These sensors could either be stereo or multiview vision cameras, lasers, time-of-flight sensors (*TOF*) and structured light sensors as stated by Murcia et al. (2018).

Many Scanning devices will use single or multiple of the above-mentioned sensors to acquire data. Once the data is obtained, it essential to have a framework for 3D data modelling and reconstruction. The principle behind 3D data reconstruction is obtained with data fusion from RGB-D sensors. This kind of sensors provide 3 channels images RGB (red, green, blue) and the depth images are mapped to each pixel. Based on this data, 3D point clouds could be generated for data reconstruction.

Similarly many other scanning devices use photogrammetry as a technique in order to do the reconstruction. In this particular project Meshroom was used as a 3D photogrammetry reconstruction software.

Moreover, several open source libraries were used in order to process and scale the draft mesh. These open source libraries include PCL and Open3D. From these several algorithms were used that aid in the data processing and these will be explained below.

2.1 Photogrammetry

Photogrammetry is described as the associated techniques with performing measurements of real-world objects and terrain features from images as mentioned by Aber et al. (2019). Many of the applications include the quantification of distances, volumes, areas, heights, 3D topographic mapping, measuring of objects, extraction of 3D pointcloud for surfaces reconstruction as well as the generation of orthophotographs and digital elevation models.(Aber et al., 2019).

In recent years, with the development of technologies pairing computer vision concepts & algorithms and photogrammetry specifically *Structure from Motion-Multiview Stereo (SfM-MVS)* has led to significant advances in 3D surface Reconstruction from images (Aber et al., 2019).

The key principle behind all the photogrammetric measurements is associated with the mathematical & geometrical reconstruction of the path of light rays from the object to the sensor camera in the exact moment of data image acquisition. Thus, the fundamental concept of photogrammetry is the understanding of geometric characteristics of a single photograph.

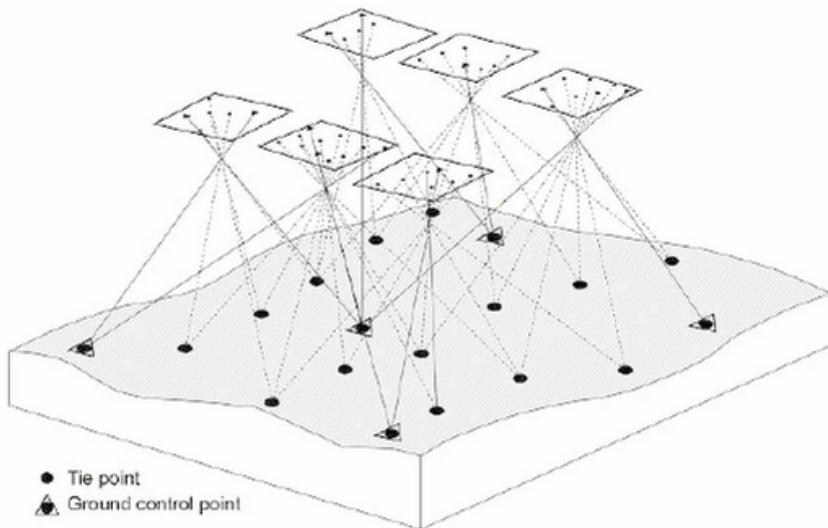


Figure 2.1: Photogrammetry Principle Illustration
(Aber et al., 2019)

2.2 Meshroom

Meshroom is a free, open-source 3D Reconstruction Software based on the AliceVision framework (AliceVision, 2021). AliceVision is a software framework based on photogrammetric computer vision, which focuses on the development of 3D Reconstruction and Camera Tracking Algorithms. It aims to provide robust software foundation with novel computer vision algorithms that can be analysed, tested and reused.

Meshwoom was developed as collobaration project between industry and academia in order to develop cutting-edge algorithms that are robust and of high quality which are crucial for production environments.

Components Pipeline

Meshroom can be downloaded and used in both Windows and Linux and it will require a powerfull CPU in order to have adequate performance. Also it will require an Nvidia GPU as it requires to run CUDA for different nodes and components. Once meshroom is downloaded it can run in either platform. When it is executed the Meshroom GUI will appear and it will be similar to figure 2.2

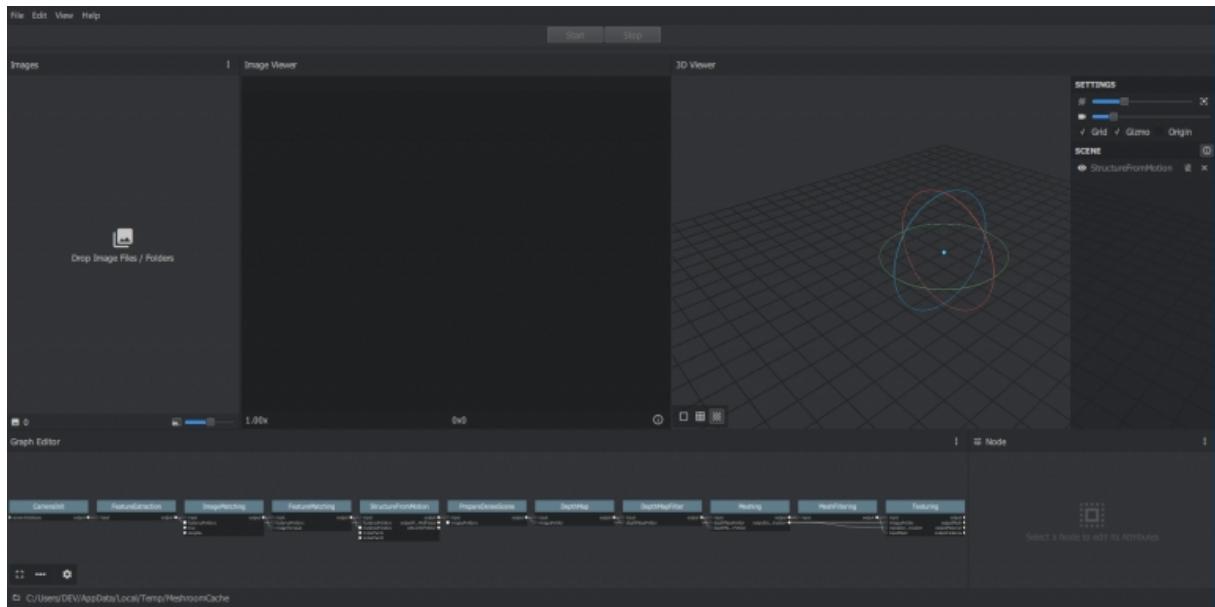


Figure 2.2: Meshroom Graphical User Interface
(AliceVision, 2021)

As mentioned before, mushroom is based on AliceVision's Framework. Therfore it follows a photogrammetric pipeline for 3D reconstruction. This pipeline will have the following components:

- Natural Feature Extraction
- Image Matching
- Features Matching
- Structure from Motion
- Depth maps estimation
- Meshing
- Texturing
- Localization

The pipeline will be included into Meshroom GUI and it will be used in the nodes. All the components in Meshroom will be embedded into different individual nodes that will be executed and produce individual results that will be send to the next node along in the pipeline. As the nodes represent specific components of the reconstruction pipeline, these should be executed in a defined order.

The nodes in the defined order to reconstruct a 3D model from images include:

1. Camera Initialization
2. Feature Extraction
3. Image Matching
4. Feature Matching
5. Structure From Motion
6. Prepare Dense Scene
7. Depth Map
8. Depth Map Filter
9. Meshing
10. Mesh Filtering
11. Texturing

2.2.1 Camera Initialization

Camera Initialization or "*CameraInit*" loads the image metadata, sensor information, camera parameters and it will generate viewpoints based on the images. It is possible to use a mixture of cameras & focal lengths. This node will generate a groups of intrinsics that are based on the images metadata, these group os intrinsics can be adjusted if the cameras have been fully calibrated. The intrinsic include the camera matrix which includes the focal length (f_x, f_y) and principal point (x_0, y_0).

$$CameraIntrinsics = \begin{pmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

The components in the node are described on table 2.1

	viewpoints (1 Element for each loaded image)
Viewpoints Input	<ul style="list-style-type: none"> - ID - Pose ID - Image Path - Intrinsic: Internal Camera Parameters (Intrinsic ID) - Rig (-1 - 200) - Rig Sub-Pose: Rig Sub-Pose Parameters (-1 - 200) - Image Metadata: (list of metadata elements)
Intrinsic Camera Intrinsics	<ul style="list-style-type: none"> (1 Element for each loaded image) - ID - Initial Focal Length: Initial Guess on the Focal Length - Focal Length: Known/Calibrated Focal Length - Camera Type: pinhole', 'radial1', 'radial3', 'brown', 'fisheye4' - #Make: Camera Make (not included in this build, commented out) - #Model: Camera Model - #Sensor Width: Camera Sensor Width - Width: Image - Width (0-10000) - Height: Image Height (0-10000) - Serial Number: Device Serial Number (camera and lens combined) - Principal Point: X (0-10000) Y(0-10000) - DistortionParams: Distortion Parameters - Locked(True/False): If the camera has been calibrated, the internal camera parameters (intrinsics) can be locked. It should improve robustness and speedup the reconstruction.
Sensor Database	Camera sensor width database path
Default Field Of View	Empirical value for the field of view in degree 45° (0°-180°)
Verbose Level	verbosity level (fatal, error, warning, info, debug, trace)
Output SFMData File	... /cameraInit.sfm

Table 2.1: Meshroom CameraInit Node

2.2.2 Feature Extraction

The key objective of this node is to extract distinctive pixel groups, which are to a certain extend invariant to a change in camera viewpoints during the image acquisition process. Thus, a feature in this scene should have similar features descriptors in all the captures images.

One of the most used feature detection method is **SIFT** (*Scale-invariant feature transform*) algorithm. The objective of SIFT is to extract discriminative patches in an initial image that later can be compared to discriminative patches of a second image, irrespective of scale, rotation and translation (Lowe, 2004). As a relevant detail is only present at a

specific scale, the extracted patches are centered around the point of interests. Therefore, the SIFT invariance can be used to deal with image transformations that occur when the viewpoints change during image acquisition.

Based on the representation of one image at different scales, SIFT is able to compute scale-space maxima of the Laplacian Representation. This is a defined image energy-based representation, using the differences of Gaussians. This Maxima is associated to the points of interest in the image. Once this is processed, it samples for each of the this maxima a square image patch, whose origin is the maximum and "x" direction is the dominant gradient at the origin as suggested by Lowe (2004) and Otero (2015). For each keypoint, a description of these patches is associated.

The description consists of statistics of gradients which is computed in regions around the keypoint. The size of the region is defined by the keypoint scale and the orientation by the dominant axis. It is also frequently stored in 128 bits.

The number of extracted features could vary as a result of texture complexity, from one image to other ones or in different image sections. Hence, a post-filtering step controls the number of extracted features to a specified limit. Furthermore, grid filtering is used to ensure an even repartition in the image.

The components in the node are described on table 2.2

Name	Description
Input	SMDdata file.
Descriptor Types	Descriptor types used to describe an image. ‘sift’, ‘sift*float’, ‘sift*upright’, ‘akaze’, ‘akaze*liop’, ‘akaze*mldb’, ‘cctag3’, ‘cctag4’, ‘sift*ocv’, ‘akaze*ocv’
Descriptor Preset	Control the ImageDescriptor configuration (low, medium, normal, high, ultra). Configuration “ultra” can take long time !
Force CPU Extraction	Use only CPU feature extraction.
Max Nb Threads	Specifies the maximum number of threads to run simultaneously (0 for automatic mode). (0-24) 0
Verbose Level	verbosity level (fatal, error, warning, info, debug, trace).
Output Folder	Output path for the features and descriptors files (*.feat, *.desc).

Table 2.2: Meshroom FeatureExtraction Node Settings

2.2.3 Image Matching

The principle behind this node is to find images that point to the same areas of interest. Therefore, image retrieval techniques are implemented with the purpose of finding images that share content without the demand of resolving all the detailed feature matches. Hence, the goal is to simplify the image in a compact image descriptor. This allows to efficiently compute the distance between all images descriptors.

A vocabulary tree is one of the widely used methodologies to generate the image descriptor. It works by passing all the extracted features descriptors into it. Then it performs a classification process, which compares their descriptors to the ones on each node of the vocabulary tree as mentioned by Nistér and Stewénius (2006). Each feature descriptor

is associated with one leaf, which can be stored with an standard index (*The index of this leaf in the tree*). Thus, the collection of leaves indices represents the image descriptor (Nistér and Stewénius, 2006).

The components in the node are described on table 2.3

Name	Description
Image	SfMData file
Features Folders	Folder(s) containing the extracted features and descriptors
Tree	Input name for the vocabulary tree file ALICEVISION_VOCTREE
Weights	Input name for the weight file, if not provided the weights will be computed on the database built with the provided set
Minimal Number of Images	Minimal number of images to use the vocabulary tree. If we have less features than this threshold, we will compute all matching combinations
Max Descriptors	Limit the number of descriptors you load per image. Zero means no limit
Nb Matches	The number of matches to retrieve for each image (If 0 it will retrieve all the matches) 50 (0-1000)
Verbose Level	verbosity level (fatal, error, warning, info, debug, trace)
Output List File	Filepath to the output file with the list of selected image pairs

Table 2.3: Meshroom Image Matching Node

2.2.4 Feature Matching

The key component of feature matching node is to be able to match all features between potential image pairs. Initially, the node performs a photometric process that creates matches between the descriptors from 2 separe images. For each feature in image "A", a list of potential features in image "B" is generated. The descriptor space is not linear and defined space, which causes uncertainty in the validity of the matches due to the absolute distance values. Furthermore, in order to remove bad candidates, an assumption process associates only one valid match in the other image. Hence, for each feature descriptor on Image "A", two of the closests descriptors in the image are used with a relative threshold between them which provides a robust criteria as mentioned by Lowe (2004).

This process will provide a photometric list of feature matching candidates. Afterwards, the images features positions are used in a geometrica filtering process that uses epipolar geometry in an outlier detection framework, which is RANSAC ("Random Sample Consensus"). Moreover, a random selection process uses a small set of feature correspondances and it computers either the fundamental or essential matrix. Once this process is completed, the number of features verifies and validates this model and iterates through the RANSAC framework (Muja and Lowe, 2009). The components in the node are described on table 2.4

Name	Description
Input	SfMData file
Features Folder	
Features Folders	Folder(s) containing the extracted features and descriptors
Image Pairs List	Path to a file which contains the list of image pairs to match
Describer Types	Describer types used to describe an image **sift**/ `sift_float`/ `sift_upright`/ `akaze`/ `akaze_liop`/ `akaze_mldb`/ `cctag3`/ `cctag4`/ `sift_ocv`/ `akaze_ocv`
Photometric Matching Method	For Scalar based regions descriptor * BRUTE_FORCE_L2: L2 BruteForce matching * ANN_L2: L2 Approximate Nearest Neighbor matching * CASCADE_HASHING_L2: L2 Cascade Hashing matching * FAST CASCADE_HASHING_L2: L2 Cascade Hashing with precomputed hashed regions (faster than CASCADE_HASHING_L2 but use more memory) For Binary based descriptor * BRUTE_FORCE_HAMMING: BruteForce Hamming matching
Geometric Estimator	Geometric estimator: (acransac: A-Contrario Ransac // loransac: LO-Ransac (only available for fundamental_matrix model)
Geometric Filter Type	Geometric validation method to filter feature matches: **fundamental_matrix** // essential_matrix // homography_matrix // homography_growing // no_filtering
Distance Ratio	Distance ratio to discard non meaningful matches 0.8 (0.0 - 1)
Max Iteration	Maximum number of iterations allowed in ransac step 2048 (1 - 20000)
Max Matches	Maximum number of matches to keep (0 - 10000)
Save Putative Matches	putative matches (True/False)
Guided Matching	the found model to improve the pairwise correspondences (True/False)
Export Debug Files	debug files (svg/ dot) (True/False)
Verbose Level	verbosity level (fatal, error, warning, info, debug, trace)
Output Folder	Path to a folder in which computed matches will be stored

Table 2.4: Meshroom Feature Matching Node

2.2.5 Structure From Motion

The Structure From Motion Node carries out the 3D points reconstruction from the input images. The key behind this node is to be able to comprehend the geometric relationship behind all the views from the input images, and decipher the rigid scene structure (3D PointCloud) along with the pose (position & orientation) and internal calibration of the cameras as mentioned by Cheng et al. (2014). This is an incremental process pipeline which associated with a growing reconstruction process. It initially computes an starting two-view point reconstruction which is iteratively extended by adding new views (Cheng et al., 2014).

It starts by fusing all feature matches between all images and store them in tracks. Each track represents a point in space, which is visible from multiple cameras. Nevertheless, at this stage of the pipeline, many outliers are present, therefore during matches fusion, incorrect tracks are removed (Fischler and Bolles, 1981).

Afterwards, the incremental algorithm chooses the best initial image pair. This choice is very important for the quality of the final reconstruction, hence it should provide robust matches and reliable geometric information as suggested by Moulon et al. (2012). Therefore, this image pair maximises the number of matches and the repartition of features correspondances in all images. On the other hand, the angle between the cameras must be able to provide with reliable geometric information. Then, it computes the fundamental matrix between these two images considering the first image as the origin of the coordinate system (Kneip et al., 2011). As the pose of the first two cameras is known, it is possible to triangulate the matching 2D features into 3D points.

Once this step finishes, the *Next Best Views Selection* algorithm selects all the images that have enough number of associations with the 3D reconstructed features. Based on 2D-3D

2.2.6 Prepare Dense Scene

This node undistorts the images and generates EXR images. The components in the node are described on table 2.6

Name	Description
Input	SfMData file
ImagesFolders	Use images from specific folder(s). Filename should be the same or the image uid.
Output File Type	Output file type for the undistorted images. (jpg, png, tif, exr)
Save Metadata	Save projections and intrinsics information in images metadata (only for .exr images).
Save Matrices Text Files	Save projections and intrinsics information in text files.
Correct images exposure	Apply a correction on images Exposure Value
Verbose Level	[‘fatal’, ‘error’, ‘warning’, ‘info’, ‘debug’, ‘trace’]
Output	MVS Configuration file (desc.Node.internalFolder + ‘mvs.ini’)
Undistorted images	List of undistorted images.

Table 2.6: Meshroom Prepare Dense Scene Node

2.2.7 Depth Map

This step in the pipeline retrieves the depth value of each pixel for all cameras that were processed by the StructureFromMotion Node. It uses Semi-Global Matching method as proposed by Hirschmuller (2005).

For each image, it selects the "N" best & closest cameras around. It chooses fronto-parallel planes based on the intersection of the optical axis with the pixels of the chosen neighboring cameras (Hirschmuller, 2005). This produces a volume $\mathbf{W}, \mathbf{H}, \mathbf{Z}$ with several depth candidates per pixel and it estimates the similarity for all of these. The Similarity is calculated by the *Zero Mean Normalized Cross-Correlation (ZNCC)* of a small patch in the principal image which is reprojected into at the other camera, which creates a volume of similarities as suggested by Strecha et al. (2006). For each neighboring image, it accumulates similarities in this volume. Nevertheless, this volume is noisy, therefore, there is a filter in each step along X & Y axis which reduces the isolated high values. Finally, it selects the local minima and replaces the selected plane index with a depth value stored into a depth map (Scharstein et al., 2001). This depth map has banding artifacts as it is based on the original selection of depth values as mentioned by Scharstein et al. (2001). Thus, a refinement process calculates the depth values with sub-pixel accuracy.

The components in the node are described on table 2.7

These series of modular libraries have a series of embedded algorithms that can be used in scenarios such as: filtering outliers from noisy data, stitch multiple 3D pointclouds, segment relevant parts of the scene & point cloud, extract keypoints and compute image descriptors to be able to recognize real world objects based on geometric appearance, among many other possible implementations.

The selected components and algorithms that were used in this project include:

- Statistical Outlier Removal Filter
- Pass Through Filter
- Voxel Grid Filtering
- Plane Model Segmentation
- KdTree
- Euclidean Cluster Extraction
- Principal Component Analysis
- Moving Least Squares
- Poisson Surface Reconstruction

2.3.1 Statistical Outlier Removal Filter

This component of the PCL library aims to remove noisy measurements (outliers) from a point cloud dataset using statistical analysis techniques (Rusu and Cousins, 2011).

Typically, Laser Scans produce point cloud datasets with various point densities. Similarly, measurement errors generate sparse outliers that can corrupt the results of the dataset. This complicates process of estimating local point cloud characteristics in particular surface normals and curvature changes. Hence, leading to erroneous values, that can create failures in the point cloud registration process. Some of these irregularities in the dataset, can be addressed with a statistical analysis performed on each point's neighborhood, and removing the points that do not meet a specified criteria as mentioned by Rusu and Cousins (2011). The sparse outlier removal is built on the computation of the distribution of point to neighbors distances of the source dataset (Rusu and Cousins, 2011). For each point, it computes the mean distance from the point to all the corresponding neighbors. This process assumes that the resulting distribution is Gaussian with a standard deviation and a mean. Then all points whose mean distances are outside an interval defined by

global distances (mean & standard deviation) can be expressed as an outlier and it can be removed from the dataset as suggested by Rusu and Cousins (2011).

Figure 2.3 illustrates the results of the sparse outlier analysis and removal process. The initial dataset is on the left and the processed result in on the right. Furthermore, the graph on the right, illustrates the mean k-nearest neighbour distances in a point neighborhood, before and after the filtering process (Rusu and Cousins, 2011).

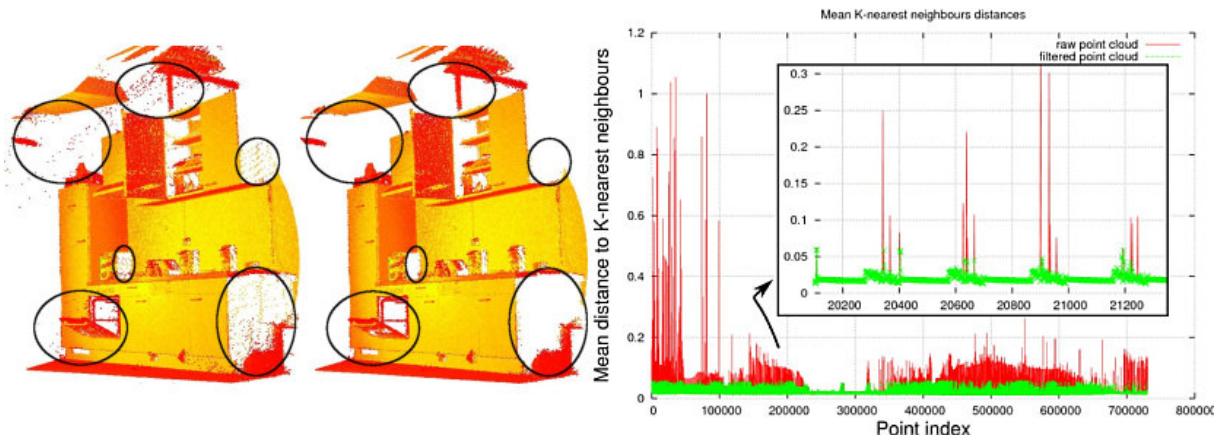


Figure 2.3: Removing outliers using a StatisticalOutlierRemoval filter
(Rusu and Cousins, 2011)

2.3.2 Pass Through Filter

This component of the PCL library has embedded two main components of the framework. The **PassThrough** components uses the base **Filter** class methods to pass through all the data that satisfies certain contrains.

PassThrough passes points in a pointcloud based on constrains for one specific field of the point type (Rusu and Cousins, 2011). It iterates through the entire input pointcloud once, and automatically filters non-finite points, including the points outside the interval specified by *setFilterLimits()*, that only applies uniquely to the configured field *setFilterFieldName()*. The component *setFilterLimits()* sets the numerical limits for the field for data filtering, whereas *setFilterFieldName()* configures the name of the field that will be used for data filtering.

Therefore, this component performs a filter along a given dimension, which essentially removes the values that are either outside or inside the configured range. The image 2.4 illustrates the process. The pointcloud has five points, after filtering the green points represents the remaining end result and the red points are the points that were removed

by the filter.

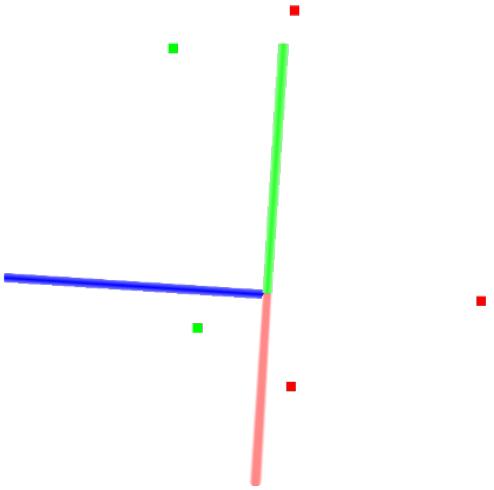


Figure 2.4: Filtering a PointCloud using a PassThrough filter
(Rusu and Cousins, 2011)

2.3.3 Voxel Grid Filtering

A voxel Grid illustrates a value on a regular grid in 3D space. A Voxel is an image of 3D space region which is limited by defined sizes, which has its corresponding nodal point coordinates in an accepted system, own form, own state parameter that demonstrates it belongs to some modelled object and its associated properties of the modelled region (Shchurova, 2015).

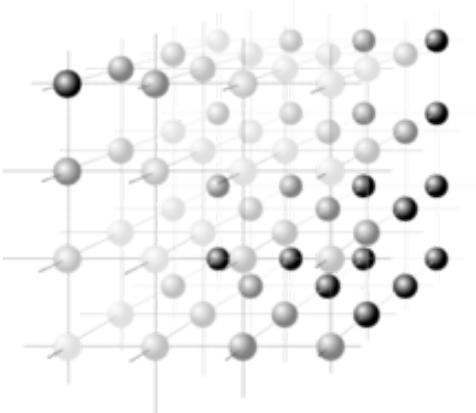


Figure 2.5: Illustration of a voxel grid containing color values
(Shchurova, 2015)

This component of the PCL framework aims to downsample (*Reduce the Number of Points*) of a point cloud dataset using a voxelized grid method. The **VoxelGrid** component,

creates a 3D Voxel grid, which in common terms refers to a set of small boxes in 3D space, over the input point cloud dataset. Once it is successfully converted into a voxel grid, then in each voxel(*3D box*), all the existing points will be approximated (*Downsampled*) with their centroid as suggested by Rusu and Cousins (2011). This approach represents the underlying surface of the point cloud dataset with high accuracy. The main component is to set a defined and Voxel leaf Size, which allows to set the Voxel Size and the Number of Voxel in the Voxel Grid. Therefore, directly influencing the downsampling process and the resulting processed point cloud.

2.3.4 Plane Model Segmentation

This component of the PCL library aims to perform a plane segmentation of a given set of points, which finds all the points within a point cloud data set that support a plane model. One of the components that this component uses is *pcl::ExtractIndices*, which aims to extract the indices from a point cloud. It is a filter that extracts a subset of points from a point cloud dataset, which are related to the indices output of a segmentation algorithm.

Similarly, it uses *pcl::SACSegmentation*, which represents the Nodelet segmentation class for Sample Consensus methods and model, in a way that creates a Nodelet wrapper which can be used for generic-purpose SAC-based Segmentation (Rusu and Cousins, 2011). The *pcl::SACSegmentation*, creates an object and sets the model & method type. Furthermore, it specifies the "distance threshold", that is used to determine the distance for point in the model, in order to be considered an inlier. The used method in this project uses the **RANSAC** method. The main advantage of RANSAC is its simplicity and robustness.

The estimated plane parameters are estimated with equation 2.2.1, where **a,b,c,d** are the model coefficient values.

$$ax + by + cz + d = 0 \quad (2.1)$$

Figure 2.6, illustrates the process below. The red points are the outliers, whereas the green points are the inliers of the found plane model.

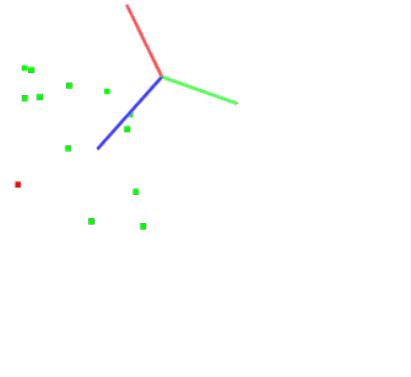


Figure 2.6: Illustration of Plane model segmentation
(Shchurova, 2015)

RANSAC

Random sample consensus (RANSAC), is an iterative method that is used to estimate parameters of a mathematical model from a set of observed data that contains outliers (Strutz, 2016). Hence, it can be interpreted as an outlier detection method. It is a non-deterministic algorithm that generates an adequate result with a specific probability, the probability increases as more iteration are performed.

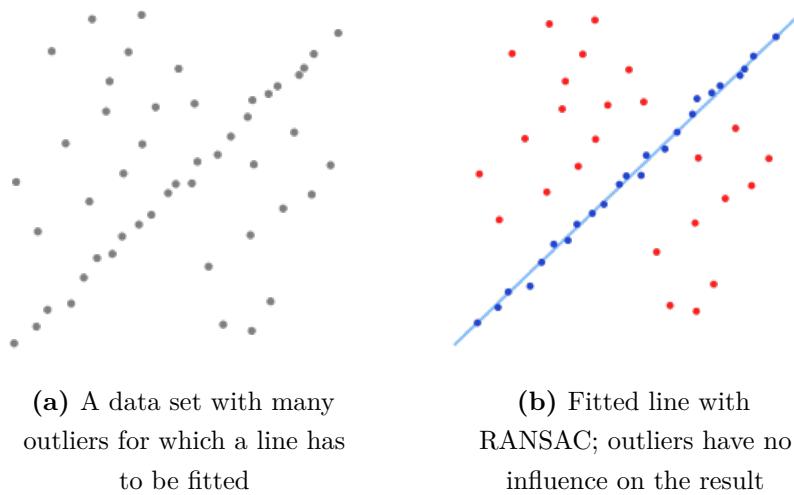


Figure 2.7: Dataset Before & After RANSAC algorithm (Strutz, 2016)

2.3.5 KdTree

In this project Kdtrees are used to find the K nearest neighbour of certain points or locations and how to find these within an specified radius.

A K-D tree (*k-dimensional tree*), is a data structure that is commonly used to organize a certain number of points in a space with k dimensions as suggested by Rusu and Cousins (2011). It is a binary search tree with other constraints imposed on it and are useful for searches for range and closest neighbours. As this project uses pointclouds, the Kdtrees would have three dimensions. Each level of a Kdtree splits all children along a specific dimension, utilising a hyperplane which is perpendicular to the corresponding axis. At the root of the KdTree, all children will be divided based on the first dimension, and each level down in the Kdtree divides on the next dimension, finally it returns to the first dimension once all other have been exhausted (Rusu and Cousins, 2011).

KdTrees are used extensively in many components of the project framework pipeline.

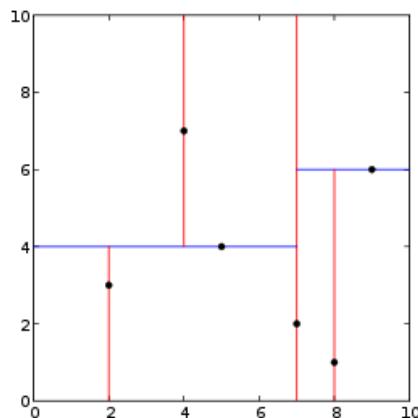


Figure 2.8: Example of a 2-dimensional k-d tree
(Rusu and Cousins, 2011)

2.3.6 Euclidean Cluster Extraction

The clustering method allows to divide a non-organized pointcloud dataset P into smaller components, in order to reduce the overall processing time for P . Point Cloud Data Clustering can be approached in an Euclidean manner and be implemented using 3D grid subdivisions of the space with fixed width boxes or a octree data structure. In this project, Kdtrees are used to find the nearest neighbours and implements a clustering technique which is similar to a flood fill algorithm (Rusu and Cousins, 2011).

The following Steps illustrates the clustering algorithm in this project with a Kdtree:

1. create a Kd-tree representation for the input point cloud dataset P ;
2. set up an empty list of *clusters* C , and a queue of the points that need to be checked

Q ;

3. then for every point $p_i \in P$, perform the following steps:
 - add p_i to the current queue Q ;
 - for every point $p_i \in Q$ perform:
 - Search for the set P_i^k of point neighbors of p_i in a sphere with radius $r < d_{th}$
 - for every neighbor $p_i^k \in P_i^k$, check if the point has already been processed, and if not add it to Q
 - when the list of all points in Q has been processed, add Q to the list of clusters C , and reset Q to an empty list
4. the algorithm terminates when all points $p_i^k \in P$ have been processed and are now part of the list of point clusters C

Figure 2.9, illustrates a cluustering output of a dataset, where different components were divided into multiple clusters

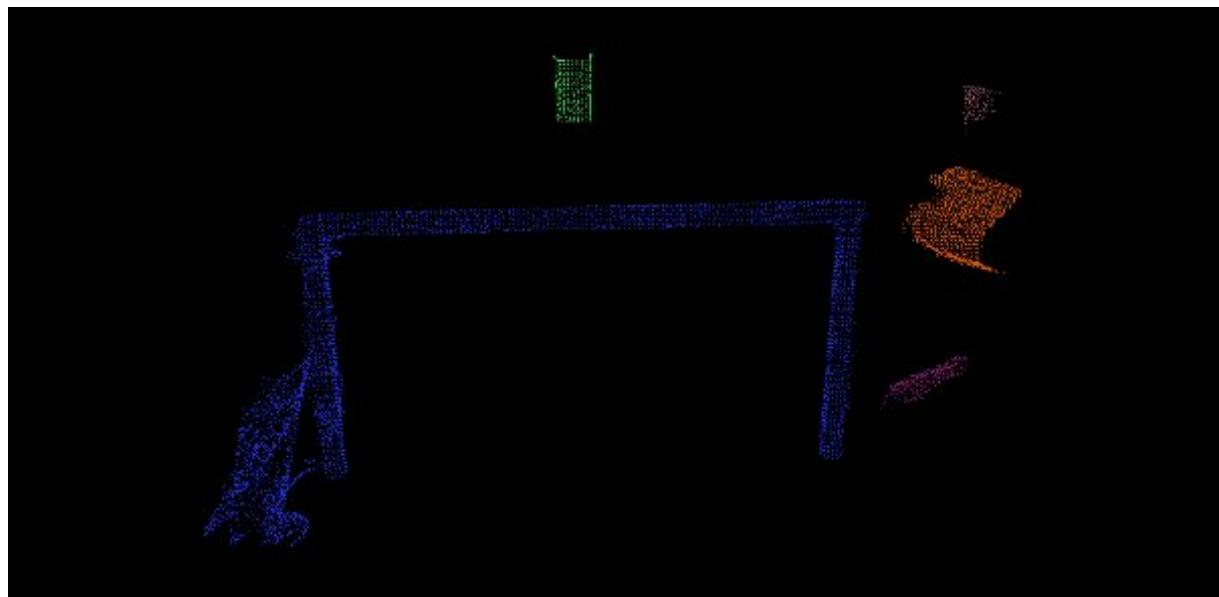


Figure 2.9: Example of point clouud cluster
(Rusu and Cousins, 2011)

2.3.7 Principal Component Analysis

The Principal Component Analysis (PCA) is the process of computing the principal components and use them to change the basis on data, using only the first few principal components and discarding the rest. The principal components is a sequence of p unit vectors of a collection of points in a real coordinate space (Artac et al., 2002).

The principal components are extracted using a singular value decomposition method, which is applied on the covariance matrix of the centered input point cloud dataset. Once the PCA analysis is performed using the PCL library, it is possible to calculate the following components:

- Mean of input data
- Eigen Vectors: Ordered set of vectors that represents the final principal components and the eigen cartesian space.
- Eigen Values: These are the correspondent loading of the Eigen Vectors in a descending order.

The Principal Component Analysis is used to determine the scale of the Two point clouds that are used in this project and properly adjust the final mesh.

2.3.8 Moving Least Squares

Moving Least Squares (MLS) is a method of reconstructing continuous functions from a set of unorganized point samples through the computation of weighted least squares, biased towards the region around the point at which the reconstructed value is requested (Levin, 1998).

MLS can be modelled considering function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and a set of sample points $S = \{(x_i; f_i) | f(x_i) = f_i\}$. Then, its MLS approximation of degree m at the point x is $\tilde{p}(x)$ where \tilde{p} reduces the weighted least square error $\sum_{i \in I} (p(x_i) - f_i)^2 \theta(\|x - x_i\|)$ over all polynomials p of degree m in \mathbb{R}^n as mentioned by Levin (1998).

The main objective of Moving Least Squares (MLS) is to smooth and resample noisy data in a surface reconstruction object. Some pointclouds have certain data irregularities, which can be caused by small errors in the distance measurements, and are complex to remove using conventional statistical analysis. It is important to account for complex surfaces and occlusions to create a complete model. Resampling algorithms such as MLS can be used to recreate the missing parts of the surface by a high order polynomial interpolation

between the surroundings data points. Resampling, allows to correct the small errors and smooth the surface of a pointcloud dataset.

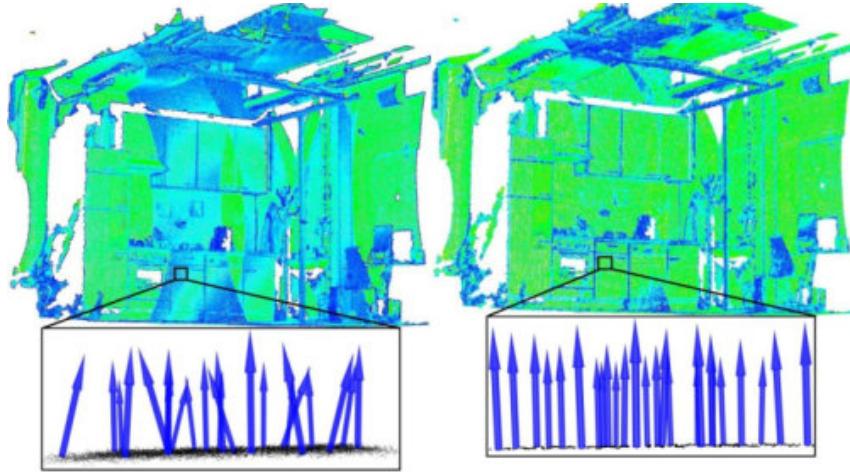


Figure 2.10: Moving Least Square Smoothing
(Rusu and Cousins, 2011)

On figure 2.10, on the left side, it is illustrated the effect of estimating the normals of a pointcloud, however, due to alignment errors the normals are noisy. On the right side of figure 2.10, the effects of the MLS algorithm is demonstrated as it smoothes the surface of the pointcloud.

In order to approximate, the surface illustrated by the local neighbours of points p_1, p_2, \dots, p_k at a point q , it uses a bivariate polynomial function, which is defined on a robust modelled reference plane. Figure 2.11, shows the curvatures at each point with the eigen value relation before and after the resampling method.

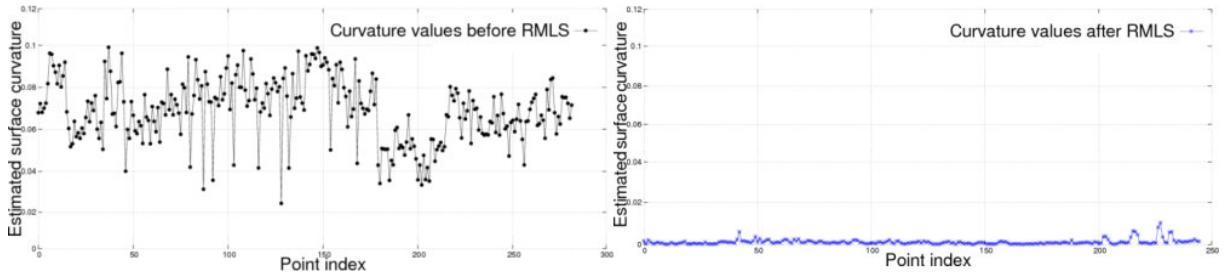


Figure 2.11: Curvatures of MLS before & after
(Rusu and Cousins, 2011)

Furthermore, the Moving Least Squares Component of the PCL library also allows for different upsample methods. These include

DISTINCT CLOUD Project the points of the distinct cloud to the MLS surface(Rusu and Cousins, 2011).

SAMPLE LOCAL PLANE The local plane of each input point will be sampled in a circular fashion using the *UpsamplingRadius* and the *UpsamplingStep* parameters(Rusu and Cousins, 2011).

RANDOM UNIFORM DENSITY The local plane of each input point will be sampled using an uniform random distribution such that the density of points is constant throughout the cloud - given by the *DesiredNumPointsinRadiusParameter*(Rusu and Cousins, 2011).

VOXEL GRID DILATION The input cloud will be inserted into a voxel grid with voxels of size *VoxelSize*.This voxel grid will be dilated (*DilationIterationNum*) times and the resulting points will be projected to the MLS surface of the closest point in the input cloud; the result is a point cloud with filled holes and a constant point density (Rusu and Cousins, 2011).

For this project, the **RANDOM UNIFORM DENSITY** method is used to upsample the MLS processed pointcloud. This method takes the parameters for a desired point cloud density within a fixed radius neighborhood as suggested by Ichim (2012). For each point, based on the density of the neighbors it will add more points on the local plane using a ramdom number generator. The random number generator will have a uniform distrubution and it will stop once the desired density is achived. Then it will replay the MLS algorithm in order to smooth the surface again.

Figure 2.12 illustrates on the left the implementation of MLS with **UNIFORM DENSITY** upsampling method on a pointcloud, whereas on the right if the raw pointcloud.

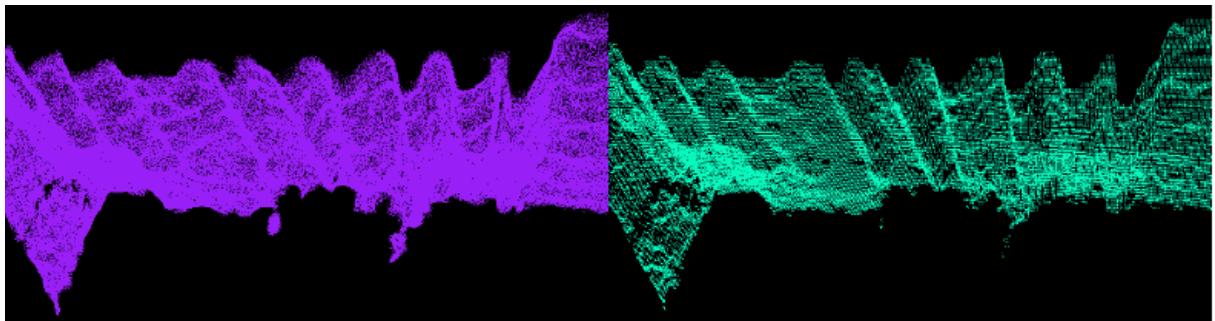


Figure 2.12: Uniform Density Upsample Method After & original
(Ichim, 2012)

2.3.9 Poisson Surface Reconstruction

In this project the Poisson Reconstruction algorithm is implemented , to reconstruct the processed point cloud. The Poisson Reconstruction algorithm demonstrates that the surface reconstruction from a set of oriented points can be associated as a spatial Poisson problem. This formulation, utilizes all the points at once, with no need of using heuristic spatial partitioning or blending, hence, making it resilient to noise as mentioned by Kazhdan et al. (2006).

The main objective of this algorithms to reconstruct a smooth surface, which is based on a large number of points p_i from a pointcloud, where each point possesses an estimate of the local surface normal n_i . It aims to create an implicit function f , whose value is zero at the points p_i and the gradient at points p_i is equal to the normal vector n_i . Kazhdan et al. (2006). The set of (p_i, n_i) is modelled as a continuous vector field V and the implicit function is found by integrating the vector field V . In complex calculations, it is possible to perform a *least-squares fit* to minimize the difference between the gradient of f and V . Kazhdan et al. (2006).

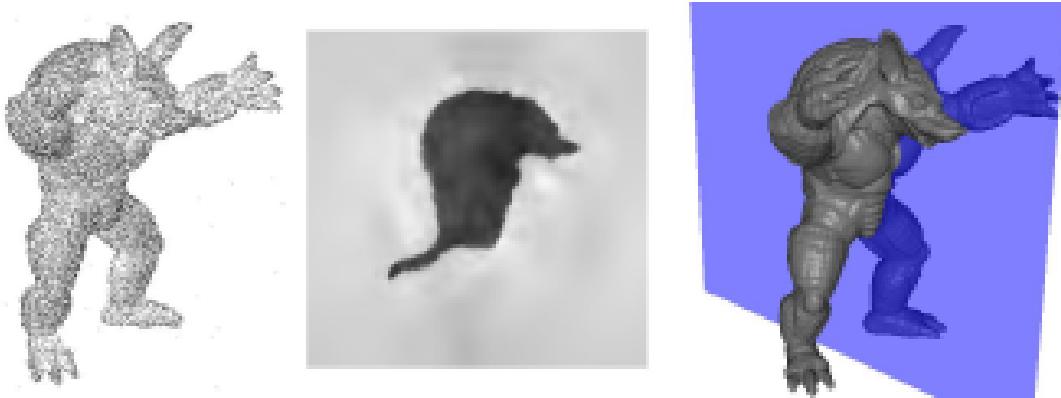


Figure 2.13: Poisson Reconstruction Example. On the left it is input pointcloud dataset, whereas on the right is output of the Poisson Surface Reconstruction Algorithm

(Kazhdan et al., 2006)

2.4 Open3D

Open3D is an open source library that supports fast development of software that handles 3D data. The Open3D frontend exposes outputs a set of carefully chosen data structures and algorithms in both Python and C++ (Zhou et al., 2018). The backend of the framework

is highly optimized and it configured for parallelization. Open3D is compatible with Linux, macOS and Windows and it can be installed via source or via packages.

The core features include:

- Simple installation via conda and pip
- 3D data structures
- 3D data processing algorithms
- Scene reconstruction
- Surface alignment
- PBR rendering
- 3D visualization
- Python binding

In this project Open3D was used as an auxiliary library used as contingency for Poisson Reconstruction. As in many circumstances the normals of a Pointcloud, might not be properly oriented *orient normals consistent tangent plane* allows to propagate the normal orientation with a minimum spanning tree as suggested by Zhou et al. (2018). After the normals is aligned it will execute the Poisson Surface Reconstruction algorithm (Kazhdan et al., 2006) and solves a regularized optimization problem to obtain a smooth surface mesh.



Figure 2.14: Open3D Poisson Reconstruction Example
(Zhou et al., 2018)

Methodology

As mentioned before, this project is associated with the software reconstruction of a 3D scanner device. The key components are divided into three processes that include:

- Data collection
- Photogrammetry Process
- Point Cloud Processing & Reconstruction

All the components mentioned above will work in a Pipeline manner and as the key process for a 3D scanner includes the Data collection. Once the Data is collected, a photogrammetry process will run in order to reconstruct the object based on the input data. Finally, the resultant reconstructed object will be processed using a Custom framework pipeline based on the PCL (*Point Cloud Library*) and produce a processed mesh. Once the mesh has finished processing it can be used in different applications such as clothes fitting, etc.

3.1 Data Collection

The data Collection process is performed with a 3D scanning RIG (device). The process will be performed with a series of twenty-eight cameras and a single Lidar. The cameras will be located and assembled in a series of eight tall poles and 4 short poles. Similarly, the Lidar will be mounted to a single Pole.

All the cameras will be controlled using Raspberry Pis. There will be a total of three Raspberry Pi's that set the execution process to acquire data from all the cameras that are attached to them, with a total number of twenty eight images. Similarly, an Nvidia Jetson Xavier is used to acquire data from the Lidar. The Nvidia Jetson Xavier will acquire a single image and a PointCloud of the scanned subject.

Cameras

The cameras used for data acquisition have the following specifications.

Model No.	HBV-1825
Model Size	62mm × 9mm × 5.68mm±0.2mm
Active Array Size	2592 x 1944
Pixel Size	1.4 μ m x1.4 μ m
shutter	rolling shutter / frame exposure
Field of View	65°

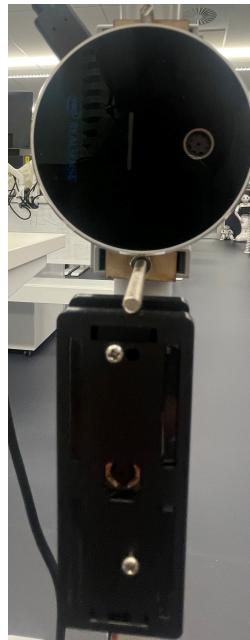
Table 3.1 Cameras Specification

As observed in figure 3.1, the cameras will have a Plastic Housing. This Housing device allows to safely place the cameras in the poles as well as protecting the electronics and connection from external factors. The cameras will be connected to a series of USB hubs that will be directly connected to the Raspberry Pi's to control the data acquisition process.



Figure 3.1: Tall Pole Cameras Location With Housing.

Lidar



The Lidar used for the data acquisition process is an Intel RealSense L515. The Intel RealSense will be located in tall pole and it will be attached in the middle section in order to maximise the field of view. It will be locked above the middle camera of a tall pole. As the Lidar has both an RGB sensor and a Depth Camera, it will be used to capture an RGB image as well as a Pointcloud with the Depth and RGB Sensor. The Lidar will be connected and controlled with a Nvidia Jetson Xavier. It will use the RealSense SDK to capture both RGB image and Point.

Table 3.2 states the specification of the Intel RealSense L515.

Figure 3.2: Lidar Mounted in Pole

Model No.	Intel RealSense L515
Depth Technology	Lidar
Ideal Range	0.25 m to 9 m
Depth Field of View (FOV):	$70^\circ \times 55^\circ (\pm 3^\circ)$
Depth output resolution:	Up to 1024×768
Depth Accuracy:	~ 5 mm to ~ 14 mm thru 9 m
RGB frame resolution	1920×1080
RGB sensor technology	Rolling Shutter
RGB sensor FOV (H \times V)	$70^\circ \times 43^\circ (\pm 3^\circ)$
RGB sensor resolution	2 MP

Table 3.2: Intel RealSense L515 Specification

3.1.1 Layout & Configuration

As mentioned before Lidar will be controlled to an Nvidia Jetson Xavier, which will make it agnostic of the Cameras acquisition process. It was previously mentioned that the Cameras will be controlled with a Series of multiple Raspberry Pi's and those will be managed via a SSH connection with Putty. Due to the limited number of USB connectors, a series of USB hubs will be used in order to control more cameras with Single Raspberry Pi's. Figure 3.3, illustrates how the twenty-eight cameras will be connected with the Different Raspberry Pi's.

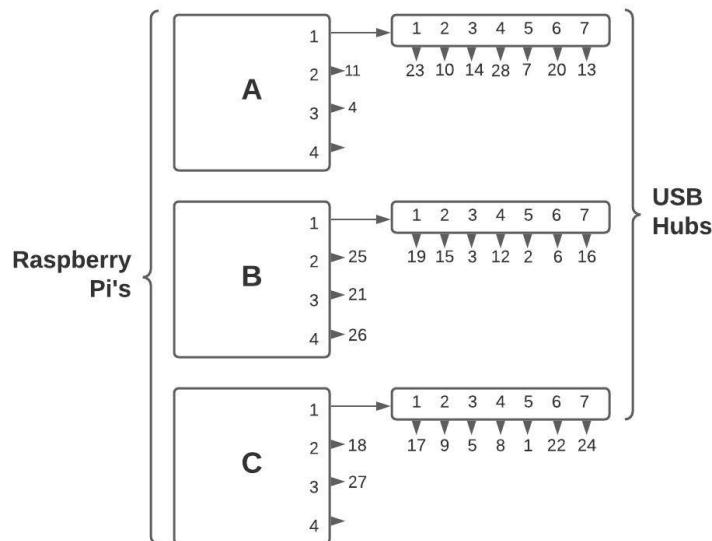


Figure 3.3: Cameras Lay-out connection

All the cameras will be mounted in the RIG in a series of Pole. There will be a total number of eight tall and four short poles. The Tall Poles will be used to attached three cameras to the Frame. The First Camera will be located on top of the pole and oriented toward the subject in order to have the best overlap with the Field of View. The Second Camera is located in the middle of the pole and parallel oriented to the scanned subject in order to have the best field of view. The third Camera is located at the bottom of the pole and it will oriented upwards toward the scanned subject in order to have the a wide coverage. This configuration can be observed on figure 3.4.

On the other side, there will be four short poles that each will hold a single camera. The aim of these poles is to gain close data as well as acquiring images from regions of interest around the body in order to increases the accuracy of the final scan.

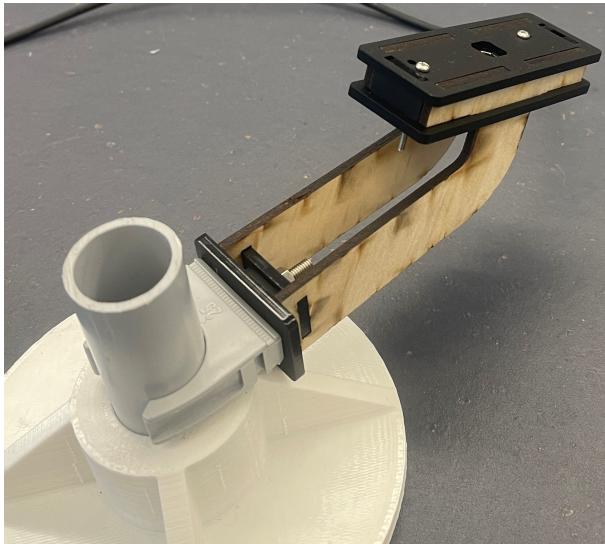


Figure 3.5 Open3D Poisson Reconstruction Example

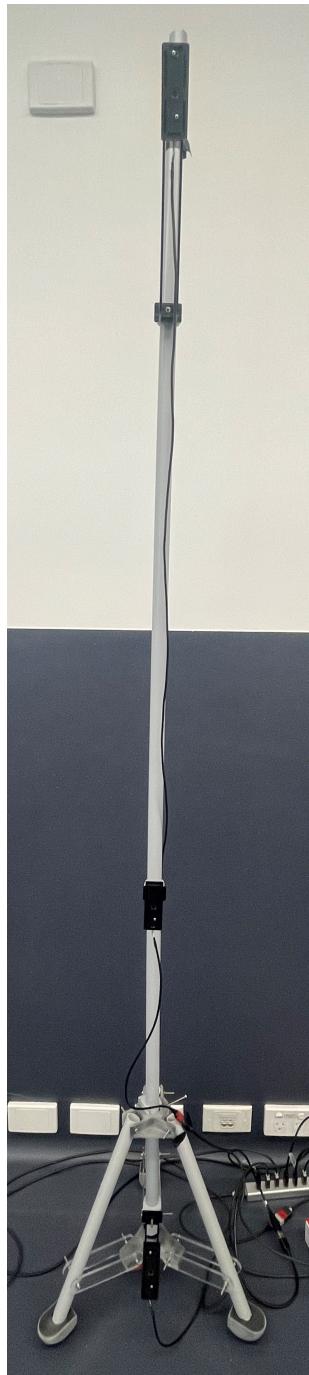


Figure 3.4: Tall Pole Cameras Configuration.

All the tall poles will be placed in a circular shape in order to capture as much data as possible. The subject (person) that will use the RIG will be located in the origin of the circle and all the tall poles will be places at a 1.5 m away from the person in order to have the best overlap between all cameras. The four short poles will be located closeup from the person that will be scanned. The four poles will be placed at 60 cm away from the person (circle origin) in order to capture close up data for complex regions and improve the overall reconstruction.



(a) RIG From View



(b) RIG lateral View

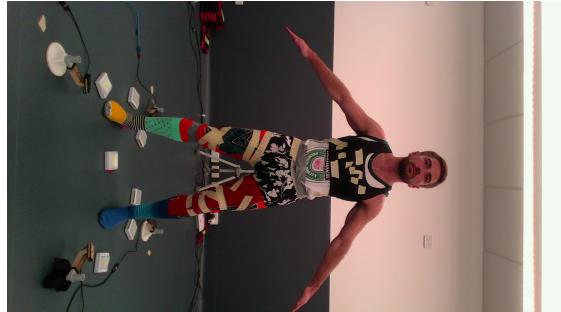
Figure 3.6: RIG scanning With Poles Locations

Execution

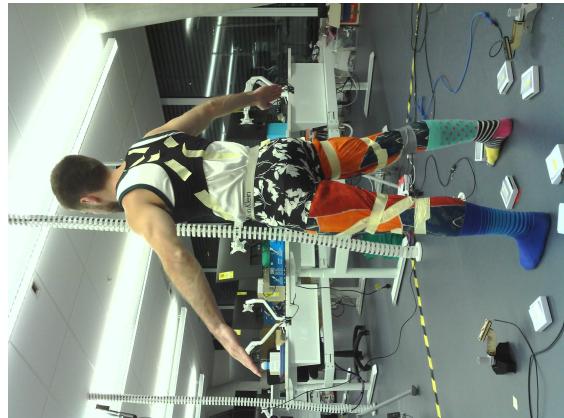
The execution process starts by connecting the user computer to the Raspberry Pi's with a SSH connection via Putty. Once a connection is established, a trial run will be performed to ensure the cameras are working as intended. After the trial run is successful, the RIG is fully operational.

Before a person enters the scan , the user is requirited to start the Nvidia Jetson Xavier and execute the Realsense SDK. Once the mentioned above are perform, a person can enter the RIG to get scanned. In parallel the administrator of the RIG will execute the script to capture data using putty. Once that script is executing, the administrator will take a Photo and PointCloud using the RealSense SDK.

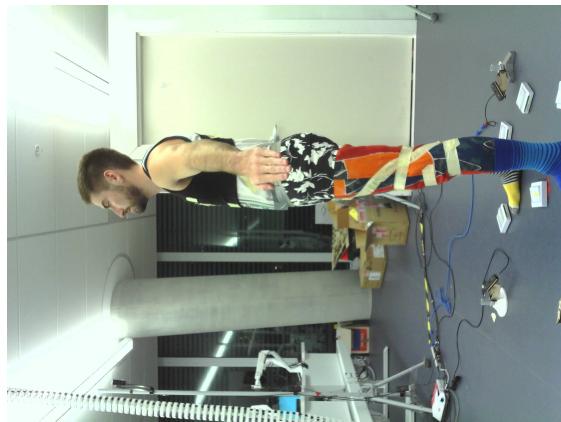
The entire process takes approximately ten seconds. Once the data is captured, it can be used in the photogrammetry pipeline for reconstruction.



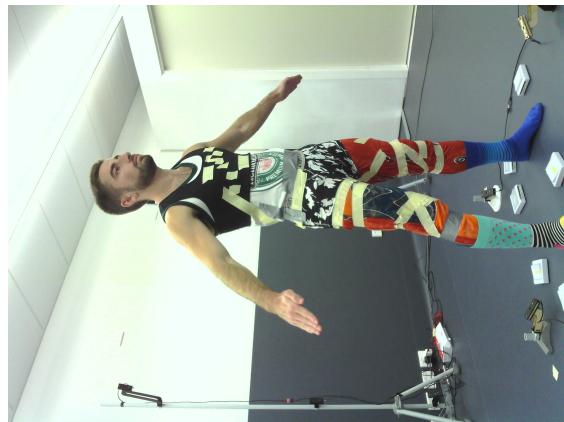
(a) Front View Lidar



(b) Camera Lateral View



(c) Camera Side View



(d) Camera Lateral View

Figure 3.7: Example from Captured Dataset

3.2 Photogrammetry Reconstruction

As mentioned before, Meshroom will be used to process the images and reconstruct the scanned person, once the data acquisition process has finalised. Meshroom will have a series of nodes as noted on section 2.2. It can be used in either windows or Linux. In this project both Meshroom version (2020 & 2021) will tested and fully work. The chosen platform was Ubuntu 18.04. It has a CUDA enabled GPU in order to speed up the execution speed in certain nodes.

After a series of multiple iterations & experimentation, the best parameters were selected and it will be explained in section 3.2.

As Meshroom, has a series of nodes that work in a Pipeline manner, all of the corresponding setting for the best results are detailed below.

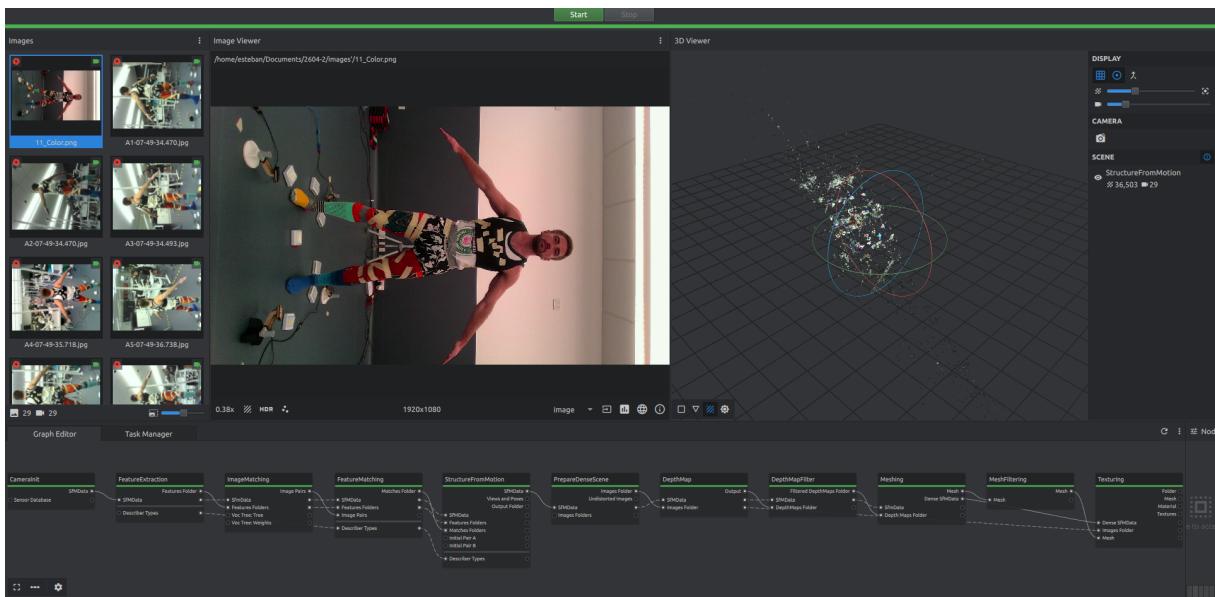


Figure 3.8: Successful Meshroom Pipeline (2021 version)

CameraInit Node

This is the Initial node of Meshroom Pipeline Reconstruction. In this node all the captured data (images) will be imported into. Once all the images have been imported into the initial node. All the available fields are illustrated on table 2.1. Nevertheless in this project, the following fields were configured, whereas the remaining stayed as default and are referenced on figure 3.9. These fields are described below

ViewPoints : In this field is neccesary to edit the **Id**, **Pose Id**, **Intrinsic**.

For each image (viewpoint), the *Id*, *Pose Id*, *Intrinsic* have to be unique in order to properly assign the camera calibration parameters from table 3.3. All of these fields will be direclty mapped with the *Intrinsics Field*. This process will have to be perform for all twenty-nine images for this project.

Intrinsics : This Field is where all the camera calibration parameters from table 3.3 will be associated to all the configure ViewPoints (Images) from the previous field.

In section, it is crucial to map the *ViewPoint Id* from the previous section, to the correct *Intrinsics Id* for this field. Similarly, it is neccesary to select the *Camera Type* as **Radial3**, as this would allow to input the Focal Lenght, Principal Point and Distorsion Parameters. Furthermore, the Width & Heigh of the Image is required to be imported. All these parameters are located on table 3.3.

On the other hand, the *Initialization Mode* has to be configures as **calibrated** and the *Locked Box* requires to be checked as the Cameras (viewpoints) have been previously fully calibrated. This process will have to be perform for all twenty-nine images for this project.

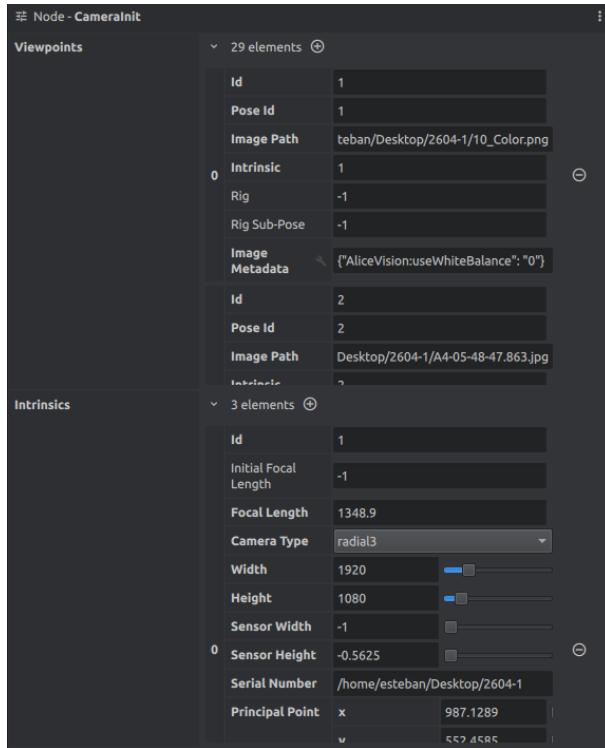


Figure 3.9: CameraInit Node Settings

FeatureExtraction Node

This will be the Second Node of Reconstruction Pipeline. All the field and settings of this node are located on table 2.2. For best results all the settings were configured as below:

Descriptor Types : The used descriptors are **SIFT** and **AKAZE**. After extensive testing it was found that a combination of *Sift* & *Akaze* will extract the most feature and enhance the quality of the reconstruction.

Descriptor Density : This field was configured as **Ultra** due to the small dataset (twenty-nine images)

Descriptor Quality : This was configured to **Ultra**, in order to retrieve the most amount of features from the images.

Contrast Filtering : Configured as **GridSort**

Grid Filtering : Checked field.

Force CPU Extraction : Unchecked Field. The reason behind this was to use CUDA in order to use the GPU to extract all the features to improve the time of processing.

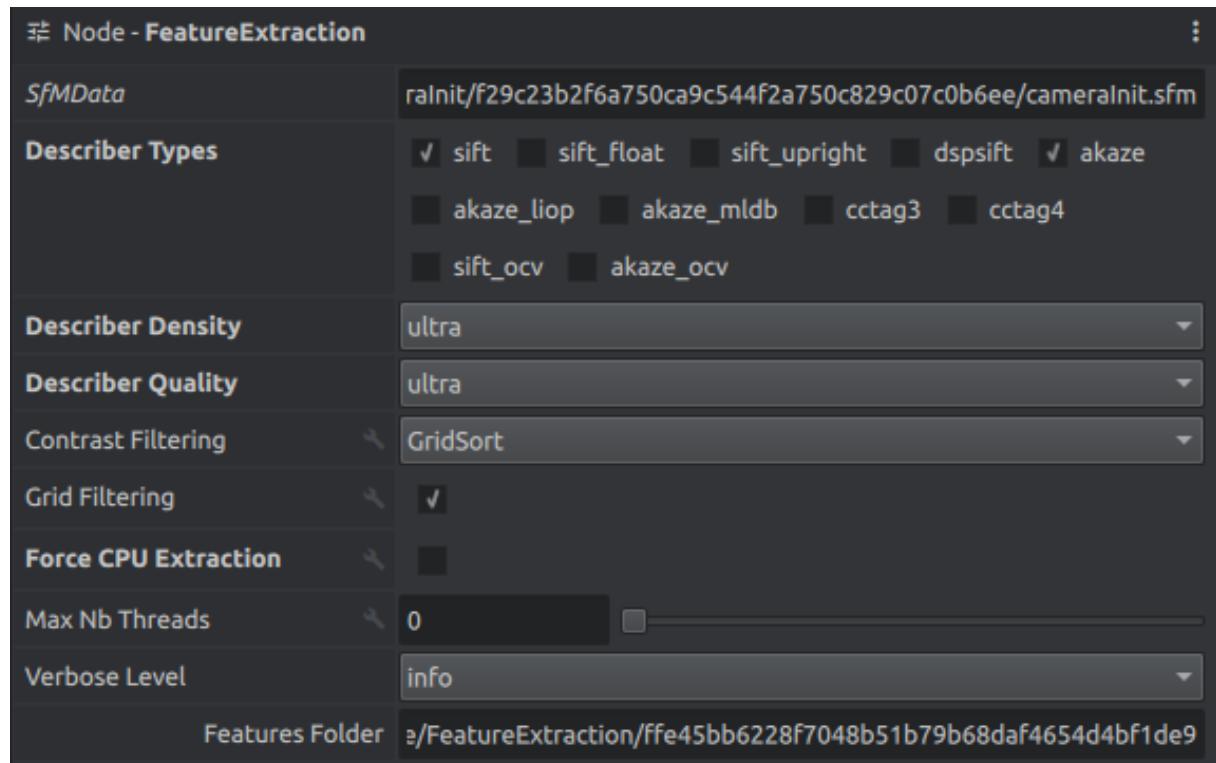


Figure 3.10: FeatureExtraction Node Settings

ImageMatching Node

This is the thid node of the Pipeline Reconstruction. All the fields & settings for this node can be seen on table 2.3. The configured fields for best results are found below:

Method : The selected method for best results is **Exhaustive**. This method will export all image pairs for image matching process.

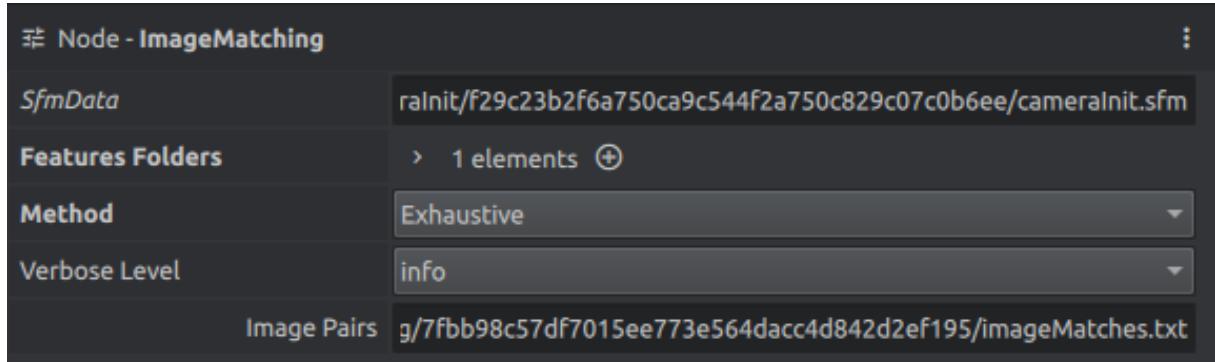


Figure 3.11: ImageMatchingNode Node Settings

FeatureMatching Node

This is the fourth node of the Reconstruction Pipeline. All the settings for this node can be refered to table 2.4. The configured settings for best results are stated below.

Descriptor Types : The Selected Desciber have to match the selected on node the *FeatureExtraction Node*. Therefore the selected descriptors will **SIFT & AKAZE**

Photometric Matching Method : The selected method is **ANN_L2** (*Approximate Nearest Neighbor Matching*).

Geometric Estimator : The selected Estimator **acransac** *A-Contrario RANSAC*.

Geometric Filter Type : The selected validation method to filter feature matches is **fundamental_matrix**.

Distance Ratio : The selected distance ratio to discard non-meaningful matches is 0.8.

Max Iteration : The selectd number of iteration for RANSAC step is 5000.

Known Poses Geometric Error : The maximum error specified for feature matching guided by information for cameras poses was selected as 5

All the other settings were configured as default and are referenced on figure 3.12.

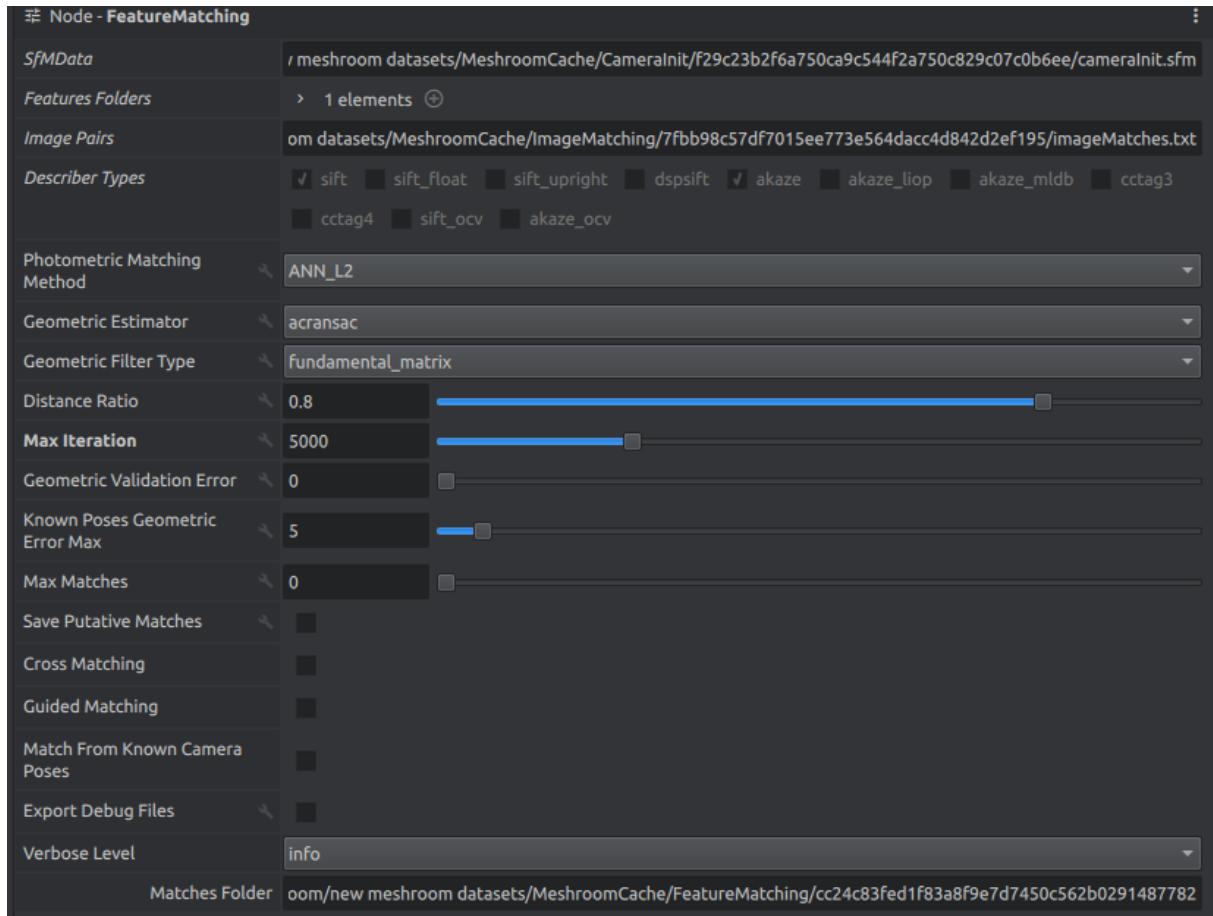


Figure 3.12: FeatureMatching Node Settings

StructureFromMotion Node

The Structure for Motion is the fifth node of the Reconstruction Pipeline. All the settings for this node can be referred to table 2.5. In this Step it is possible to visualise the Structure along with the valid Viewpoints (images) and number of feature matches. Furthermore it is possible to observe the number of cameras that match the valid viewpoints along with its corresponding Pose. As observed in figure 3.13, there are a total of 27883 feature matches and 29 valid viewpoints. Therefore, all the input data was valid and the initial reconstruction process started.

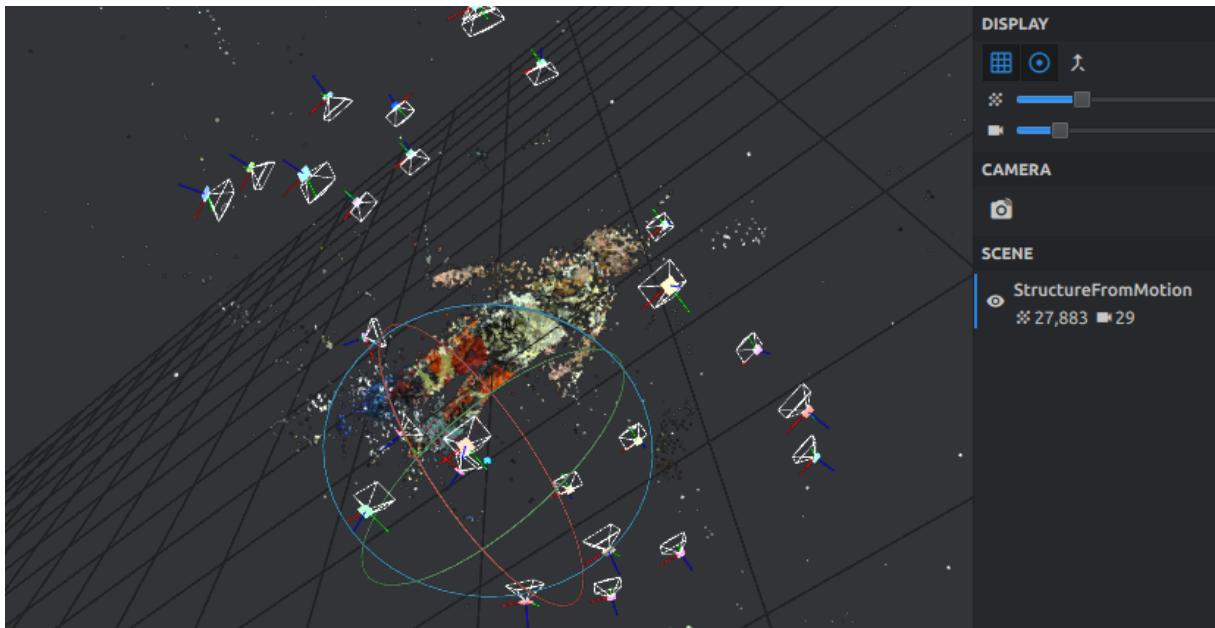


Figure 3.13: Structure From Motion Output

The configured settings for best results are stated below.

Descriptor Type : The Descriptor Type has to match the ones used in previous steps. Hence, **Sift & Akaze** were selected.

Localizer Estimation : The estimator selected to localise the cameras poses is **acransac**.

Observation Constraint : The Selected Method for Observation Constrain Optimization is **Basic** .

Localizer Max Ransac Iterations : For an Optimal Results the selected number of iterations is 15000

LocalBA Graph Distance : The selected Graph-Distance Limit is 1

Min Input Track Length : Configured as 2

Min Observation For Triangulation : Configured as 2

Min Angle For Triangulation : Configured as 1

Min Angle For LandMark : Configured as 1

Max Reprojection Error : Configured as 4

Min Angle Initial Pair : Configured as 1

Min Angle Initial Pair : Configured as 40

Use Rig Constant : This field is selected.

Force Lock of All Intrinsic Camera Parameters : This field is selected in order to force all the cameras intrinsics, as the cameras have been fully calibrated.

All the other fields were configured as default and are referenced on figure 3.14.

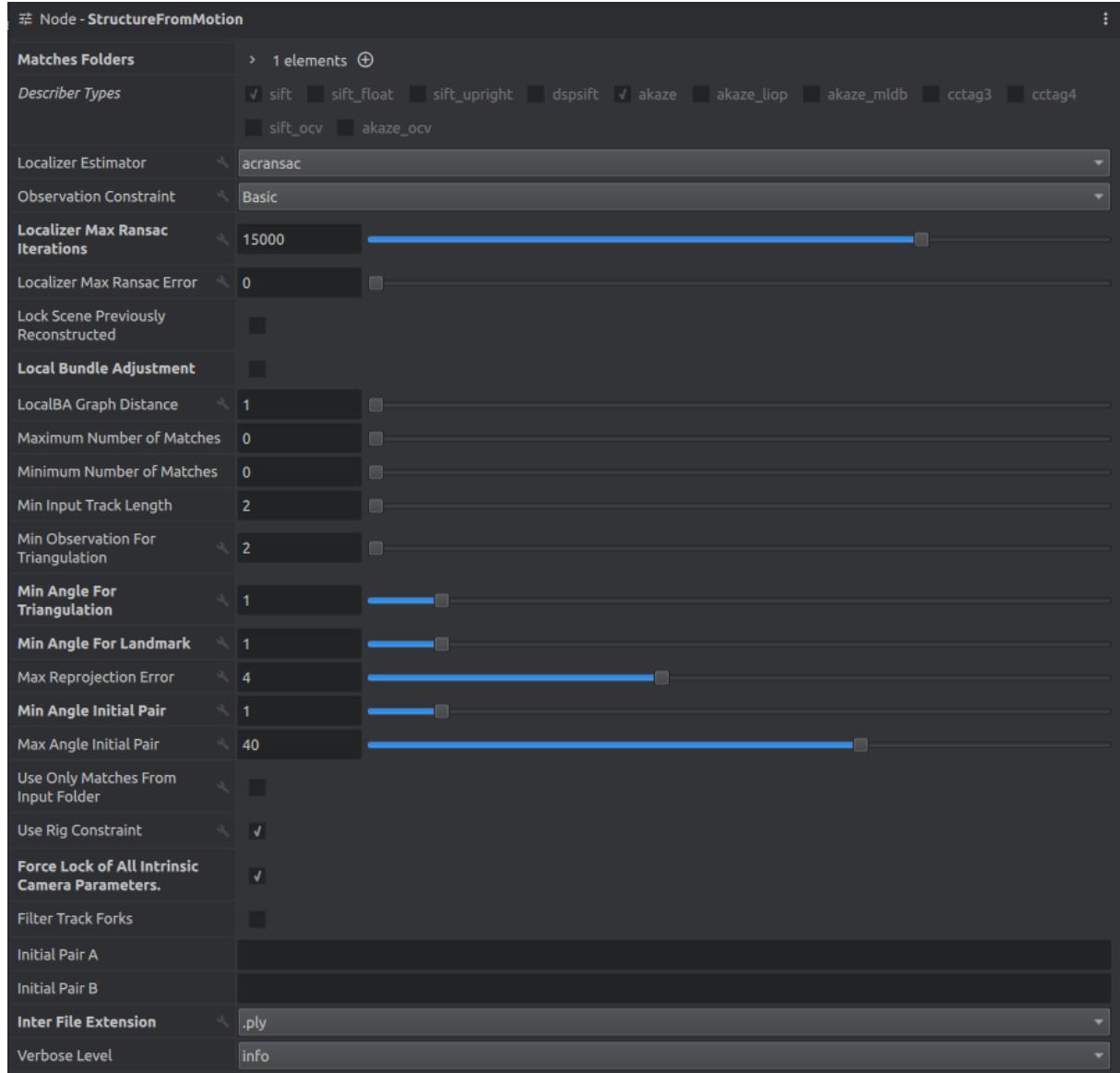


Figure 3.14: Structure From Motion Settings

PrepareDenseScene Node

The PrepareDenseScene is the sixth node of the Reconstruction. All the settings for this node can be referred to table 2.6. The configured settings for best results are stated below.

Output File Type : The chosen File Type for undistorted images is **PNG**.

All the remaining parameters stayed as default and are referenced on figure 3.15.

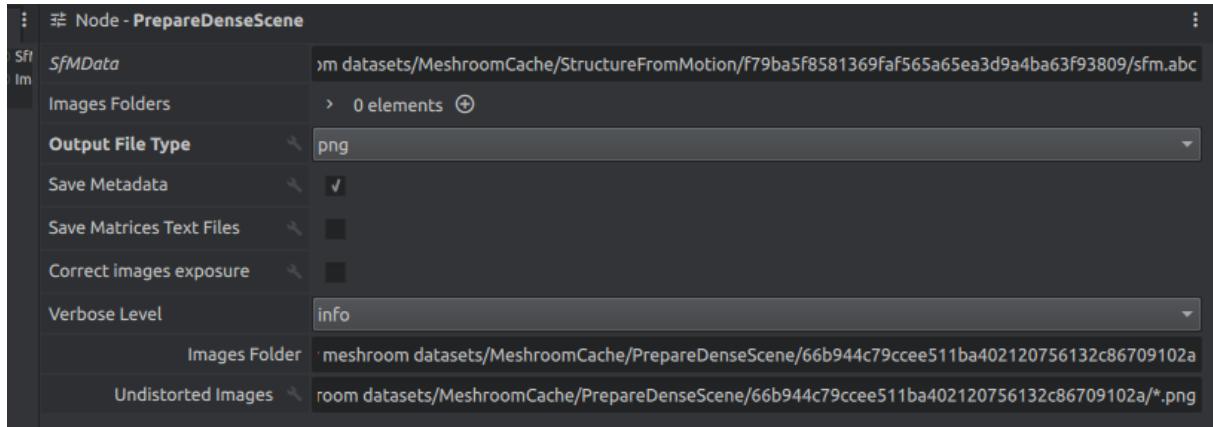


Figure 3.15: PrepareDenseScene Settings

DepthMap Node

DepthMap Node is the Seventh Node of Meshroom Reconstruction Pipeline. All the settings for this node can be referred on table 2.7. The configured settings for best results are stated below.

DownScale : The Downscale Image factor is set to 1 for the best results.

SGM: Nb Neighbour Cameras : The Semi Global Matching Number of Neighbour Cameras selected is 10

Refine: Nb Neighbour Cameras : The selected Refine Number of Neighbour Cameras is 6.

All the remaining parameters stayed as default and are referenced on figure 3.16.

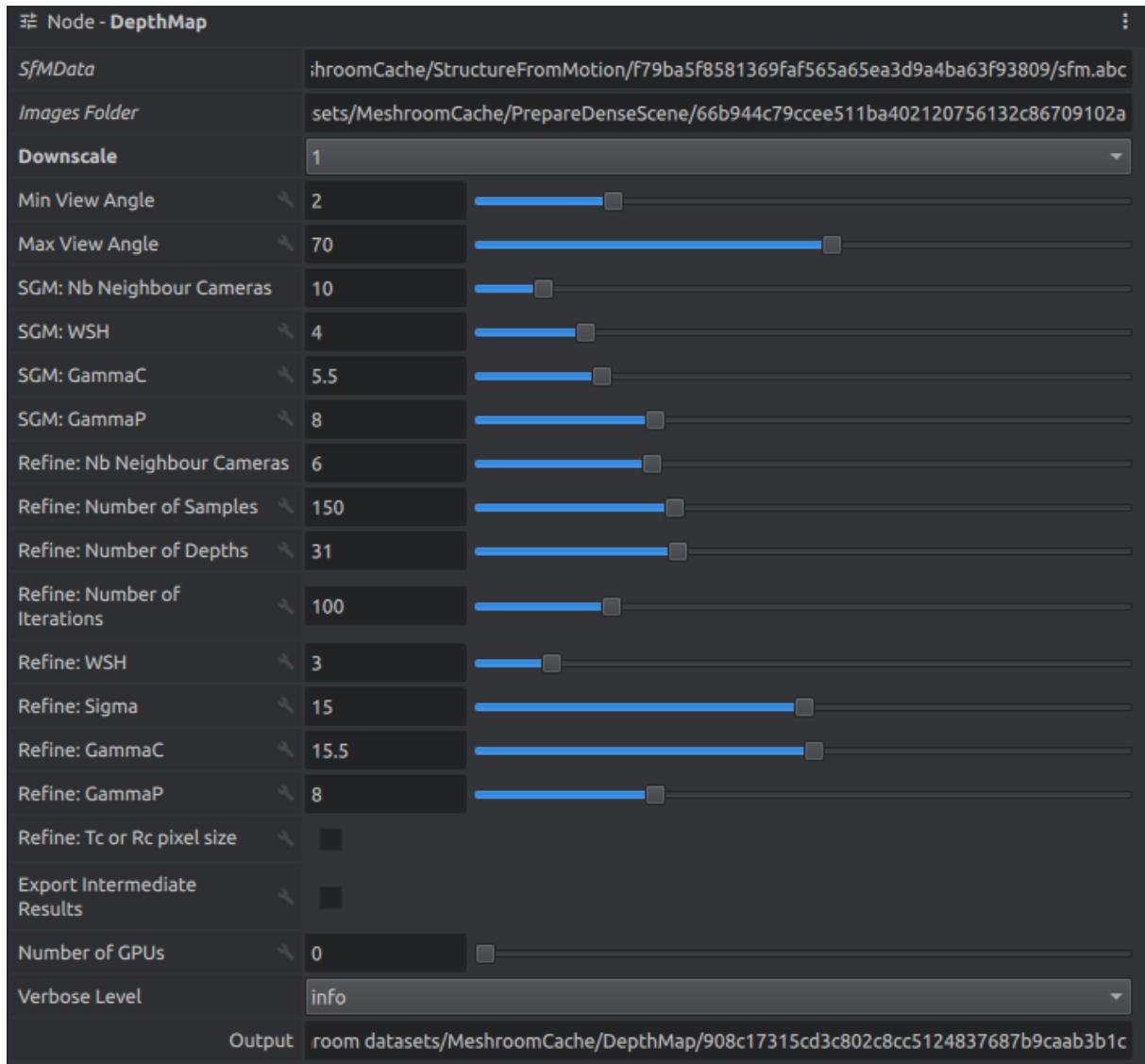


Figure 3.16: DepthMap Node Settings

DepthMapFilter Node

DepthMap Node is the Eighth Node of the Meshroom Reconstruction Pipeline. All the settings for this node can be referred on table 2.8. The configured settings for the best results are stated below.

Min View Angle : The selected Minimum angle between two views is 2.

Max View Angle : The selected Maximum angle between two views is 70.

Number of Nearest Cameras : The selected Number of nearest cameras used for filtering is 10.

Min Consistent Cameras : The selected Number of minimum constant cameras is 3.

Min Consistent Cameras Bad Similarity : The Selected minum Number of consistent cameras with weak similarity value is 4

All the remaining settings are configured as default and are referenced on figure 3.17.

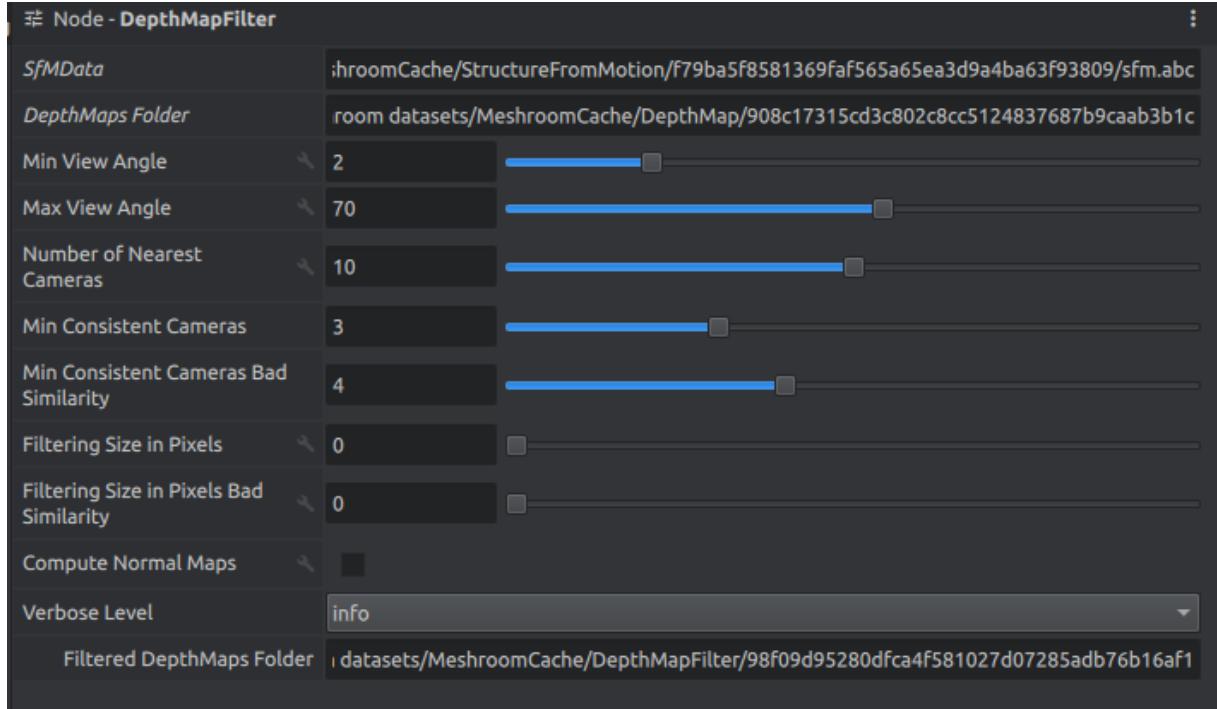


Figure 3.17: DepthMapFilter Node Settings

Meshering Node

The Meshering is the ninth step of the Meshroom Pipeline Reconstruction. This node will take all the results from the previous steps and it will produce a mesh file (*.obj file*) that can be later processed on the PCL framework. All the settings for this node can be referenced to table 2.9.

The configured setting for best results are configured below.

Estimate Space From SFM : This setting was selected in order to estimate the space from the results of the structure from motion node.

Min Obsevation For SFM Space Estimation : The minimum observation selected value for estimation of structure from motion is 1.

Min Obervation Angle For SFM Space Estimation : The selected value for angle

for SFM space estimation is 45. This angle should be modified based on the desired result.

Min Step : The selected step used to load the depth values from depth maps is 1

Partitioning : **singleBlock** was selected for best results.

Repartition : **MultiResolution** was selected for best results.

Refine Fuse : This setting was selected in order to refine the depth map fusion with new pixel size which is defined by the angle and similarity scores.

Weakly Supported Surface Support : This setting was selected in order to improve support of weakly supported surfaces.

Add Landmarks to the Dense Point Cloud : This field was selected in order to add Structure From Motion (SFM) Landmarks to the dense point cloud and improve the resulting mesh.

Colorize Output : This field was selected to generate a colorized output point cloud and mesh.

Add Mask Helper Points : This field was selected in order to add helper points on the outline of the depth maps masks.

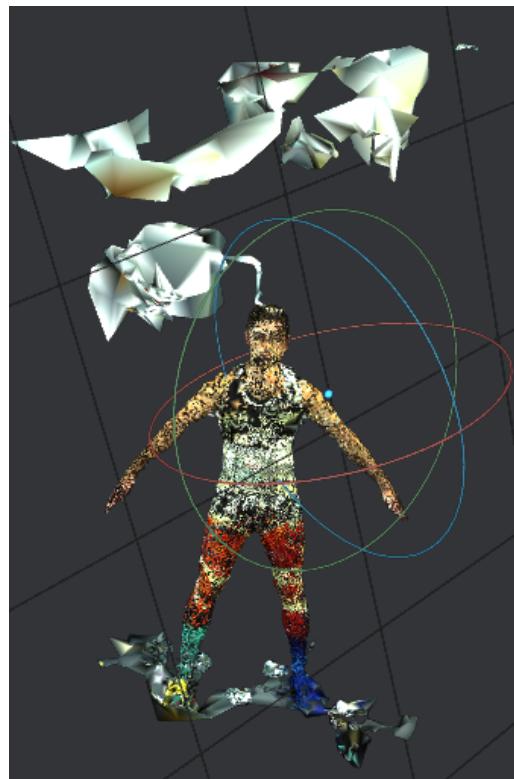
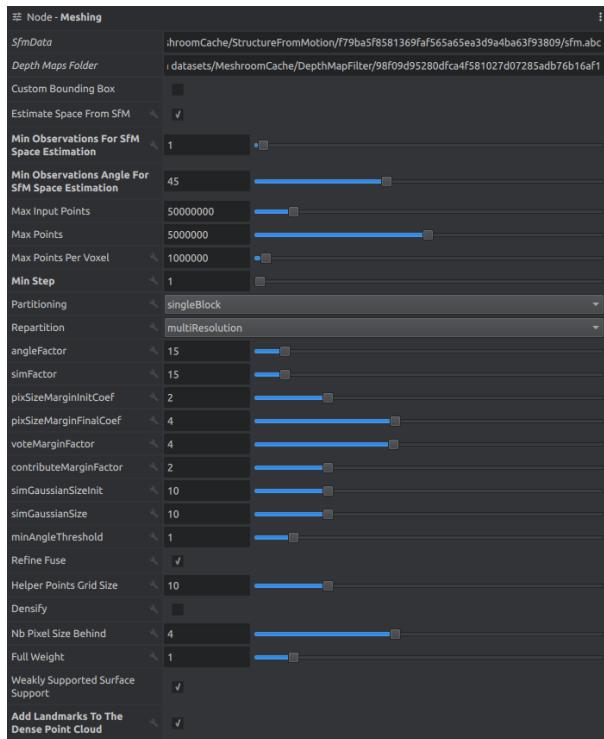
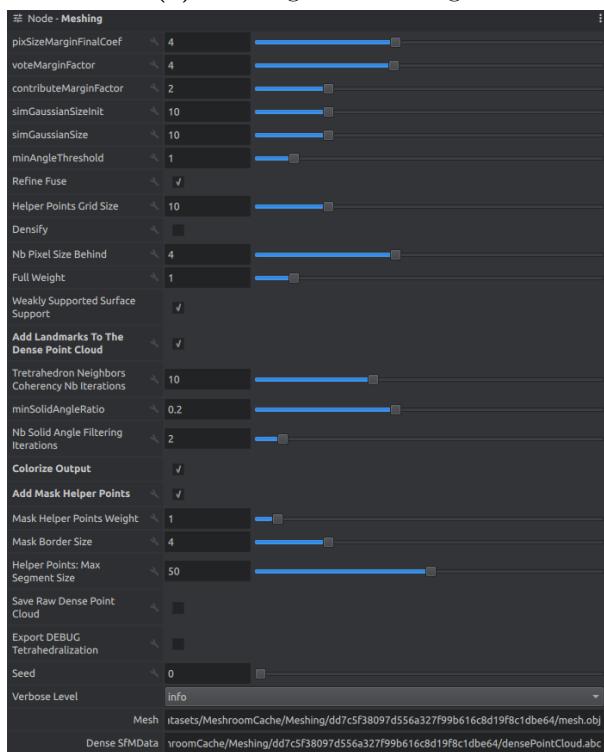


Figure 3.18: Meshing Node Output



(a) Meshing Node Settings



(b) Meshing Node Settings

Figure 3.19: Entire Meshing Node Settings

All the remaining settings are configured as default and are referenced on figures 3.19a and 3.19b.

MeshFiltering Node

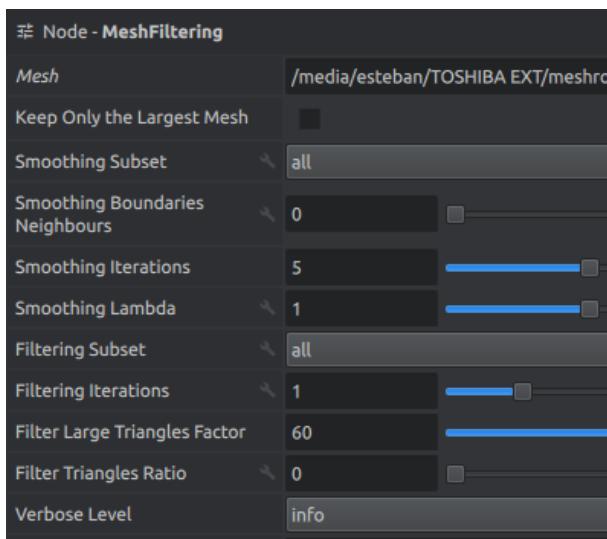
The MeshFiltering Node is the tenth step of the reconstruction pipeline. It smooths the resulting point cloud and mesh. The output of this step will be used in the Data processing framework based on the PCL library. All the settings for this node can be referenced to table 2.10.

The configured setting for best results are configured below.

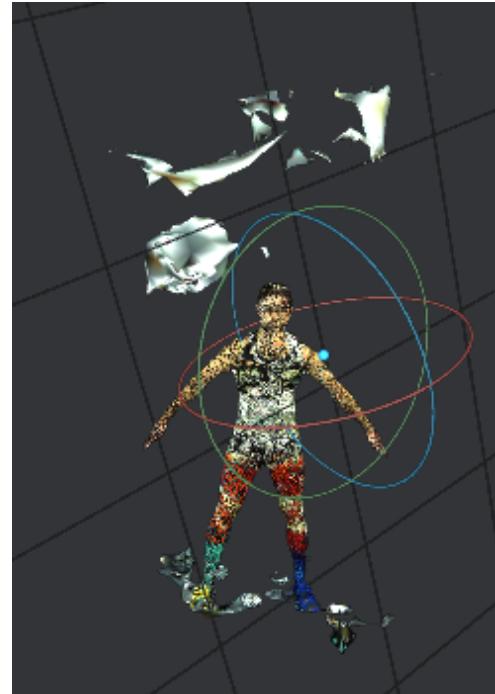
Smoothing Iterations : The selected number of smoothing iterations for best results is 5.

Filter Large Triangles Factors : The selected factor to remove all large triangles is 60.

All the remaining settings are configured as default and are referenced on figure 3.20a.



(a) MeshFiltering Node Settings



(b) MeshFiltering Node Output

Figure 3.20: MeshFiltering Node Settings & Output

Texturing Node

The Texturing Node is the Final Step of the Meshroom Reconstruction Pipeline. It generates a textures mesh. All the settings for this node can be referenced to table 2.11.

The configured settings for best results are configured below.

Texture Side : The selected output texture size is 8192.

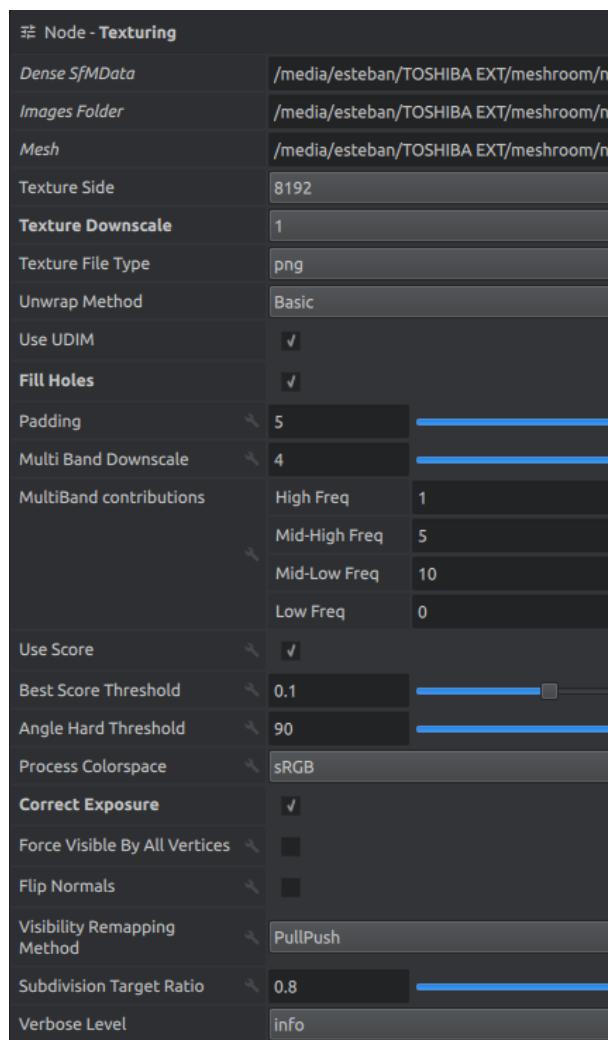
Texture Downscale : The selected texture downscale factor is 1.

Texture File Type : The selected texture file type is **PNG**.

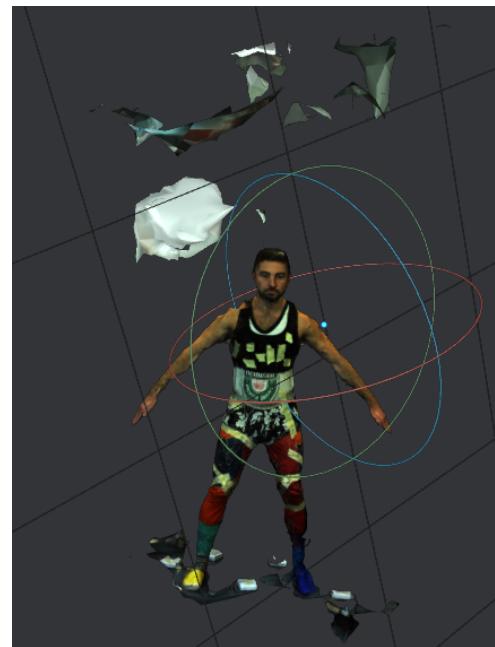
Fill Holes : This setting was selected in order to fill texture holes with plausible values.

Correct Exposure : This setting was selected in order to uniformize the exposure value from the images.

All the remaining settings are configured as default and are referenced on figure 3.21a.



(a) Texturing Node Settings



(b) Texturing Node Output

Figure 3.21: Texturing Node Settings & Output

3.3 Data Processing Framework

Results

Conclusion

Recommendations

Bibliography

- Aber, J. S., Marzolff, I., Ries, J. B., and Aber, S. E. (2019). Chapter 3 - principles of photogrammetry. In Aber, J. S., Marzolff, I., Ries, J. B., and Aber, S. E., editors, *Small-Format Aerial Photography and UAS Imagery (Second Edition)*, pages 19–38. Academic Press, second edition edition.
- AliceVision (2021). Meshroom manual.
- Allene, C., Pons, J.-P., and Keriven, R. (2009). Keriven r.: Seamless image-based texture atlases using multi-band blending. pages 1 – 4.
- Artac, M., Jogan, M., and Leonardis, A. (2002). Incremental pca for on-line visual learning and recognition. In *Object recognition supported by user interaction for service robots*, volume 3, pages 781–784 vol.3.
- Baumberg, A. (2003). Blending images for texturing 3d models. *Proceedings of the British Machine Vision Conference*.
- Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of in-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on pattern analysis and machine intelligence*, 26:1124–37.
- Burt, P. and Adelson, E. (1983). A multiresolution spline with application to image mosaics. *ACM Trans. Graph.*, 2:217–236.
- Cheng, J., Leng, C., Wu, J., Cui, H., and Lu, H. (2014). Fast and accurate image matching with cascade hashing for 3d reconstruction. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.
- Daanen, H. A. and Psikuta, A. (2018). 3d body scanning. *Automation in Garment Manufacturing*, pages 237–252.

- Digumarti, S. T., Chaurasia, G., Taneja, A., Siegwart, R., Thomas, A., and Beardsley, P. (2016). Underwater 3d capture using a low-cost commercial depth camera. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9.
- Fang, Z., Zhao, S., Wen, S., and Zhang, Y. (2018). A real-time 3d perception and reconstruction system based on a 2d laser scanner. *Journal of Sensors*, 2018:1–14.
- Fernández Alcantarilla, P. (2013). Fast explicit diffusion for accelerated features in nonlinear scale spaces.
- Fischler, M. and Bolles, R. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395.
- Gao, W. and Tedrake, R. (2019). Surfelwarp: Efficient non-volumetric single view dynamic reconstruction. *CoRR*, abs/1904.13073.
- Gautier, Q., Garrison, T., Rushton, F., Bouck, N., Lo, E., Tueller, P., Schurgers, C., and Kastner, R. (2020). Low-cost 3d scanning systems for cultural heritage documentation. *Journal of Cultural Heritage Management and Sustainable Development*, ahead-of-print.
- Gonzalez-Rodriguez, A. (2020). Growing pains of ill-sized clothing for both consumers and brands.
- Hirschmuller, H. (2005). Accurate and efficient stereo processing by semi-global matching and mutual information. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, volume 2, pages 807–814 vol. 2.
- Ichim, A. I. (2012).
- Intel (2020a). Depth camera d435i intel realsense depth and tracking cameras.
- Intel (2020b). Intel nuc mini pcs2020.
- Jancosek, M. and Pajdla, T. (2011). Multi-view reconstruction preserving weakly-supported surfaces. In *CVPR 2011*, pages 3121–3128.
- Jancosek, M. and Pajdla, T. (2014). Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces. *International Scholarly Research Notices*, 2014:1–20.
- Jo, W., Kannan, S. S., Cha, G.-E., Lee, A., and Min, B.-C. (2020). Rosbag-based multimodal affective dataset for emotional and cognitive states.

- Kazhdan, M., Bolitho, M., and Hoppe, H. (2006). Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, SGP '06, page 61–70, Goslar, DEU. Eurographics Association.
- Keller, M., Lefloch, D., Lambers, M., Izadi, S., Weyrich, T., and Kolb, A. (2013). Real-time 3d reconstruction in dynamic scenes using point-based fusion. page 8.
- Kneip, L., Scaramuzza, D., and Siegwart, R. (2011). A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. pages 2969–2976.
- Lepetit, V., Moreno-Noguer, F., and Fua, P. (2008). Epnp: An accurate o(n) solution to the pnp problem. *International Journal of Computer Vision*, 81:155–166.
- Levin, D. (1998). The approximation power of moving least-squares. *Math. Comput.*, 67(224):1517–1531.
- Levy, B., Petitjean, S., Ray, N., and Maillot, J. (2002). Least squares conformal maps for automatic texture atlas generation. *ACM Trans. Graph.*, 21:362–371.
- Li, Y., Wang, S., Tian, Q., and Ding, X. (2015). A survey of recent advances in visual feature detection. *Neurocomputing*, 149:736–751.
- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–.
- Magnenat-Thalmann, N., Seo, H., and Cordier, F. (2004). Automatic modeling of virtual humans and body clothing. *Journal of Computer Science and Technology*, 19(5):575–584.
- Moulon, P., Monasse, P., and Marlet, R. (2012). Adaptive structure from motion with a contrario model estimation.
- Muja, M. and Lowe, D. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. volume 1, pages 331–340.
- Murcia, H., Monroy, M. F., and Mora, L. F. (2018). *3D Scene Reconstruction Based on a 2D Moving LiDAR*.
- Newcombe, R. A., Fox, D., and Seitz, S. M. (2015). Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Nister, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–770.

- Nistér, D. and Stewénius, H. (2006). Scalable recognition with a vocabulary tree. volume 2, pages 2161 – 2168.
- Otero, I. (2015). *Anatomy of the SIFT method*. PhD thesis.
- Rusu, R. B. and Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China.
- Scharstein, D., Szeliski, R., and Zabih, R. (2001). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*, pages 131–140.
- Shah, R., Deshpande, A., and Narayanan, P. (2014). Multistage sfm: Revisiting incremental structure from motion. pages 417–424.
- Shchurova, C. I. (2015). A methodology to design a 3d graphic editor for micro-modeling of fiber-reinforced composite parts. *Advances in Engineering Software*, 90:76–82.
- Siena, F., Byrom, B., Watts, P., and Breedon, P. (2018a). Utilising the intel realsense camera for measuring health outcomes in clinical research. *Journal of Medical Systems*, 42.
- Siena, F., Byrom, B., Watts, P., and Breedon, P. (2018b). Utilising the intel realsense camera for measuring health outcomes in clinical research. *Journal of Medical Systems*, 42.
- Slavcheva, M., Baust, M., and Ilic, S. (2018). Sobolevfusion: 3d reconstruction of scenes undergoing free non-rigid motion. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Spahiu, T., Shehi, E., and Piperi, E. (2014). *Extracting body dimensions from 3D body scanning*.
- Strecha, C., Fransens, R., and Van Gool, L. (2006). Combined depth and outlier estimation in multi-view stereo. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2394–2401.
- Strutz, T. (2016). *Data Fitting and Uncertainty (2nd edition)*.
- Sturm, J., Bylow, E., Kahl, F., and Cremers, D. (2013). Copyme3d: Scanning and printing persons in 3d. *Lecture Notes in Computer Science*, pages 405–414.
- Treleaven, P. and Wells, J. (2007). 3d body scanning and healthcare applications. *Computer*, 40(7):28–34.

- Vedaldi, A. and Fulkerson, B. (2010). Vlfeat: An open and portable library of computer vision algorithms. volume 3, pages 1469–1472.
- Yu, G. and Morel, J.-M. (2011). Asift: An algorithm for fully affine invariant comparison. *Image Processing On Line*, 1.
- Zhou, Q.-Y., Park, J., and Koltun, V. (2018). Open3D: A modern library for 3D data processing. *arXiv:1801.09847*.

Appendix

Risk	Affected Phase	Severity	Likelihood	Consequences	Mitigation Plan
Lack of resources access for sensors	Research, Testing, Development and Deployment	Medium-High	Unlikely	Time setback and delays in the milestones for the delivery which could hinder the final demonstration and showcase.	Ensure that all sensors and hardware are purchased accordingly . Ensure that all SDK and driver are installed in workstations
Pandemic Progression	All Phases	Medium	Very Unlikely	The situation might trigger further adjustments to the scope of the project. As well as supervision availability	Moving all communications channel to MS teams and Zoom.
Supervisor unavailability and Low Guidance	All Phases	High	Very Likely	Lack of support could trigger a delay in the project competition	Establish as Well defined communication method and Look for other sources of guidance.
Excessive testing for Data Fusion framework and data modelling	Development and deployment	Medium	Unlikely	Loss of time for the final integration process for the 3D data reconstruction model.	Establish a reasonable time frame for testing all data visualization techniques.
No Finalization of developing framework for Data fusion and 3D reconstruction	Development	High	Unlikely	This will cause that the project wont finalise successful at the end	Try to gather as much sources as possible in order to get finalise the project.
Inability to fully integrate with the components from the other teams	Deployment	Medium	Unlikely	The consequence is that the project will get delay and the components will work independently but not combined.	Try to develop a an organised schedule that will give the proper timing for the correct integration with the other components and teams.

Table A.1: Risk Matrix

Subject	Contact	Channel	Discussion Topic	Frequency	Notes
Supervisor Meeting 1. Initial Project Discussion	Dr Teresa Vidal Calleja	Zoom	Introduce the proposed topic and discussion about how the components were labelled and divided.	One time occurrence 15/7/2020	Agreement on division of different components of the project. Establish requires skills.
Initial team discussion	Teresa Vidal Calleja, Asher Katz, Mark Liu	Zoom, Teams	Initial discussion about different components and Scope of the project	One time meeting 29/7/2020	Introduction of all team members and project discussion.
Team Meeting	Teresa Vidal Calleja, Asher Katz, Mark Liu	Zoom Teams	Weekly Discussion for all topics regarding the project and advancement.	Weekly 1 hour meeting every Tuesday until 20/6/2021	General Discussion about project status and updates.
Initial Data Acquisition Testing	Teresa Vidal Calleja, Asher Katz, Mark Liu	In Person	Initial scanning of model. Initial glance at sensor model.	One time occurrence 25/8/2020 3 Hours meeting	Initial testing with Intel real sense d435i
Engineering Research Preparation consultation	Xi Jin	Zoom	Assessment Task consultations and feedback	28/7/2020 18/8/2020 29/9/2020	Assessment task related questions and overview
Discussion on Software and hardware for Scanner	Asher Katz	TBA	General discussion and planning for integration and deployment	TBA	Lay out and planning or integration of the hardware component and 3D data reconstruction component
Final Report Submission	Teresa Vidal Calleja	Face to Face or zoom	Get final approval of project from supervisor	Single occurrence TBA	Get Final mark and sign from supervisor.
Project Showcase	Academic panel	TBA	Presentation of project	TBA	TBD
TOTAL HOURS COMPLETED				Expected	Actual
				65 Hours	TBD

Table A.2: Communication Plan