School
of
Electronics and Communication

V sem Mini Project
on

VOLUMETRIC CONTENT GENERATION

By:

1. **Sameer Kulmi** USN: 01FE17BEC158
2. **Dinesh Dhotrad** USN: 01FE17BCS069
3. **Siddharth Katageri** USN: 01FE17BCS093
4. **Shivaraj Chattannavar** USN: 01FE17BEC191

Semester: V

Under the Guidance of

Prof. Ramesh Tabib

K.L.E SOCIETY'S
KLE Technological University,
HUBBALLI-580031
2019-2020

SCHOOL OF ELECTRONICS AND COMMUNICATION

CERTIFICATE

This is to certify that project entitled **" Volumetric Content Generation "** is a bonafide work carried out by **Sameer Kulmi(01FE17BEC158) Dinesh Dhotrad(01FE17BCS069) Siddharth Katageri (01FE17BCS093) Shivaraj Chattannavar (01FE17BCS191)**. The project report has been approved as it satisfies the requirements with respect to the mini project work prescribed by the university curriculum for BE (V semester) in School of Electronics and Communications of KLE technological University for the academic year 2019-2020.

Prof. Ramesh Tabib	**Dr. Nalini C. Iyer**	**Prof. N. H. Ayachit**
Guide	**Head of School**	**Registrar**

External Viva:

Name of Examiners **Signature with date**

1.

2.

ACKNOWLEDGMENT

Any accomplishment requires the effort of many people and this work is not different. The successful completion of our project work is the outcome of invaluable contribution and assistance of several people. We offer our sincere thanks and remain indebted to our teacher, our guide to Prof. Ramesh Tabib whose inspiring stands gave us an immense pleasure to do the work. Their constant guidance and support was the key which kept us moving ahead steadily even during critical times. We would like to also thank Dr.Uma Mudengudi, Prof. Ujwala Patil and Prof. Chaitra Desai for guiding us in various aspects of our project.
We are grateful to Dr. Nalini C. Iyer, Head of School of Electronics and Communication, for all the facilities provided in carrying out our project. We extend our sincere thanks to the review committee for their cooperation in solving the critical problems that we come across during the implementation of the project. We extend our heartfelt thanks to the staff members of our department for their support. We sincerely thank the non-teaching staff of our department for their kind cooperation and making us available all the requirements on time. We also thank our parents and friends for their moral support which boosted up in critical times.

-The project team

ABSTRACT

In this paper, we propose a pipeline for real-time volumetric 3D reconstruction of object using multiple views. The object can be dynamic, complex and non-rigid in nature.The generated 3D model serve as fundamental in AR/VR content generation.

A multiple RGB-D camera setup is made to capture the 360° view of object. These cameras are to be calibrated in order to know the relative pose between each other. TSAI camera calibration method is incorporated for calibration of RGB camera. The depth maps are aligned with respect to RGB image and the error in depth camera is addressed, as it plays essential role in order to generate a perfect 3D model. All the depth maps are then fused to generate a single 3D mesh based on truncated signed distance function(TSDF), a volumetric scene representation that allows for integration of multiple depth images taken from different viewpoints.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Volumetric content generation is the 3D model representation of any object, person, or a scene. Multiple RGB-D cameras are used to capture the 360° view of the object. The object is dynamic and non-rigid in nature, that is, the object can change its shape with time and deformities can occur in its shape. Microsoft Kinect or Intel Realsense cameras are used to capture RGB-D information.
The calibration of multiple cameras is one of the main challenges. The RGB cameras are calibrated using TSAI camera calibration method. This gives us the relative pose of each camera with other, which will help in fusion of meshes from different cameras. Depth cameras are calibrated and aligned with respect to RGB image so as to improve the ill-posed model to become well-posed.
Later, the depth maps from multiple cameras are fused together using truncated signed distance function(TSDF). It takes aligned depth map and RGB image as input along with the camera parameters and camera pose. This generates a mesh from given depth map and RGB image.

1.1 Motivation

In recent years, a significant growth can be seen in use of Augmented Reality(AR) and Virtual Reality(VR) which has increased the demand for a high quality 3D content. Volumetric capturing of a scene/ object can be used to view it's 360° virtual 3D object and it serves as fundamental for AR and VR content generation. Also, the virtual 3D object can be viewed from any point of view. This enhances the user experience and this can be applied to any field, like to capture a dance by a person, or to capture a sports-person in action.
Our motive is to develop a system for a real-time 3D reconstruction.This can lead towards new possibilities such as ability to watch a sports event, remote concert live in full 3D, even the ability to communicate in real-time by remotely capturing of person or object using AR/VR gadgets.

1.2 Objectives

1.Perform extensive literature survey and provide analysis on the existing methods for camera calibration and pose estimation.
2.Perform literature survey on different materials required for camera.
3.Designing a calibration object for calibration of multi cameras mounted in 360-degree.

4.Implement an algorithm for back projection and error computation for fine tuning of the computed poses and depth.

1.3 Literature survey

Sl.no	Title with Conference/Journal details	Year	Addressed Problem	Advantages/ Challenges
1.	**DoubleFusion** (IEEE Conference on Computer Vision and Pattern Recognition (CVPR))	2018	Real-time Capture of Human Performances with Inner Body Shapes from a Single Depth Sensor	Advantages: Real-time system that combines volumetric dynamic template fitting to simultaneously reconstruct detailed geometry from a single depth camera Challenges: Single depth camera should be calibrated to the object so that it tracks the moment of the object instead of whole object creation
2.	**Motion2Fusion** (c 2017 Association for Computing Machinery)	2017	Real-time Volumetric Performance Capture	Advantages: The ability to model the complex pose, shape, appearance and motions of humans in 3D. Challenges: Depth details are very compact of facial expressions there should be fine tuning of the 3D model to see the emotions.
3.	**Fusion4D** (SIGGRAPH '16 Technical Paper, July 24-28)	2016	Real-time Performance Capture of Challenging Scenes	Advantages: Use of RGBD camera help in getting data of RGB, Depth and Segmentation which helps in better processing. Challenges: Since there is use of more than 10 RGBD cameras handling incoming data and processing it real-time.

1.4 Problem statement

Volumetric Content Generation: Implementing a multi-camera calibration method and multi-camera pose estimation technique towards volumetric data capture and generation.

Figure 1.1: Black box

1.5 Organization of the report

Chapter 2 contains details on System design of the project which includes block diagram, morphological chart and design of system. Chapter 3 contain the implementation details of the system along with flow chart and algorithm. Chapter 4 Contain the results of the project followed by Conclusions and future scope in Chapter 5.

Chapter 2

System design

In this chapter, we discuss on the architecture of the proposed system, its block diagram, modules, interfaces used.

2.1 Functional block diagram

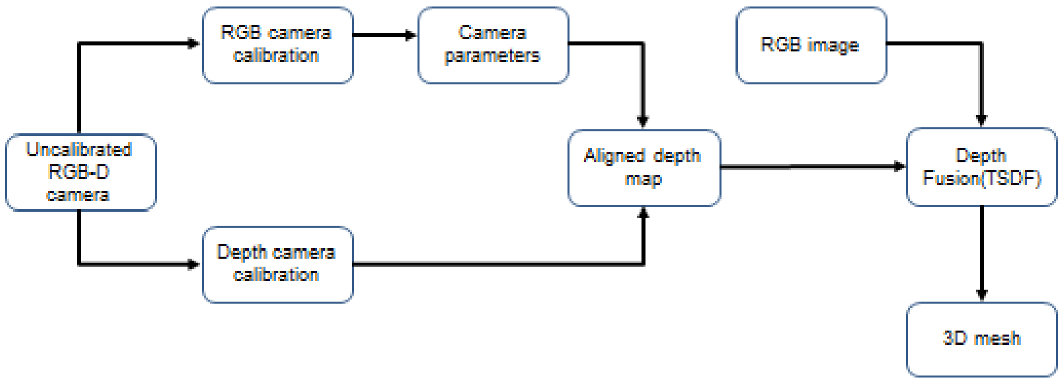

Figure 2.1: Pipeline

Figure 2.1 shows the pipeline of system. Initially, the calibration of the RGB-D camera has to be done. The RGB cameras are calibrated using TSAI calibration method.The depth camera is calibrated and the depth maps are aligned with respect to RGB images so as to generate a well-posed 3D model. Later, the aligned depth maps are fused together with help of truncated signed distance function(TSDF). This is a very robust technique for depth fusion. RGB images give the required color and texture to the mesh generated by TSDF. So the output will be a 3D mesh generated via depth map, RGB image, camera pose, extrinsic and intrinsic parameters.

2.2 Morphological chart

Morphological chart describes different mechanism that can be used for a particular function.Table 2.1 shows the morphological chart.
For camera calibration, listed are two methods: TSAI and Zhang calibration methods. For

Table 2.1: Morphological chart

Functions/Means	**Option 1**	**Option 2**
Camera calibration	TSAI calibration	Zhang calibration
Depth fusion	TSDF	Stereo fusion

a multi camera calibration setup, Zhang methodology as drawbacks like, change of center of object point, which makes it less efficient for the present use. Whereas, TSAI calibration method uses a 3 dimensional calibrator object, which can be seen by all the cameras at a time and calibration is possible.
For the depth fusion, we incorporate TSDF method. This method can be implemented on GPU for faster computations, and parallel processing is also possible. TSDF is proven to be robust algorithm.

2.3 Design alternatives

This section talks about one of the alternative design for the intended pipeline.

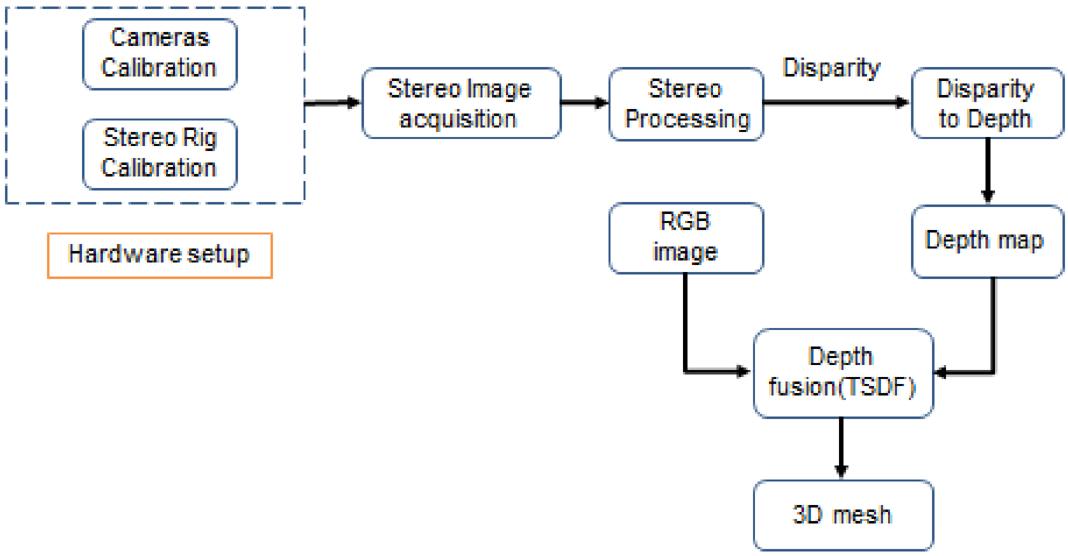

Figure 2.2: Alternative design

Figure 2.2 shows the alternative design for the system. The process here is as follows: Initially, the stereo camera rig is calibrated using Zhang camera calibration method. This requires capturing of multiple checkerboard images from the two stereo cameras. This calibration method gives the K,R and T matrices of the camera. Now, the stereo images of a particular object are acquired, followed by various processing methods to obtain the depth map. This is computationally hard way of obtaining the depth maps. Later, the depth maps are fused together to obtain a 3D mesh using TSDF algorithm.

2.4 Final design

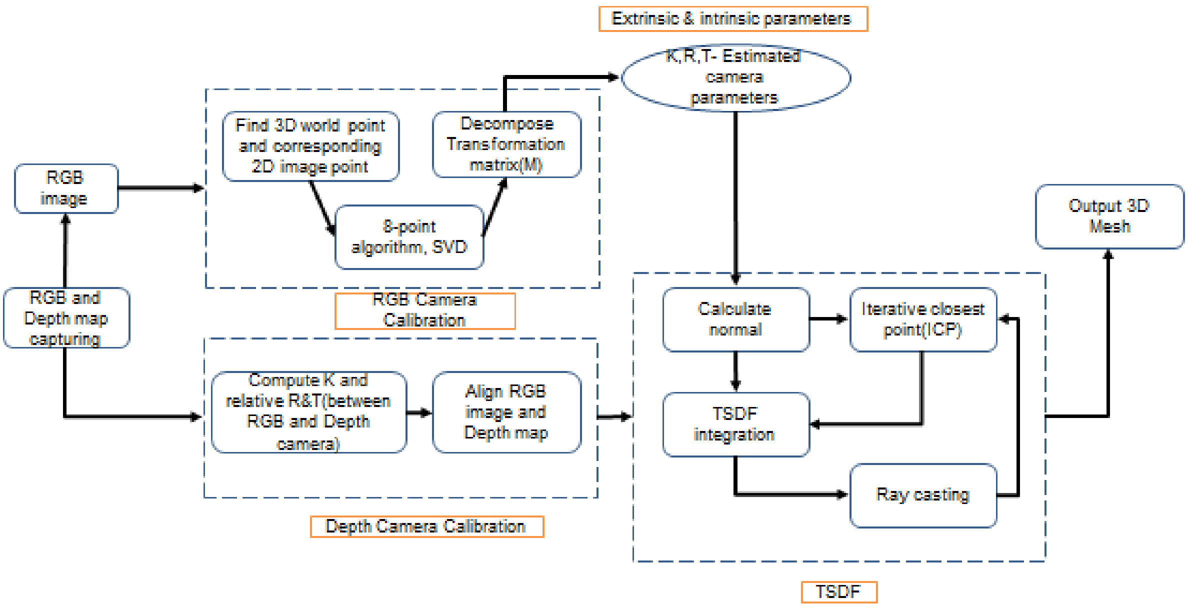

Figure 2.3: Final system deign

Figure 2.3 shows the final design of system. A multiple camera rig of RGB-D cameras is setup. The first step is to perform calibraiton of RGB cameras in order to know the camera parameters(intrinsic and extrinsic). Depth cameras are calibrated as [2]. The depth map alignment with respect to RGB image plays a crucial role because, if there is no alignment, the depth map is distorted as compared to RGB images which results in a ill-posed 3D model generation. After alignment, the camera parameters along with camera pose and the aligned depth maps are passed to the TSDF block. This block renders a 3D mesh by fusing multiple depth images.

Chapter 3

Implementation Details

In this chapter, we discuss on the implementation details of the final design.

3.1 Specifications and final system architecture (White Box)

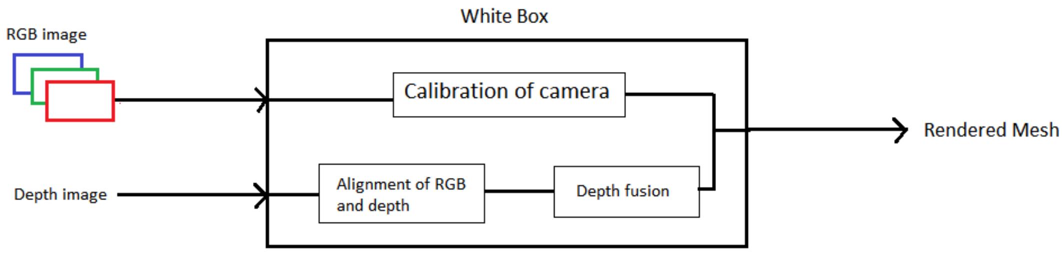

Figure 3.1: White box

3.2 Algorithm

3.2.1 Calibration and Alignment

The RGB cameras are calibrated using TSAI camera calibration method. Firstly, a image of the calibrator object is captured from camera. Using Edge detection and corner point detection algorithms, we find the corners of the checkerboard in calibrator object as shown in figure3.3.

The depth cameras are calibrated with help of the relative R and T between depth and RGB camera on a Kinect v1. This is followed by the alignment of depth map and RGB image. Alignment is a major process which helps to generate a well-posed 3D model. If there is misalignment between the RGB and depth map, then the 3D model wont be accurate as shown in figure3.4.

(a) Controlled calibrator object (b) Points detected

Figure 3.2: Calibrator object

(a) RGB image (b) Depth map

Figure 3.3: Input data

.5

Figure 3.4: Misaligned mesh generation

3.2.2 Depth Fusion

Truncated signed distance function(TSDF) is incorporated for depth fusion. This algorithm takes aligned depth map and RGB image as input along with camera intrinsic parameters and camera pose. The example of a mesh generated after depth fusion is shows in figure

Figure 3.5: Generated 3D mesh

Chapter 4

Optimization

In this chapter, we discuss on the optimization we used on the design alternatives to design of the final pipeline of system.As discussed in last chapter we have one design alternative and a final design. This final design is the the optimizer view of given alternate design.

4.0.1 Optimization in Depth map computation

The major optimization is in process to capture Depth map of a scene. We can obtain depth map via homography process. This requires a stereo pair of images, in which we calculate the disparity between them and followed by depth estimation. This process is tedious as compared to calculation of depth directly from a RGB-D camera. This makes computation of depth easy as compared to stereo pair depth generation.

4.0.2 Optimization in Multi-camera calibration

There are two methods to calibrate a camera: Zhang and TSAI calibration methods.
When we use Zhang method , the calibrator object is a 2D planar checkerboard. For multi-camera calibration, the object center changes as we move the object for capturing it from different view points. Now if we estimate the R and T matrix for new camera it won't be in the same plane. Thus, we move to TSAI camera calibration method, in order to address this problem.

Chapter 5

Results and Discussions

5.1 Result Analysis and Discussion on optimization

5.1.1 Calibration of multiple cameras:

A controlled structured calibration object in built for performing the TSAI camera calibration. We can observe the points which are detected in the image (corner points). Now if we know the image co-ordinate, we can get its corresponding world co-ordinate as it is a structured object. The points which are selected must be on more the one plane, in order to solve the set of equations

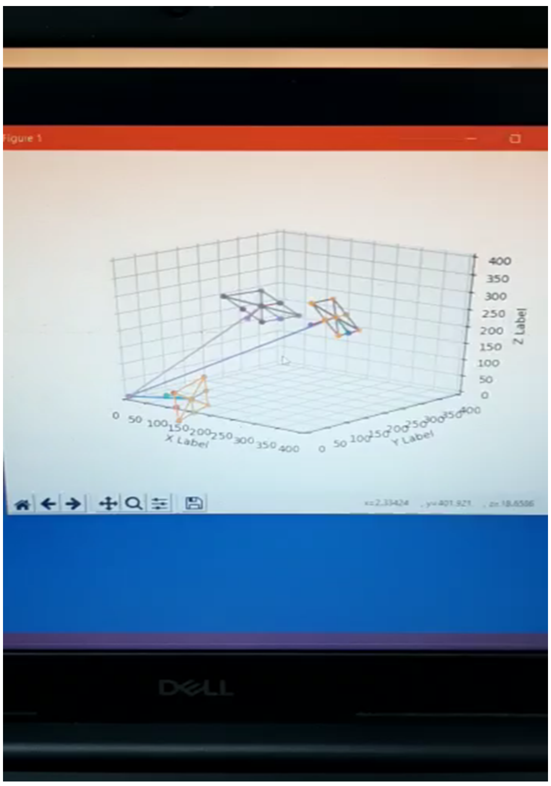

Figure 5.1: Extrinsic parameter visualization

Multiple images are taken from the same camera at different moment of time. Figure5.1, shows the visualization of extrinsic parameters of the camera, specifically rotation and translation, which is obtained after TSAI calibration.

5.1.2 Fusing multiple depth images:

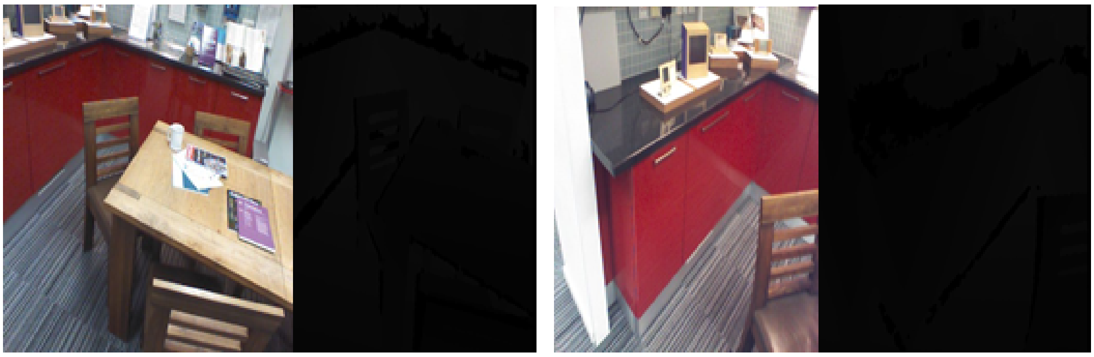

Figure 5.2: Multiple RGB and Depth images

100 such frames as shown in Figure 5.2 of a kitchen scene having the RGB image and Depth image are taken as a dataset. Each frame's corresponding pose is also known. Using all these inputs volume is reconstruction of the kitchen scene by fusing all the depth images shown below.

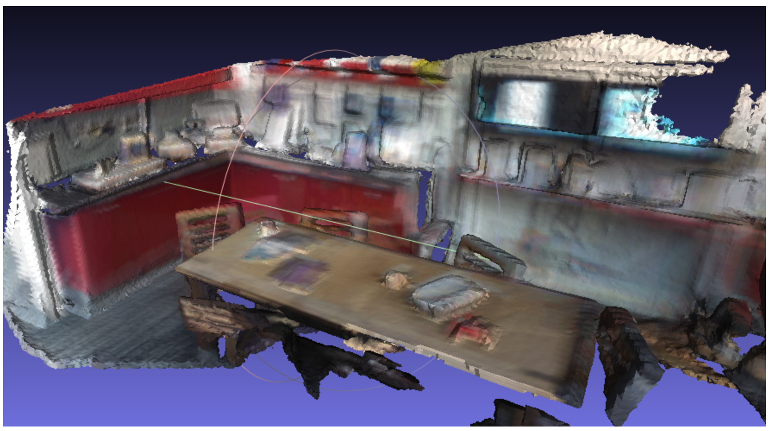

Figure 5.3: Fused mesh

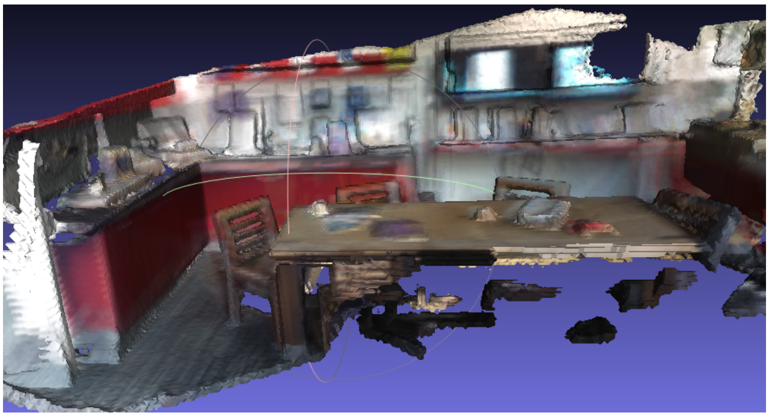

Figure 5.4: Fused mesh

Chapter 6

Conclusions and future scope

6.1 Conclusion

Our present work includes, multiple RGB camera calibration but single depth camera calibration. Single view point is used to generate a volumetric representation for a static object currently. As we face occlusion problem, which results in hole generation we need to move to a multiple camera setup.

6.2 Future scope

6.2.1 Application in the societal context

Future scope of our project would be to introduce a multiple camera setup. As all cameras are independent of each other a method for synchronizing the inputs from all the cameras should be proposed. The depth camera error should to addressed as well which is one of the major challenges to solve as it directly affects the quality of the mesh generated.

6.3 References

1.Motion2Fusion (2017).
2.KinectFusion (2011).
3.Truncated Signed Distance Function: Experiments on Voxel Size.
4.TSDF Volume Reconstruction.