STRUCTURE FROM MOTION 3D SCANNER

By

Muhammad Ifham Shahid

Areeba Majeed

Muhammad Talha Tahir

Syeda Sana Riaz

Submitted to the Faculty of Computer Science, Military College of Signals
National University of Sciences and Technology, Rawalpindi in partial fulfillment for
the requirements of a B.E Degree in Software Engineering
JUNE 2020

In the name of Allah, the Most Beneficent, the Most Merciful

ABSTRACT

Technology has a deep social impact on our lives in today's world. With the increase in technology, humans are more prone to comfort compared to what they were some years ago. Structure from Motion (SfM) being one of the emerging technologies of the present era is proving to be useful in several fields. It is like the holy grail of multiple view geometry. The idea proposed is of a Structure from motion technique for creating 3D models of a body or an object.

Computational methods to generate three-dimensional models from photographs have received increasing attention in the last years. Nowadays, although they are not substitutive to laser scanners, they are a viable alternative in cases where using a laser scanner is not possible. These methods are also attractive because they do not need the acquisition of expensive equipment.

A desktop application has been developed for everyone to create 3D models of their bodies; a standard procedure being followed normally which requires customers to upload 140-200 images having a 40-70% overlap on an application. The result will be displayed after 48 hours on the screen from where the user will be able to download a ~180cm 3D model of a human body. The result of this evaluation revealed that application has potential benefits to promote online shopping and help retailers enhance their online business.

CERTIFICATE FOR CORRECTNESS AND APPROVAL

It is certified that work present in the thesis –Structure from Motion 3D scanner is carried out by Muhammad Ifham Shahid, Areeba Majeed, Muhammad Talha Tahir and Syeda Sana Riaz under supervision of Dr. Muqeem Sheri for fractional satisfaction of Degree of Bachelor of Computer Software Engineering is right and affirmed. It is fully ample, in scope and excellence, for the degree of Bachelor of Computer Software Engineering from Military College of Signals, National University of Sciences and Technology (NUST). And is original with 18% plagiarism.

Approved By:

Signature:
Asst. Prof Dr. Ahmed Muqeem Sheri
(Main Supervisor)

Signature:
Asst. Prof. Dr. Saddaf Rubab
(Co-Supervisor)

Department of CSE, MCS

Dated: 11th July 2020

DECLARATION

No bit of the work introduced in this proposition has been submitted on the side of another grant or capability either in this establishment or anyplace else.

DEDICATION

With great pleasure, we dedicate our work to our parents, colleagues, and most importantly to our supervisor, Dr. Ahmed Muqeem Sheri, and co-supervisor Dr. Saddaf Rubab who provided us great support, confidence and suggestions throughout the project.

ACKNOWLEDGEMENTS

There is no accomplishment without the desire of ALLAH Almighty. We are thankful to ALLAH, who has given us direction, quality, and empowered us to achieve this undertaking. Whatever we have accomplished, we owe it to Him, in totality. We are likewise appreciative of our folks and family and well-wishers for their excellent help and their basic audits. We might want to thank our supervisor. Dr. Ahmed Muqeem Sheri and co-supervisor Dr. Saddaf Rubab, for their consistent direction and inspiration throughout our venture. Without their assistance, we would have not had the option to achieve anything.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

1. **SFM:** Structure from motion
2. **LIDAR:** Light detection and ranging
3. **RANSAC:** Random sample consensus
4. **SRS:** Software requirement specifications
5. **SDS:** Software design specifications
6. **MVS:** Multi view stereo
7. **LIOP:** Local intensity order pattern
8. **CUDA:** Compute unified device architecture
9. **GPU**: Graphics processing units
10. **SIFT:** Scale invariant feature transform
11. **OPENCV:** Open source computer vision library
12. **OPENGL:** Open graphics library
13. **PNG:** Portable network graphics
14. **JPG:** Joint photographic experts group
15. **ReLU:** Rectified Linear Units

Chapter 1
Introduction

1.1. Overview

The main idea behind the project is to reduce the cost of 3D scanners. Scanners that are present currently in the market uses LIDAR technology or stereo vision. Stereo Vision uses two cameras mounted separately taking different pictures to reconstruct a 3D model. Our goal is to develop an application that will use monocular images and use them for 3D reconstruction. This scanner can be used easily by a person after a week of training and will produce 3D models.

1.2. Problem Statement

Scanners that are currently available in the market are very expensive and require LIDAR technology. LIDAR is way too expensive to be affordable by any industry or layman. LIDAR technology is very expensive and even after spending a lot of money maintenance cost is too much which make these scanners practically very difficult to be obtained by modeling industries.

1.3. Approach

A desktop application will be developed for the generation of a 3D point cloud, the user will have to upload 140-200 images of the object (taken from angles) having a 40-70% overlap to the application. The application will apply SfM technique on the images and develop a three-dimensional point cloud from the images.

1.4. Scope

Structure from Motion is intended to create a 3D point cloud of the human body or any object, so that they can with the help of no training technique, create 3D point clouds This feature will help greatly in the modeling industry. The system will produce 3D point clouds of a human body (~180 cm) that can be processed further to help designers of the fashion industry.

1.5. Objectives

The main objective of this system is an efficient generation of 3D point-cloud, by using structure from motion technique.

During this project, all the aspects of software engineering are covered i.e. survey and feasibility analysis, requirement gathering, architectural and detailed design, implementation and testing along with documentation (SRS, SDS, Test Document, Final

Report, and User manual). Students are also expected to develop extensive knowledge and technical skills in the following fields:

1. Interacting with 3D objects.

2. Programming in a 3D environment.

3. Social behavior analysis.

4. Open-cv and other open-source software libraries' understanding.

1.6. Deliverables

Sr No.	Tasks	Deliverables
1	Literature Review	Literature Survey
2	Requirements Specification	Software Requirements Specification document (SRS)
3	Detailed Design	Software Design Specification document (SDS)
4	Implementation	Project demonstration
5	Testing	Evaluation plan and test document
6	Training	Deployment plan
7	Deployment	Complete application with the necessary documentation

Table 1: Deliverables

1.7. Justification for Selection of Topic

In Pakistan, most of the companies are working on 3D modeling but none of them have ever thought about making Online shopping easier and they take more time to create a 3D model which causes delays. Our idea is to create a 3D model using structure from motion technique.

1.8. Overview of the Document

This document shows the complete working process of our application SoloScan. It starts with the literature review which shows past work done in a similar field, requirement analysis of the system, system architecture which highlights the modules of the software and represents the system in the form of a component diagram, Use Case Diagram, Sequence Diagram and general design of the system. Then it will move on to discuss the detailed Description of all the components involved. Further, the dependencies of the system and its relationship with other products and the capacity of it to be reused will be discussed. In the end test cases and any future work, proposal has been presented.

1.9. Document Conventions

Headings are prioritized in a numbered fashion, the highest priority heading having a single-digit and subsequent headings having more numbers, according to their level. Font used is Times New Roman. All the main headings are of size 18 and bold. All the second level sub-headings are of size 16 and bold. All the further sub-headings are of size 14 and bold. All references in this document are provided where necessary, however, where not present, the meaning is self-explanatory. All ambiguous terms have been clarified in the glossary at the end of this document.

1.10. Intended Audience

This document is intended for:

1. **Developers:** (Project Group)

To be certain that they are building up the correct venture that satisfies the necessities gave in this report.

2. **Testers:** (Project Group, Supervisor)

To have a definite rundown of the features and capacities that must react as indicated by prerequisites.

3. **Users:**

To get acquainted with the possibility of the task and how to utilize/react in disappointment circumstances and propose different highlights that would make it considerably progressively useful.

4. **Documentation writers:** (Project Group)

To recognize what features and how they need to clarify. What innovations are required, how the framework will react in every client's activity, what conceivable framework disappointments may occur, and what are the answers for each one of those disappointments, and so forth.

5. Project Supervisor: (Dr. Ahmed Muqeem Sheri, Dr. Saddaf Rubab)

This document will be used by the project supervisor to check whether all the requirements have been understood and, in the end, whether the requirements have been implemented properly and completely.

6. Project Evaluators: (CSE Dept. MCS)

To know the scope of the project and evaluate the project throughout the development of grading.

Chapter 2
Literature Review

A detailed description of the projects previously carried out in this context will be discussed in this section.

With the advancement in 3D modeling, many useful projects have been developed. Undergraduate Students of MCS, NUST developed a project named 3D Profile Scanner, whose purpose was to create a 3D model using a laser source to produce a light source and stepper motor to rotate belts. It was a hardware-based project.

2.1 Overview of Projects:

Detail about work done in this field is given as:

2.1.1 Structure from Motion to phenotyping in-field crops from 3D modeling

In January 2015, the creator proposed a technique through which a crop column 3D model is assembled and serves as a reason for recovering plant auxiliary boundaries. This model is figured utilizing Structure from Motion with RGB pictures gained by deciphering a solitary camera along the column. At that point, to evaluate plant tallness and leaf region, plant and foundation are segregated by a hearty strategy that utilizes both shading and stature data to deal with low-differentiated areas. The 3D model was scaled, and the plant surface was at long last approximated utilizing a triangular work. [1]

2.1.2 3D point cloud generation with visual features

In this paper, the essayist proposed a strategy to deliver a meager Euclidean 3D point cloud utilizing a computerized sound system vision-based technique that utilizes two video streams being caught by two calibrated cameras. They utilized such a procedure by which SURF highlights were being recognized consequently and coordinated b/w each pair of sound system video outlines. By utilizing triangulation 3D directions of the coordinated component focuses were then determined. The identified SURF highlights in two progressive video outlines were naturally coordinated and the RANSAC calculation was utilized to dispose of confounds. The quaternion movement estimation strategy was then used alongside pack alteration improvement to enlist progressive point clouds. The technique was tried on a database of foundation sound system video streams. [2]

2.1.3 3D Point Cloud Generation with Multiple View Images

In this article, the author explained one of the structures from motion techniques, known as incremental structure from motion, and introduced open-source software related to structure from motion briefly. Their experiments showed 3D point cloud generation results from executing an incremental structure from motion algorithm of simplified bundle adjustment without outlier rejection, depending on the different number of multiple view images. They used ORB feature for correspondence matching between pairs of multiple view images, skipped the outlier rejection process, and applied a simplified version of the bundle adjustment to the algorithm. [3]

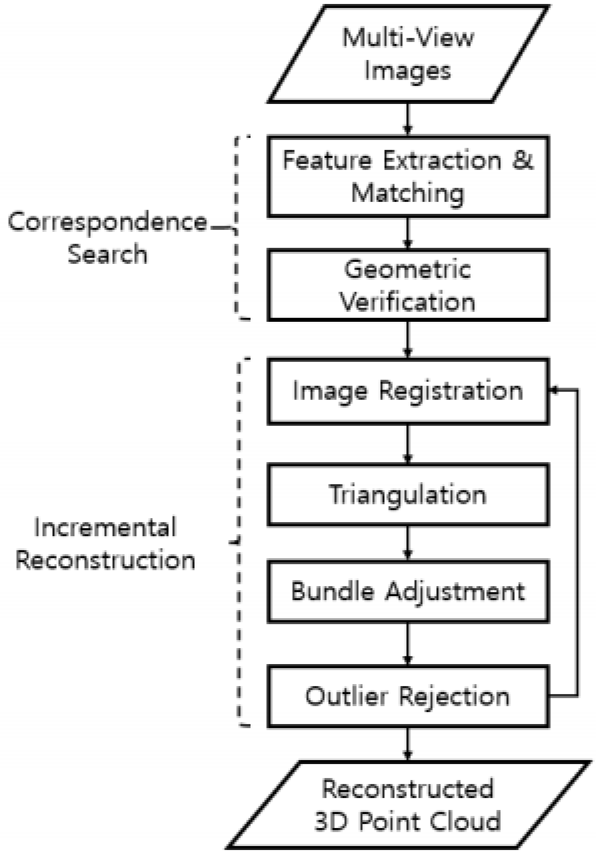

Figure 1- Incremental model

Results obtained were as follows:

Figure 2- 3D reconstrction of a chair

2.1.4 3D Object Reconstruction using a Single Image

In this article, the writer proposed a technique for the 3D model using crude age come closer from a solitary picture, producing a straight-forward type of yield i.e., point cloud arranges. The ground-truth shape for an info picture may be questionable, they planned a design, misfortune capacity, and learning worldview that are novel and viable. They made a contingent shape sampler, equipped for foreseeing various conceivable 3D point mists from an info picture. In their venture, the framework demonstrated a solid exhibition for the fruition of 3D shapes by giving a precision of 90%. [4]

2.1.5 Advanced point cloud generation for photogrammetric modeling

This exploration-based article was disseminated by S. Soule in September 2002. This article portrays strategies for the 3D point cloud. By consolidating existing tie focuses utilized in the group change with a flood-fill coordinating procedure that utilizes a tie point as a seed, the calculation makes it conceivable to create point clouds and surfaces consequently for an assortment of complex genuine scenes. The exhibition evaluation is on-going, yet the first outcomes show that, although computationally costly, the technique can be used in a viable setting with little administrator mediation. They worked on improving the versatile windows and on fitting different surfaces during coordinating with the goal that composite item types can be sent to the CAD condition as a choice to point clouds, which are costly to control. [5]

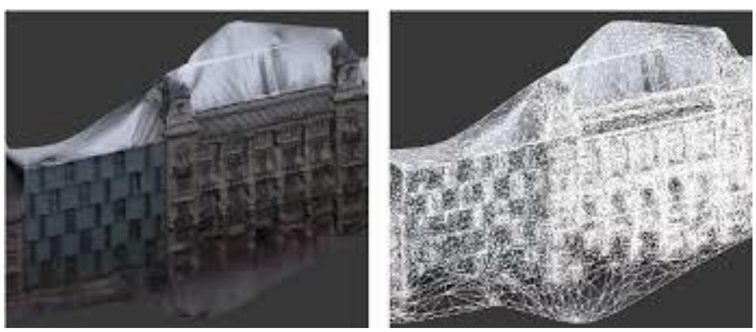

Figure 3- Photogrammetric model

2.1.6 3D Building Modeling using SfM with Telescopic Pole Aerial Photography

This article is circulated in October 2012 by Adam J Mathews and Jennifer Jensen. They used Structure from motion technique to reconstruct historical buildings in San Antonio. To improve the overall quality of the 3D model, they used a low-cost telescopic pole providing low-altitude oblique aerial photography. They created two separate models:

1. Using on-ground photographs
2. Using both on-ground and low-altitude photographs.

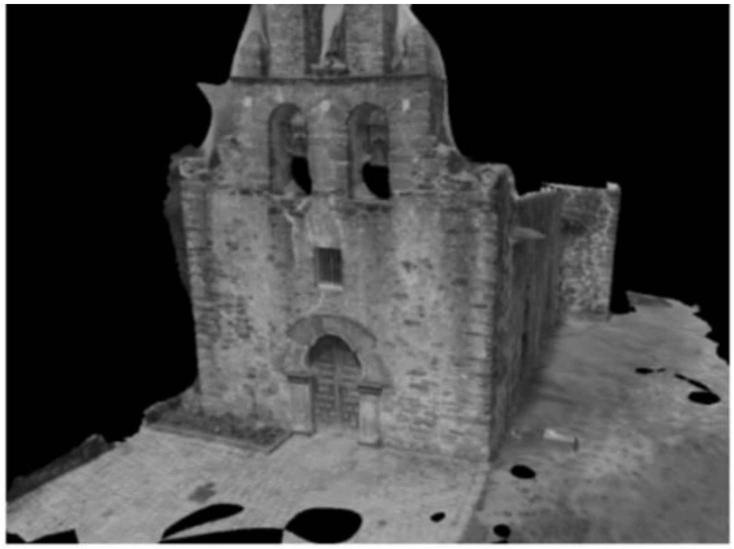

Figure 4- 3D Mesh Model

Figure 5- Ground-based vs Mesh model

They utilized some high angle photographs to help in georeferencing the created model because of the improved capture of ground control points. [6]

2.2 Fields:

For the completion of this project, a detailed study in the following fields as required.

These Fields are:

2.2.1 Structure from Motion

Structure from Motion (SfM) as applied in the geosciences isn't so much a solitary method as a work process utilizing numerous calculations created from three-dimensional (3D) computer vision, customary photogrammetry, and increasingly ordinary overview strategies. The work process is normally known as SfM-MVS, to represent the Multi-View Stereo (MVS) calculations utilized in the last stages. It incorporates the following advances.

(*i*) Detecting picture highlights or key points
(ii) *D*istinguishing correspondences between these key points on various pictures
(iii) *S*ifting these connections to expel geometrically conflicting key point correspondences
(iv) "SfM" or the at the same time assessing 3D scene geometry, camera present, and inner camera boundaries through a pack change
(v) Scaling and georeferencing the resultant scene geometry
(vi) Optimizing the boundaries recognized in the group change utilizing realized ground control focuses (GCPs)
(vi) grouping picture sets for effective preparing
(vii) Applying MVS calculations.

2.2.2 Structure from Motion Method

SfM is the most famous and widely used subject in computer vision for structuring the moving objects.

2.2.2.1 Software

Different SFM software is available commercially. Some of them are as follows:

2.2.2.1.1 Regard3D:

Regard3D is a structure-from-motion program. That implies it can make 3D models from objects utilizing a progression of photos taken of this object from various perspectives.
To acquire a 3D model, the accompanying advances are performed:

1. For each picture, highlights (in some cases likewise called key points) are identified. Key points are points in an object that have a high likelihood to be found in various pictures of a similar item, for instance, corners, edges, and so on. Regard3D utilizes A-KAZE for this reason.

2. For each component, a scientific descriptor is determined. This descriptor has the trademark that descriptors of a similar point in an item in various pictures (seen from various perspectives) are comparable. Regard3D utilizes LIOP (Local Intensity Order Pattern) for this reason.

3. The descriptors from various pictures are coordinated and geometrically separated. The consequence of this progression is an assortment of matches between each picture pair.

4. Presently "tracks" are determined. For each element that is a piece of a match in a picture pair, it is looked additionally in different pictures. A track is created from highlights if these highlights fulfill a few conditions, for instance, a track is seen in at any rate 3 pictures.

5. The subsequent stage is the triangulation stage. All the matches of all the picture sets are utilized to ascertain:

6. The 3D position and normal for the "camera", for example where each picture was shot and the visual attributes of the camera

7. For each "track" 3D position is determined

8. The consequence of the triangulation stage is a sparse point cloud. To acquire an increasingly thick point cloud ("densification"), a few calculations can be utilized.

9. The last advance is classified as "Surface age". The point clouds are utilized to create a surface, either with shaded vertices or with a surface.

Figure 6- Regard3D model

2.2.2.2.2 Mesh room

This is also an online 3D modeling software that requires GPU and CUDA and follows somehow similar steps like Regard3D.

2.2.3 3D Modeling:

Imaging is one of the most significant devices. It has its application in the clinical and building field. Three-dimensional (3D) imaging has been created in the early 1990s and has increased a valuable spot in dentistry, particularly in orthodontics. Huge scope image-based 3D modeling has been a significant objective of PC vision, empowering a wide scope of utilizations including computer-generated reality, picture-based restriction, and self-governing route. One of the most assorted information hotspots for displaying is Internet photograph assortments. The points of this writing survey are, to sum up, the present status of the 3D imaging methods.

2.2.3.1 3D laser scanning

- 3D laser scanning is significant and is costly because of the innovation in laser technology yet it provides better outcomes. However, this procedure has a few detriments for 3D checking. For example:
- The procedure is delayed to such an extent that bending happens in the examined picture.
- While the scanner spins around the body, the individual should remain unmoving for one moment or more. Because of the potential individual development and security issues identified with laser, laser filtering is extremely hard to get advanced models.
- Safety issues are significant, for example, presenting eyes to the laser bar, especially in developing kids
- There is a powerlessness to catch the delicate tissue surface, which brings about troubles in distinguishing proof of milestones because of surface shading.

2.2.3.2. Structured light technique

The structured light technique is one of the non-contact estimation strategies utilized for high goals and highly sensitive 3D modeling. In this technique, a projector, a camera, and a PC are utilized. By preparing the pictures got by the camera, the object is 3D demonstrated. The light force that is transmitted from the projector by and large not a straight capacity of the significant info. This messes brilliance up in the examples anticipated. Hence, pictures got from the camera needs to the gamma corrected.

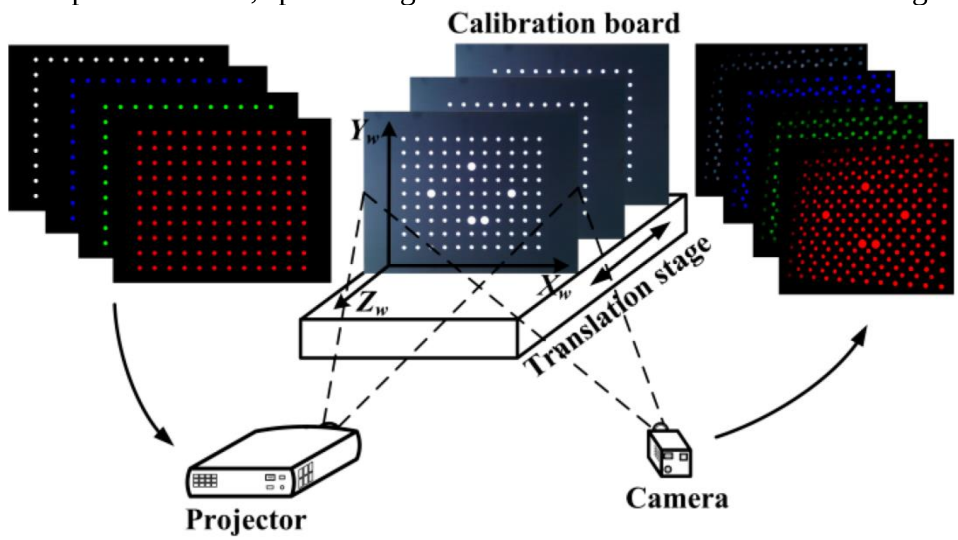

Figure 7- Structured Light

2.2.3.3 Stereophotogrammetry

Stereophotogrammetry incorporates capturing a 3D object from 2 distinctive coplanar planes to secure a 3D remaking of the pictures. This procedure has demonstrated to be viable in the face show. It notices to the private case with 2 cameras, orchestrated as a stereo pair, are utilized to recuperate 3D separations of highlights on the outside of the face. The procedure has been actualized clinically by utilizing a versatile stereometric camera. Ras et al have built up a stereo photogrammetric framework that presents the 3D directions of any picked facial milestones. This framework incorporates 2 synchronized semi-metric cameras introduced on a diagram with a separation of 50 cm among them and found merged with an angle of 15.

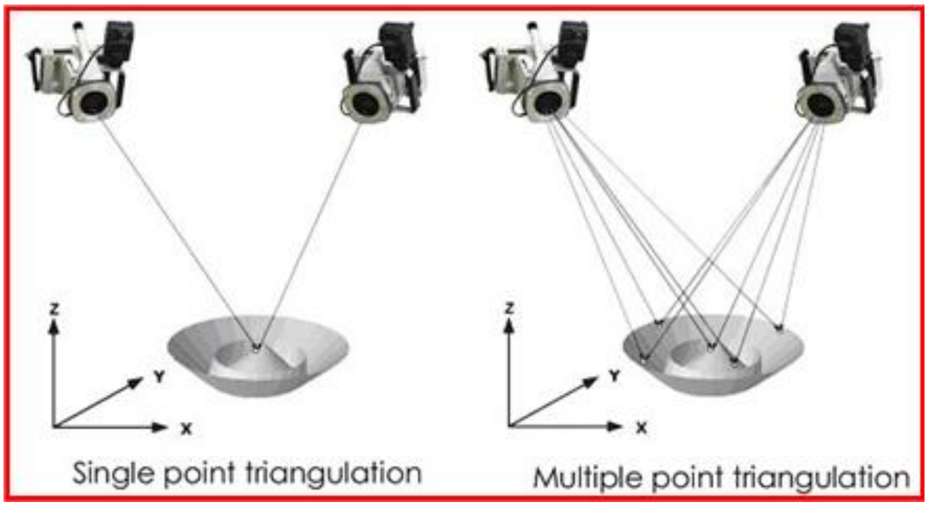

Figure 8- Stereophotogrammetry

Chapter 3
Software Req. Specification (SRS)

3.1 Introduction

The foundation of the Software Requirements Specification (SRS) gives an outline of the whole SRS with reason, scope, definitions, abbreviations, contractions, references, and diagram of the SRS. This document aims to present a detailed description of the project SfM 3D Scanner which uses Structure from Motion Technique to create a 3D point cloud of an object (size ~180 cm). The point by point necessities of the 3D Scanner is given in this document.

3.1.1 Purpose

This document covers the software requirement specifications for the project. The idea of the project is to produce a three-dimensional point cloud of an object using a single camera. The idea behind this project is to change the stereo vision system of producing 3D models to a monovision system. The software will take two-dimensional photos of an object and produce the three-dimensional point cloud by applying the Structure from Motion technique. This document is intended to lay out the highlights and prerequisites of the scanner, to fill in as a manual for the engineers on one hand and a product approval archive for the forthcoming customer on the other.

3.2 Overall Description

3.2.1 Product Perspective

The main idea behind the project is to reduce the cost of 3D scanners. Scanners that are present currently in the market uses LIDAR technology or stereo vision. LIDAR technology is very expensive and even after spending a lot of money maintenance cost is too much which make these scanners practically very difficult to be obtained by hospitals. Stereo Vision uses two cameras mounted separately taking different pictures to reconstruct a 3D model. Our goal is to develop an application that will use monocular images and use them for 3D reconstruction. This scanner can be used easily by a person after a week of training and will produce 3D models.

3.2.2 Product Functions

The main features of Structure from Motion 3D Scanner are highlighted below:

1. Images are uploaded to the application.
2. 3D reconstruction is done using the images.

3. Texture can be added to the point cloud that will increase the detail of the model.
4. The system will display the 3D point cloud of the object (human body).

3.2.3 User Classes and Characteristics

3.2.3.1 Summary of User Classes

The following section describes the types of users of the 3D Scanner. There are explanations of the user followed by the interactions the user(s) shall be able to make with the software.

3.2.3.2 Modeling Industry

The user intended to be targeted by our product is the ones in the modeling industry. The user will upload numerous photos of a human body (150-200), the system will generate a 3D point cloud of the object (mainly human body).

3.2.4 Operating environment

3.2.4.1 Hardware

The 3D scanner uses the following hardware

- **Digital Single Lens Reflex Camera:** The DSLR camera will be used to take photos of the object.
- **GeForce GTX 1080 Ti:** The graphics card will be used in fast processing for three-dimensional reconstruction.

3.2.4.2 Software

- Linux
- OpenCV, SciPy, OpenGL, NumPy
- IDE: Python IDE (python3)

3.2.5 Design and Implementation Constraints

- The scanner will only process images if the surroundings of the object are constant in every image i.e., objects in surroundings are not moved.
- Images will have surroundings and unimportant objects too.
- The object shall be of ~180 cm size (human size).

3.2.6 User Documentation

A user manual will be given to the clients in which separate guidelines will be offered by the specific user i.e., patients, gatekeepers, designers, and analyzers. It will incorporate the subtleties of the product working. Assist archives will also be a part of the framework. The undertaking report will likewise be accessible for the clients which will highlight the product's features, working, and methodology.

3.2.7 Assumptions and Dependencies

Surrounding objects shall not be moved, they must be in the same place in every photo. Images shall be taken with 6DoF poses and calibrated cameras.

3.3 External Interfaces Requirements

3.3.1 User Interfaces

The user will be able to view a three-dimensional point cloud of an object.

3.3.2 Hardware Interfaces

- Multiple images will be taken from the DSLR camera in real-time with a 40-70% overlap.
- The images will be used by application software for creating a 3D point cloud.

3.3.3 Software Interfaces

- OpenCV and SciPy will be used to process images.
- NumPy will be used in mathematical computations of the 3D point cloud.
- OpenGL will be used for the creation of 3D vector graphics.

3.4 System Features

This segment delineates organizing the functional requirements for the project SfM 3D Scanner by system features:

- **Image Acquisition:** Images are uploaded to the system.

- **Feature Detection & Matching:** Key points of images will be detected, and outlier points will be removed. Key points matching is done using an algorithm either SIFT, SURF, or ORB.

- **Triangulation:** The camera position will be estimated using two neighboring images.
- **3D Point Cloud Reconstruction:** The user is asked to input the algorithm with which they intend to calculate the 3D point cloud. The algorithm is used to compute the 3D point cloud of the intended object.
- **Display 3D point cloud:** Three-dimensional point cloud will be displayed on the screen after they are computed.

3.4.1 Image Acquisition

3.4.1.1 Description

This feature enables the system to acquire images from a camera set up on a tripod stand. These images will be fed into the system for further processing.

3.4.1.2 Stimulus/Response Sequences

3.4.1.2.1 Normal Path: Images sent for processing
Preconditions Images are selected for uploading.
Interactions The captured images will be sent to the system for processing.
Post Conditions Photos are uploaded successfully.
Categorization **Criticality**: High **Probability of Defects**: Medium **Risk**: High

Table 2: Image Acquisition

3.4.1.3 Functional Requirement

The system shall be able to acquire the images for further processing.

3.4.2 Features Detection

3.4.2.1 Description

This feature takes all images as input for further operations.

3.4.2.2 Stimulus/Request Sequence

3.4.2.2.1 Normal Path: Successfully transmitted to the system
Preconditions The images are loaded to the system for further processing.
Interactions The distinct features are detected.
Post Conditions The system can display the detected features of each image.
Categorization **Criticality**: Medium **Probability of Defects**: Medium **Risk**: Medium

Table 3: Features Detection

3.4.3 Key points Matching

3.4.3.1 Description

The images obtained from the previous stage will be used to find the key points of correspondence for every image pair.

3.4.3.2 Stimulus/Response Sequence

3.4.3.2.1 Normal Path: Key points matched
Preconditions Distinct Features are detected.
Interactions The same features are matched from two consecutive images.
Post Conditions The system can display the matched features of two images.
Categorization **Criticality**: High **Probability of Defects**: High **Risk**: High

Table 4: Key points Matching (a)

4.3.2.2 Exceptional Path: Key points not matched
Preconditions Some of the images have unmatched key points.
Interactions All neighbor points are not located at the same distance to each other.
Post Conditions Filter image pairs that have a small number of matched key points.
Categorization **Criticality**: High **Probability of Defects**: Low **Risk**: High

Table 5: Key points Matching (b)

3.4.3.3 Functional Requirements

The system will be able to display the matched features of all in pairs of two.

3.4.4 Triangulation

3.4.4.1 Description

The matched features are used to triangulate the position of the camera.

3.4.4.2 Sequence/Response Sequences

3.4.4.2.1 Normal Path: Estimation of Camera position
Preconditions Key points are matched.
Interaction Two matched images are used to estimate the camera position.
Post Conditions The system will be able to display the camera position.
Categorization **Frequency**: High **Criticality**: High **Probability of Defects**: High **Risk**: High

Table 6: Triangulation (a)

3.4.4.2.2 Exceptional Path: Camera position not matched
Preconditions There is an error in the camera matrix.
Interactions An error message is displayed
Post Conditions The camera matrix is to be calculated again.
Categorization **Criticality**: High **Probability of Defects**: Low **Risk**: High

Table 7: Triangulation (b)

3.4.4.3 Functional Requirement

The system shall be able to estimate the positions from where the images were taken.

3.4.5 Reconstruction of 3D points

3.4.5.1 Description

The estimated camera positions are used to generate the sparse and dense points of the object.

3.4.5.2 Sequence/Response Sequences

3.4.5.2.1 Normal Path: Reconstruction of 3D points
Preconditions The camera positions and camera matrix are known to the system.
Interaction Determine sparse and dense points from the given images.
Post Conditions 3D point cloud will be constructed and can be displayed by the system.
Categorization **Frequency**: High **Criticality**: High **Probability of Defects**: Medium **Risk**: High

Table 8: Reconstruction of 3D points(a)

3.4.5.2.2 Exceptional Path: 3D point cloud gives an irregular shape
Preconditions Camera positions and camera matrix are known to the system.
Interactions The correlation between key points is calculated.
Post Conditions Only sparse reconstruction
Categorization **Criticality**: High **Probability of Defects**: Low **Risk**: High

Table 9: Reconstruction of 3D points (b)

3.4.5.3 Functional Requirement

A complete 3D point cloud of an object (~180cm) shall be displayed by the system

3.4.6 Display 3D point cloud

3.4.6.1 Description

The system will be able to display the 3D point cloud created in previous steps.

3.4.6.2 Sequence/Response Sequences

3.4.6.2.1 Normal Path: 3D point cloud generated
Preconditions 3D point cloud generated.
Post Conditions The user will be able to see a regularly shaped and clear 3D point cloud of the human body.
Categorization **Frequency**: High **Criticality**: High **Probability of Defects**: Medium **Risk**: High

Table 10: Display 3D Point Cloud (a)

3.4.6.2.2 Exceptional Path: 3D point cloud gives an irregular shape
Preconditions The 3D point cloud is not generated
Interactions An error message is displayed
Post Conditions Only sparse reconstruction
Categorization **Criticality**: High **Probability of Defects**: Low **Risk**: High

Table 11: Display 3D Point Cloud (b)

3.4.6.3 Functional Requirement

A complete 3D point cloud of an object (~180cm) shall be displayed by the system which can be zoomed in or out.

3.5 Non-Functional Requirements

3.5.1 Performance Requirements

3.5.1.1 Response Time

The system shall be working within 1 minute of opening it.

3.5.1.2 Platform

The system application shall be compatible with Linux (Ubuntu/Debian).

3.5.1.3 Efficiency

The system shall be able to compute dense and sparse points in no more than 30 minutes.

3.5.2 Software Quality Attributes

3.5.2.1 Usability

The system shall have not more than 2 interface frames that can be interchanged with just one click.

3.5.2.2 Accuracy

The system shall provide 90% accuracy to make the project more useful for the modeling industry.

3.5.2.3 Legal

The system will follow customer privacy policy strictly.

3.5.2.4 Reliability

The system shall be able to work in a normal way after restarting due to an error.

3.5.2.5 Ease of Use

The experts will need training of one day to completely understand the system.

3.5.2.6 Modifiability

The system shall require no more than one week to resolve any error.

3.5.2.7 Operating Constraint

The system requires a graphics card for processing in 3D reconstruction.

Chapter 4
Design and Development

4.1 Introduction

The introduction of the Software Design Document (SDD) provides an overview of the entire SDD with purpose, scope, definitions, acronyms, abbreviations, references, and overview of the SDD. This document aims to present a detailed description of the project SfM 3D Scanner which uses Structure from Motion Technique to create a 3D point cloud of an object (size ~180 cm). The detailed designed structure of the 3D Scanner is provided in this document.

4.2 Purpose

This document covers the software design specifications for our project. The purpose of this document is to understand each component and module of the project. It will provide information about the relationship between each module and how they are interconnected. The report is expected to educate partners regarding the subtleties of the structure and the plan procedure. It is intended to outline the features, structure, and engineering of a 3D scanner, to fill in as a manual for designers also, the target group.

4.3 Project Scope

The project will help the designer who intends to create a three-dimensional model of the model. The expert will take photos of an object from different angles, the photos will have a 40-70% overlap. The project has three basic modules. The first module includes taking photos of a static object from different angles. The second module is to apply the SfM technique on the photos. The third module includes the generation of the 3D point cloud of the object. The 3D model will be used to check when the designer intends to see how the newly designed clothes will look on the model.

4.4 System Architecture Description

This section provides a detailed system architecture of SoloScan overview of system modules, their structure and relationships are described in this section. User interfaces and related issues are also discussed.

4.4.1 Structure and Relationships

This section covers the overall technical description of SoloScan. It shows the working of application from the perspective of different point-of-views and shows a relationship between different components.

System Block Diagram

The diagram(s) show the higher-level description of the application(s), generic working of the application(s), and interaction with the user.

A user interacts with the web page online. The system reads user input images and applies SfM techniques to create a 3D point cloud of an object.

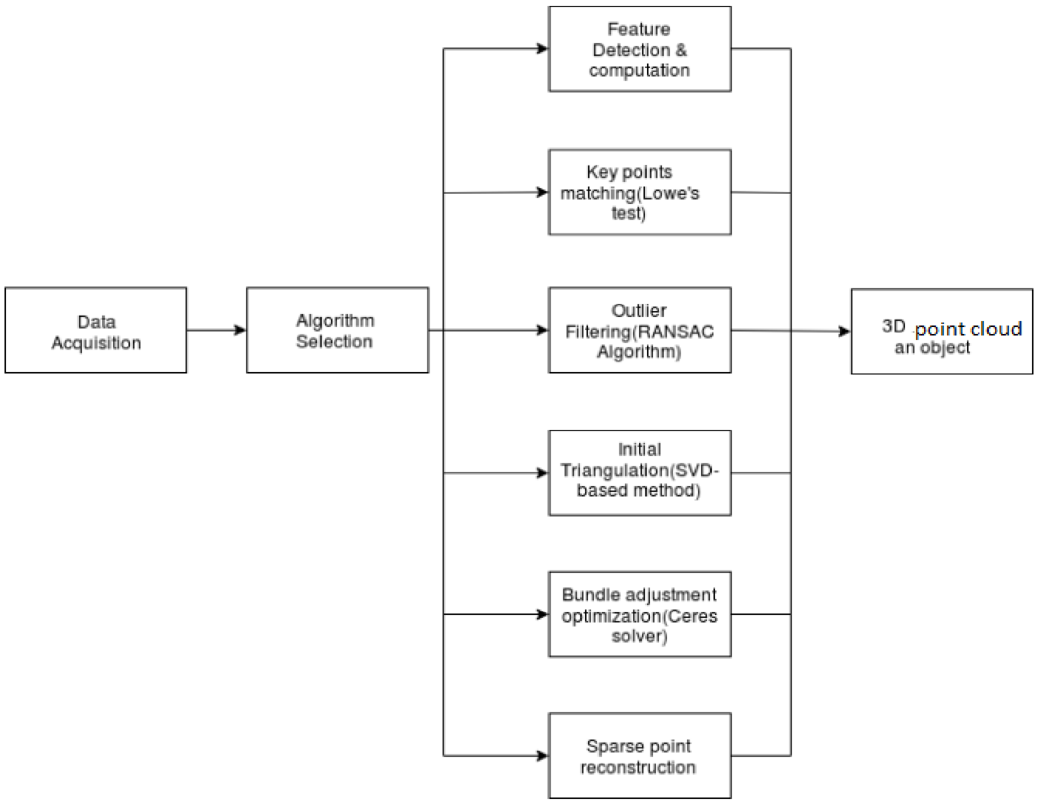

Figure 9-Block Diagram

4.4.2 Design Rationale

After the acquisition of data, the system will use an inbuilt algorithm for feature extraction. Following techniques will be used for 3D reconstruction:

1. Feature detection (SIFT algorithm)
2. Key points matching (Lowe's ratio test).
3. Outlier filtering (RANSAC algorithm)
4. Initial triangulation (OpenCV function)
5. Construction of 3D point cloud (OpenGL)
6. Display 3D point cloud

4.5 Data Design

4.5.1 Data Description

The information domains of our system will be passed from file storage at the backend of the program saved by the system. When a new project will start to convert 2D images into a 3D point cloud, a new folder will be created, and all upcoming files will be stored in that folder. When the system asks for any data or information to process it can access that folder and can take the data from the files stored at that place. The basic file handling will be used to save and use the data in the program. At the stage of the creation of the new project of the generation of the 3D point cloud, the files will be created at located space. Whenever the system has to store any data/ information it will access files from a specific location and store the particular type of data in file format. Similarly, when the system needs any data to manipulate and use in the next step it will access the files and fetch the data from those files, process the data, and again store it in the form of file storage at a given location.

There is no database used in the whole process. The system just needs the file handling to store and fetch the data for further processing.

4.5.2 Data Dictionary

The data structure is stored in the form of data storage files and folders. Initially when the project is created a folder is made at the default place along with the five folders in which all the data structure is going to be used in the future. Then step by step the data structure will be stored and fetched wherever it will be needed.

The system entities along with types and description are as follows:

System entities	Input parameter	Output parameter
Image acquisition functionalities:	no input data required	images array
Load images()	-	images array
CheckExtension()	Images array	string
CheckQuantity()	Images array	integer
SaveImages()	Images array	-
MoveToNext()	Reference of next window	-
Feature detection functionalities:	Input images array	Featured points array
ShowMenu()	-	-
KeyPointExtraction()	-	-
GetImageArray()	-	Images array
Cv2.xfeature2D.sift.create()	-	Sift object
DetectandCompile()	images,sift,interpolation	Integer
SavefeaturedArray()	images array	-
Showkeypoints ()	featured points array	-

Key points matching functionalities:	Feature points array	Key points match and percentage match
Keypointsmatching ()	-	-
Getfeaturedarray()	-	Featured points array
Matchkeypoints()	Featured points array	Integers matrices
Getoverlap()	Key points array	Integer
Displaymatches()	Key points array	-
Savematches()	Key points array	-
Triangulation functionalities:	Images array	Triangular and distance matrices of images
Triangulation()	-	
Getimagearray()	-	Images array
Gettriangularcomponents()	Images array	Triangular array matrices
Getangles()	Images array	Angle array matrices
Savetriangulationmatrices()	Distance matrices	-
Showdistancematrices()	Distances matrices	-
Sparse point calculation functionalities:	Triangulation matrices and images array	Sparse matrices of all images
Sparsepointcalculation	-	
Gethmatches	-	-
Getimagesarray	-	Matches array
Getoverlappedpairs	Images array	Images array
Gettiepointcloud	Overlapped pairs array	Overlapped pairs array
Displaysparsepointcloud	Sparse point cloud matrices	-
Getdepthmaps	Triangular matrices, images array	-
Getdensepointcloud		Depth matrices
displaydensepointcloud	Depth matrices	Dense point cloud matrices
Mergedensepointcloud	Dense point cloud matrices	-
Save3Dpointcloud	Dense point cloud matrices	3D point cloud(.PCL)
	3D point cloud	-
Display 3D point cloud functionalities:	Sparse triangulation matrices and sparse matrices	3D point cloud
Get3Dpointcloud	-	3D point cloud matrices
Display3Dpointcloud	3D point cloud matrices (.PCL)	-

Table 12: Data Dictionary

Description:

At the beginning of the program when the project will create the folder. Five subfile will be created at default system memory for each step where other files of the program are saved. All the subfolders will be named according to their functionalities.

Image acquisition

The user will be asked to load the images. The user will browse the images and proceed to the next step. The input images will be taken one by one in an array, check the input data type. If they are PNG or JPG files, the array will be saved in the backend storage at the image acquisition folder in the form Files in the project folder for further fetching and using in further processes.

Feature detection

The system will access the JPG images files from the image acquisition folder and get the images from the folder in the form of an array named image array. Then the system will extract the features of all the images of an array by the method detectAndCompile() which takes the input of an image array and a SIFT object by which it extracts the features of all the images.

Finally, in this step, we will get the featured array of images.

When the featured points of all the images will be obtained the system will make an array of all the images and store in the folder of feature detection backend in the form files at the project folder so that it may use in further processes.

Key points matching

The system will access the files of the project folder in the feature detection subfolder and fetches the featured points array file. Next, it matches the features of all the images one by one sequentially and gets percentage matches. The system will make an array of all matched features. Then the system stores it in the form of a matrix with percentage matching of respective images in the folder of key point matches at the project folder for further usage and fetching.

Triangulation

In this step, the system fetches the images array from the image acquisition folder of the project. Then the system will get the angles and all the triangular components of images in the images array. After making the triangulation matrices again the system will store the triangular components and distance matrices of images backend in the form of the file at the triangulation folder in the project.

3D point cloud generation

In this step, the system will get the triangulation matrices, and matched featured array of all the images from the file storage and get the maximum overlapping pairs of images to make the sparse matrices of the images. The next step depth of the image pairs will be calculated by taking the triangulation matrices and angles of the images with their matched pairs as at an input. The system will then design the dense point cloud matrices of the images. Then all the dense point cloud will be merged to make a 3D structured point cloud of the object. The 3D point cloud structure will be saved in the backend in 3D point cloud folder located at the project folder

Display 3D point cloud

In this step, the system will fetch the 3D point cloud matrix of the images display 3D point cloud of the object from the folder of the 3D point cloud. The 3D point cloud finally will be displayed to the user.

4.6 Component Design

Conceptual Architecture is "Context" for the system's use.

4.6.1 Overview of modules/components

This subsection will present different segments and subsystems. The acquisition of data, that is in the form of images is obtained using our input component and processed by our Model Application Component, which applies some computation to optimize the data and send it to our output component to produce a final visualization of the 3D point cloud.

4.6.1.1 Input Component

The input module will take multiple 2D images in either ".png" or ".jpeg" format and pass the dataset to the Model Application component.

4.6.1.2 Model Application Component

The model application component analyzes the data and removes outliers and sends the optimized data to the output module.

4.6.1.3 Output Module

The output module will give the visualization of the 3D point cloud.

4.6.2 Component Diagram

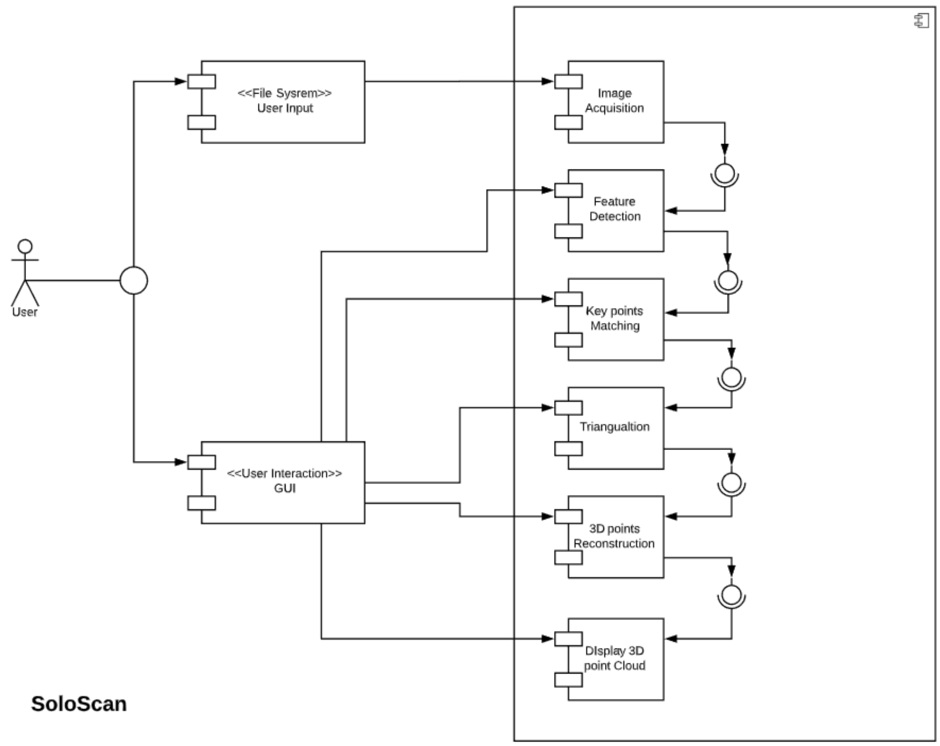

Figure 10-Component Diagram

4.6.3 State transition diagram:

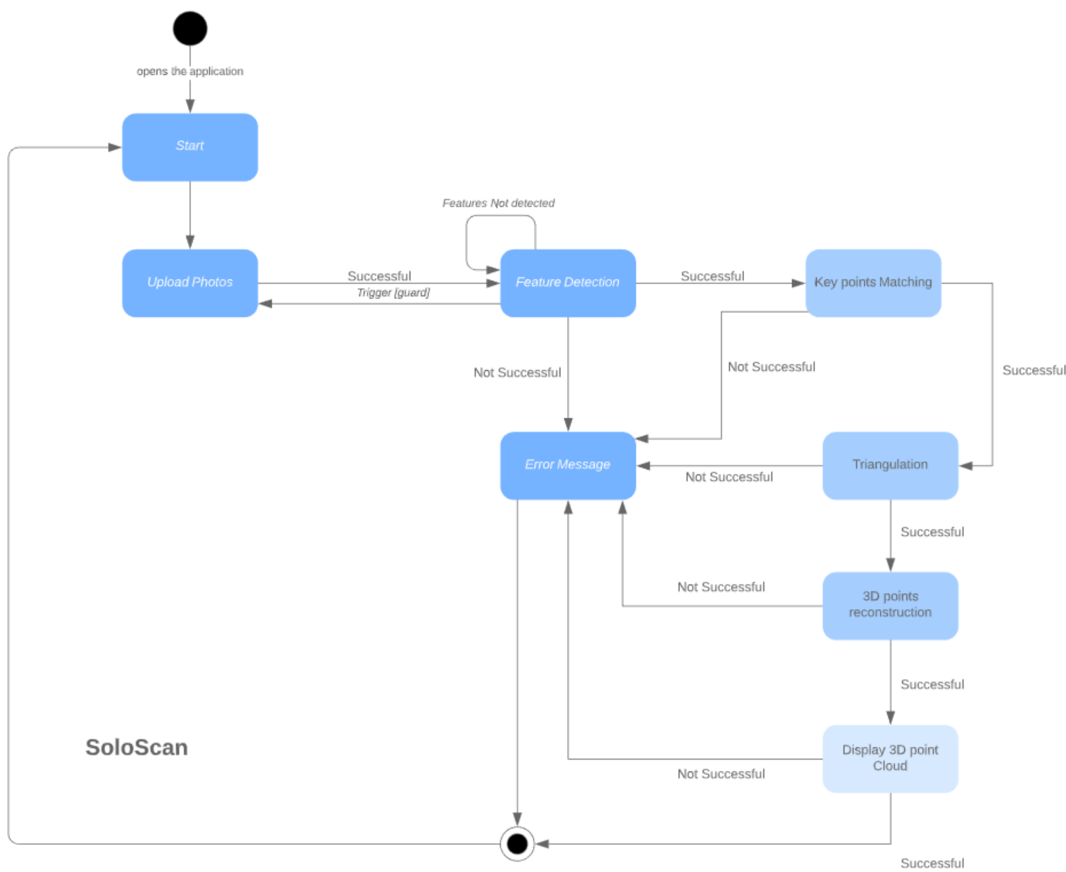

Figure 11-State Diagram

4.6.4 Use Case Diagram:

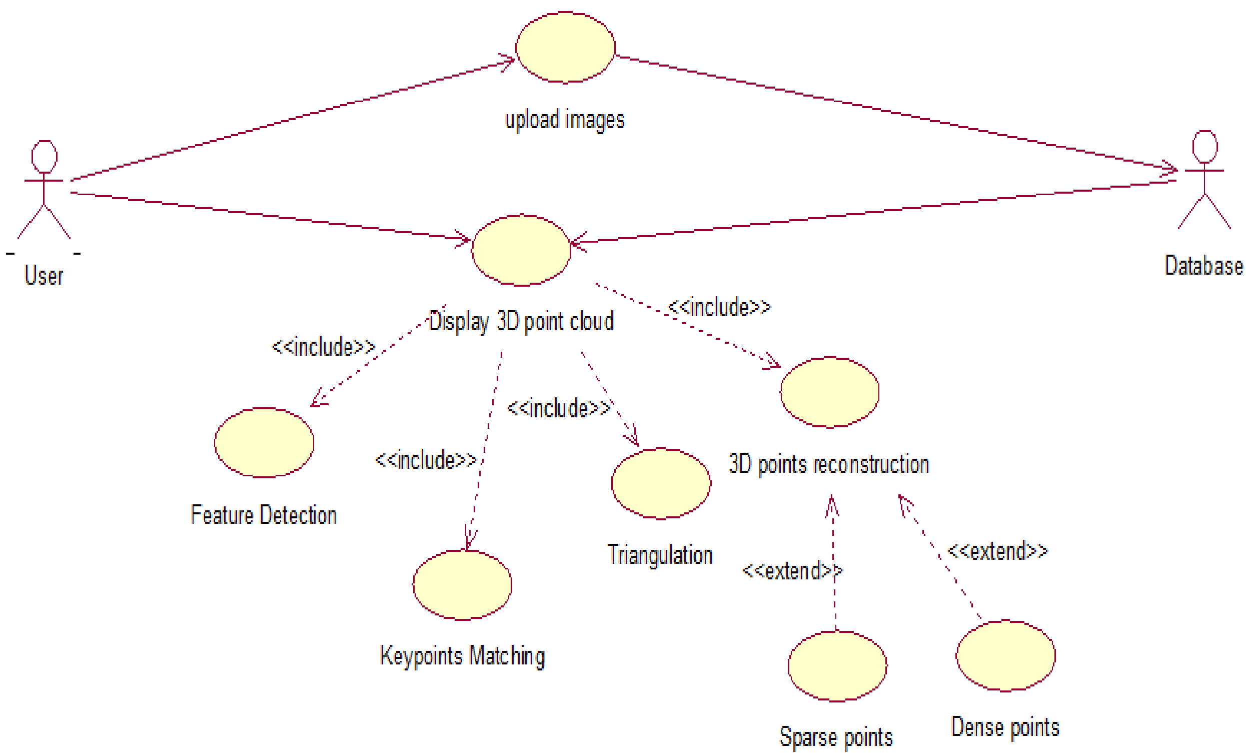

Figure 12-Use Case Diagram

4.6.5 Activity Diagram:

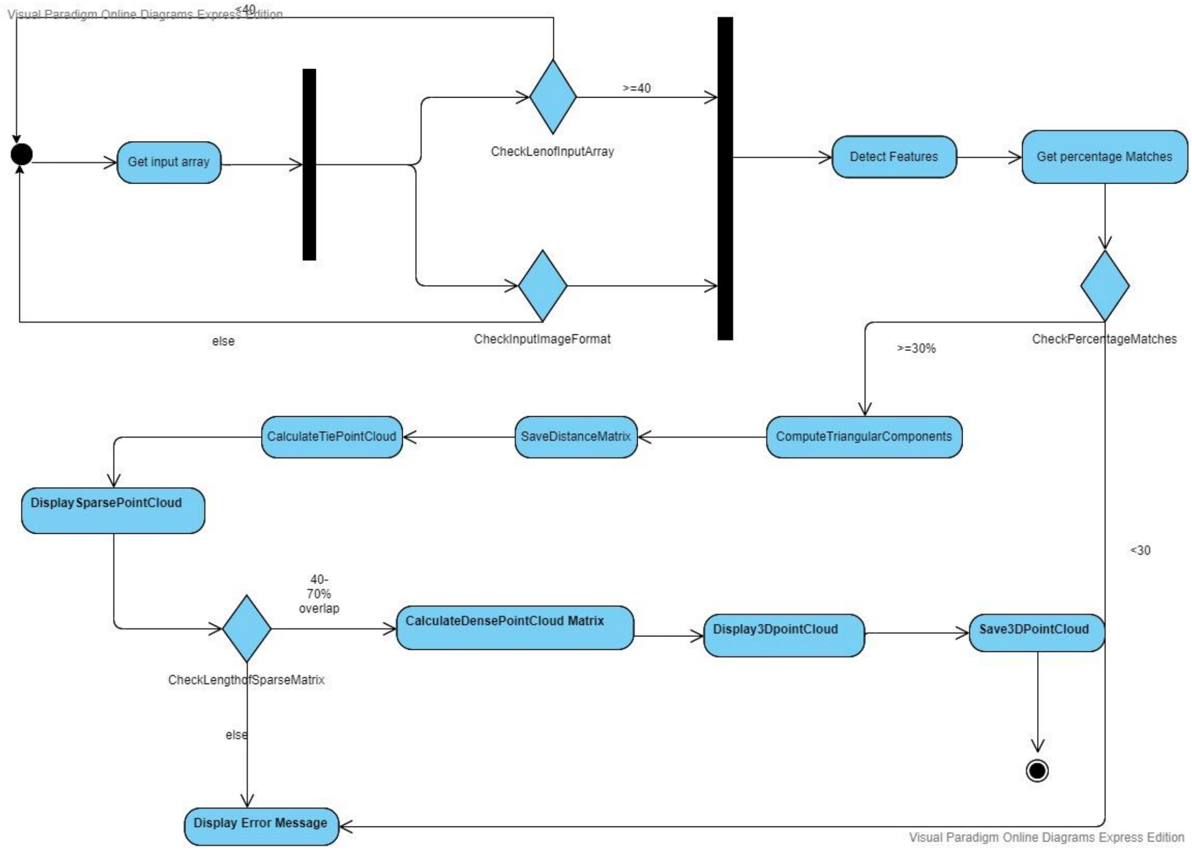

Figure 13-Activity Diagram

4.6.6 Sequence Diagram:

4.6.6.1 Image Acquisition

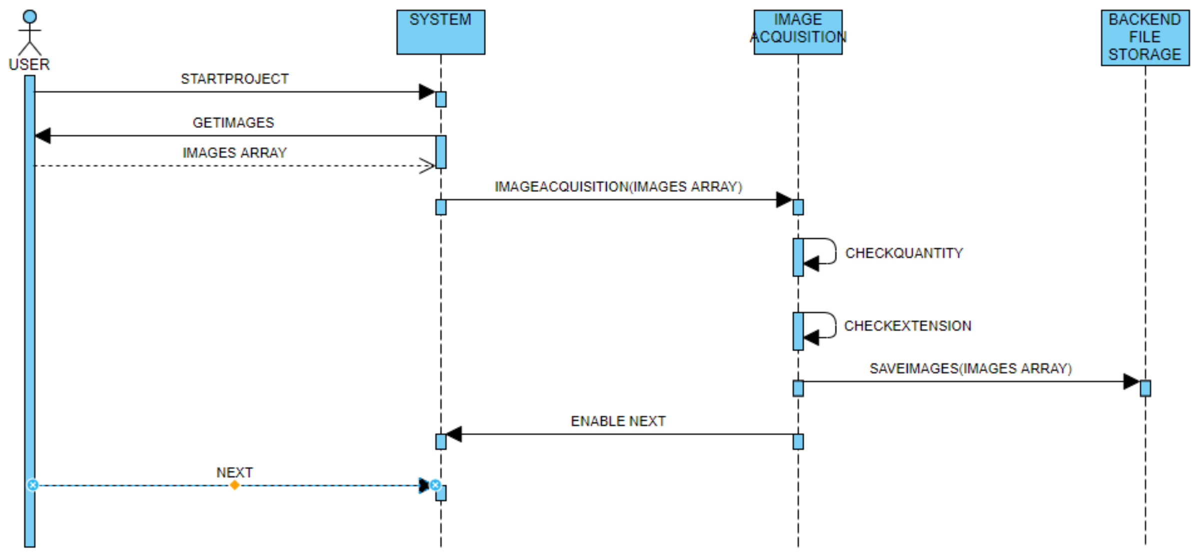

Figure 14-Sequence Diagram (Image Acquisition)

4.6.6.2 Feature Detection

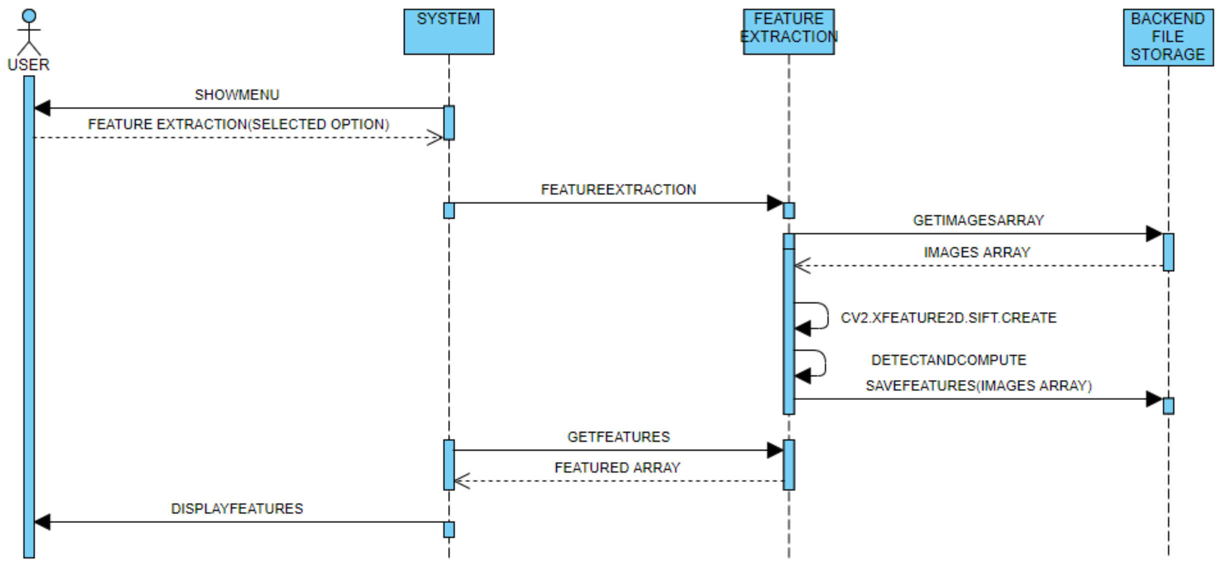

Figure 15-Sequence Diagram (Feature Detection)

4.6.6.3 Key Points Matching

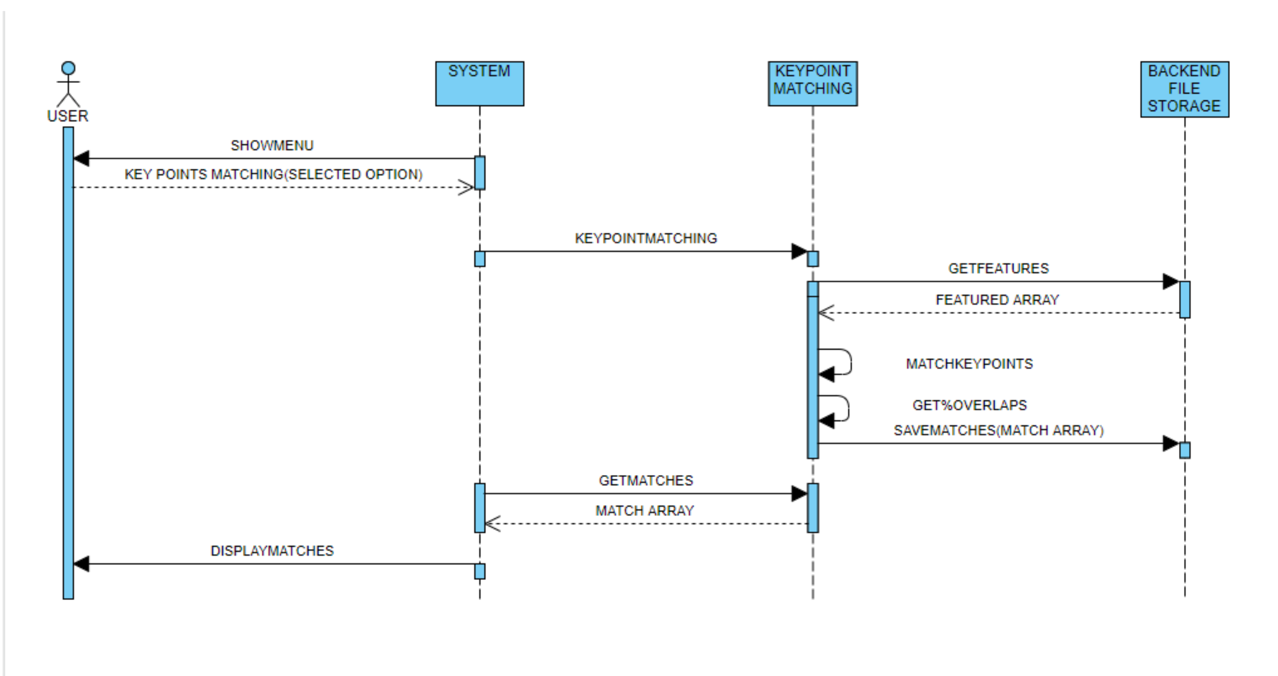

Figure 16-Sequence Diagram (Key points matching)

4.6.6.4 Triangulation

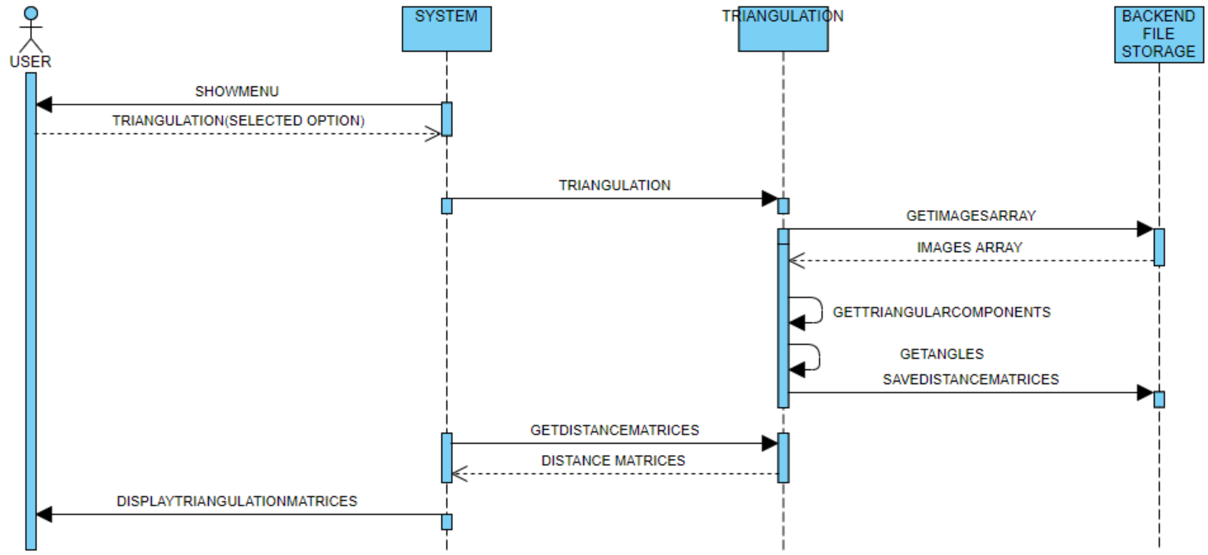

Figure 17-Sequence Diagram (Triangulation)

4.6.6.5 3D point Cloud Generation

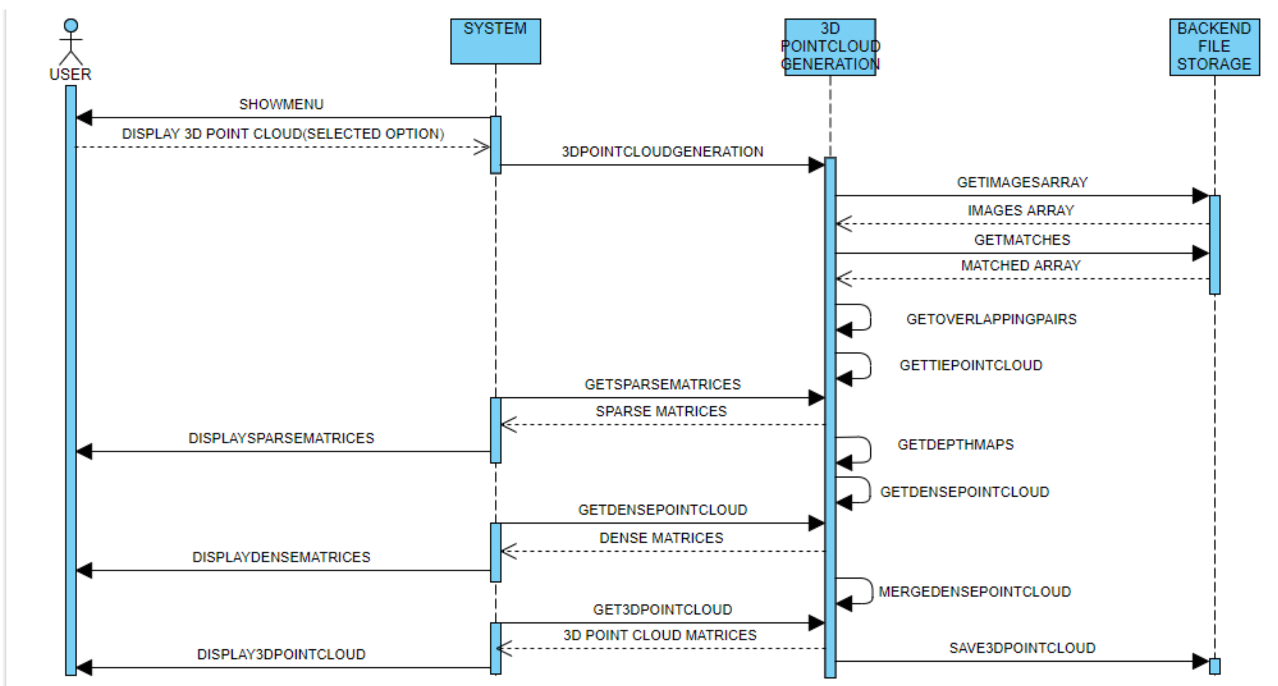

Figure 18-Sequence Diagram (3D point cloud generation)

4.6.6.6 Display 3D point Cloud

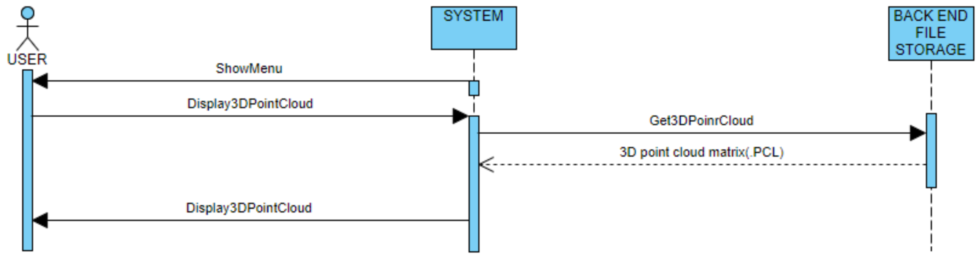

Figure 19-Sequence Diagram (Display 3d Point cloud)

4.7 Class Diagram:

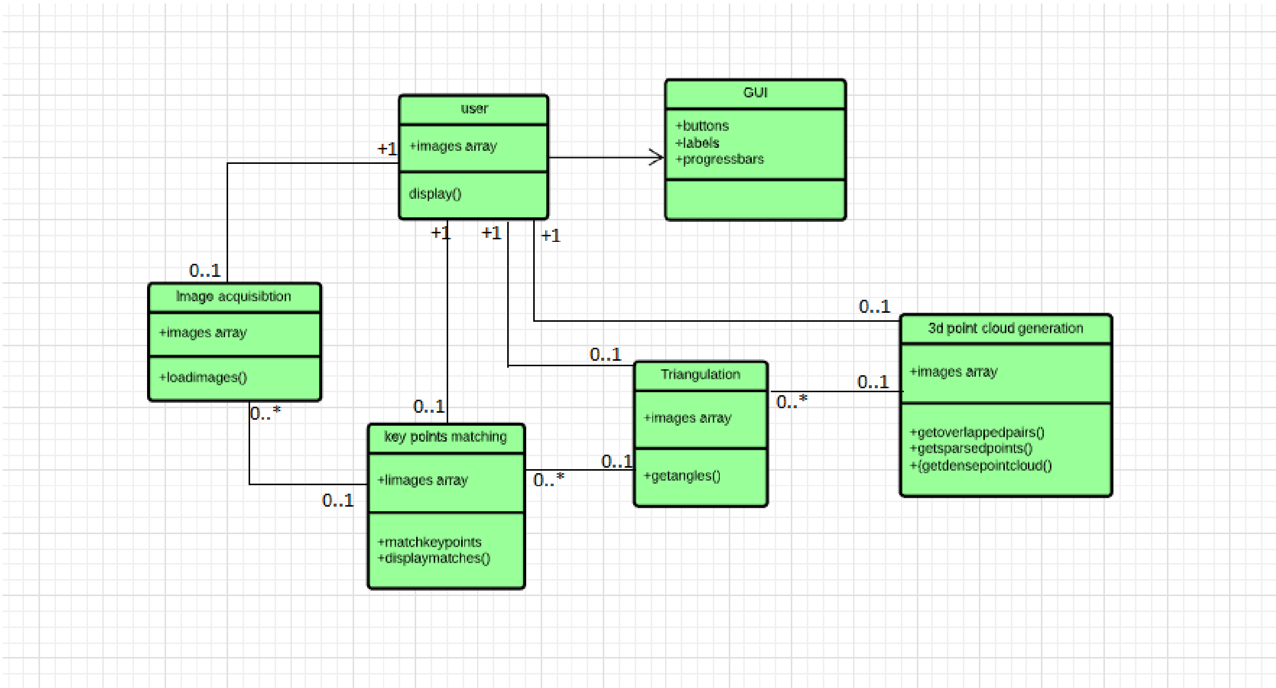

Figure 20-Class Diagram

Description:

Class Description

Name	Description
Graphical User Interface	It is the main class of the system, the front end through which the user shall interact with the system.
File handler	It is the data store containing the trained data i.e. 3D point clouds of objects. The input images are mapped against this handler.
User	This is a user class where users can interact with the system and see what features are available to use and what are not.
Feature Extraction	This class involves getting all the images having 40-60% overlap and check their feature points. Feature points are checked using the SIFT algorithm. Feature points will be displayed to the user.
Key points Matching	It involves the matching of key points in every image. Matched key points can also be viewed by a user if he/she wants to.
Triangulation	Camera angles are being triangulated here. The distance matrix is used for this purpose.
Sparse Point Cloud	This class gives an overview of the sparse point cloud.
Outlier points	This class gives detail about the removal of outlier points from the dataset.
Dense Point Cloud	It is an output class. It shows the 3D point cloud of an object.

Table 13: Class Description

4.8 Reuse and Relationships with other Products

SoloScan is not based on any previous systems neither it's an extension of any other applications at any level. But it can be evolved into a bigger and more complex system with more features and functionality. Developers can also reuse some of the modules of the system. The application can also be enhanced to further include more activities. It can also be further enhanced by developing an android application to create 3D models n to make it more impressive.

4.9 Human Interface Design

4.9.1 Overview of User Interface

Our user interface consists of three major interfaces:

1. Image input or image uploading by the user

2. Loading images

3. Functionalities user can perform

The user will be able to observe the output or result of each step.

First interface page

In the first interface page, the user is allowed to load the images by choosing the images files as shown in the figure below;

Figure 21-Interface page 1

After image acquisition, whenever the user clicks the “NEXT” button it will take him/her to the new window which shows the loading of images.

Second interface page

The loading windows show the processing of images being uploaded to the system for the creation of a 3D point cloud as shown in the figure below;

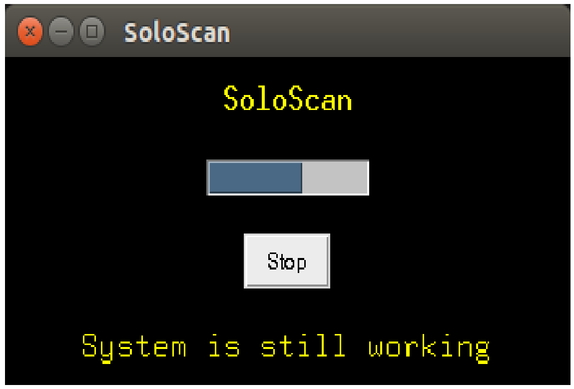

Figure 22-Interface page 2

Once the loading is complete, a 3D point cloud will be displayed on the next window. It will open a new window containing all the functionalities of software product

Third interface page

On the third interface, the user will be shown four options that are performed sequentially by the user as shown in the figure below;

Figure 23-Interface page 3

1. Feature detection
2. Key points matching

3. Triangulation
4. Display 3D point cloud

The user will be able to see the output of every step on the screen. After completion of each step, the user performs the output generated and displayed on the screen to the user. On the backend, the result is saved in the files.

Description:

- **Feature detection**

In this option, the features of all the images will be extracted and detected. When the user selects the option, images being uploaded at a previous interface will be processed one by one and the system will extract the feature points. It will then display the feature points to the user.

- **Key points matching**

In this step, all the feature points/feature points are being matched sequentially. After the selection of this option, the system will match the features of all the images sequentially and find the percentage of similarity. The system will check the matching points whether they fulfill the SfM requirement i.e. 40-70 percent overlapping. If the images fulfill the requirement, the system will display all the matched key points to the user and save them at the backend. The matched key points of the images will be displayed on the screen and the user will be asked to proceed for the next step.

- **Triangulation**

In this step, the triangulation matrix of all the images will be made by epipolar geometry and the triangular matrices of all the images will be saved on backend. The angles and the triangular components of the images will be extracted and displayed to the user. Then the user will be asked for further steps.

- **Display 3D point cloud:**

The user when performing this functionality, the system fetches the 3D point cloud of the object and displays it to the user on the screen. It will be the final step of the program.

Chapter 5
Project Test and Evaluation

5.1 Model Evaluation:

In this segment of theory, we are going to introduce our model Evaluation calculation of structure from motion, which navigates the model diagram to find the consistent nodes for 3D recreation of pictures. This calculation begins from a known reliable node and it navigates the diagram along with different ways until all the M input pictures are covered individually. During the procedure, the model in the beginning node bit by bit develops to fit the pictures in resulting groups.

5.1.2 Model development:

During the procedure, the model in the beginning node bit by bit develops to fit the pictures in resulting groups.

5.1.3 Consistency of Models:

Consistency is the essential however not a sufficient condition to reconstruct right 3D models. Think about an extreme case, where the pictures are in a clustered shape share a similar perspective, a reliable model can be still produced however doesn't look like the underlying ground truth model of the group.

5.1.4 Reliability of Models:

In our structure, we require a predictable model to be reliable. Every 3D point will be obvious to at least M cameras, and all pairwise edges between these cameras are inside a predefined extend.

5.1.5 Model Creation:

Given that reliable clusters as the beginning of the node, we first utilize incremental rigid SFM procedure to reconstruct the 3D model. We play out the projective recreations, trailed by the auto-alignment accepting the zero-skew, unit angle proportion, and principle point at the starting point. At long last, we play out the matric bundle adjustment to recuperate the 3D model for the beginning node.

5.1.6 Model Acceptance:

Accepting that a solitary reliable cluster as a beginning stage isn't prohibitive presumption in practice. For instance, working with faces, we can easily find multiple pictures of one individual's face; working with human body enunciation, we can generally find numerous pictures of common posture.

5.1.7 Model Reduction:

After Evaluation, each picture is related to at least one 3D model rely upon what number of visited predictable clusters it has a place with. A significant number of these models will be comparative. It is alluring to lessen the number of models; both speak to the picture assortment and to appraise a one of a kind of models for each picture. we depict a coarse method to decrease the models, which will be utilized a beginning stage for the algorithm.

5.1.8 Model Reconstruction Algorithm:

Specifically, we first utilized a basic K-mean calculation to partition reconstruction 3D models into K gatherings, where K is as of now set by the user and depends on how much the objective item disfigures and explains, i.e., an article that distorts significantly will require a bigger number of bases than an almost unbending article. The mean states of each gathering fill in as K basis shapes for all the remade 3D models.

5.1.9 Model Assessment:

For each picture, we scanned for the best-fit premise shape, which has the base normal projection blunder in present estimation as for the picture. We at that point relegate this premise shape and the assessed posture to the picture. Utilizing these premise shapes and assessed acts like an underlying worth, we next depict a progressively exact answer for 3D model reduction.

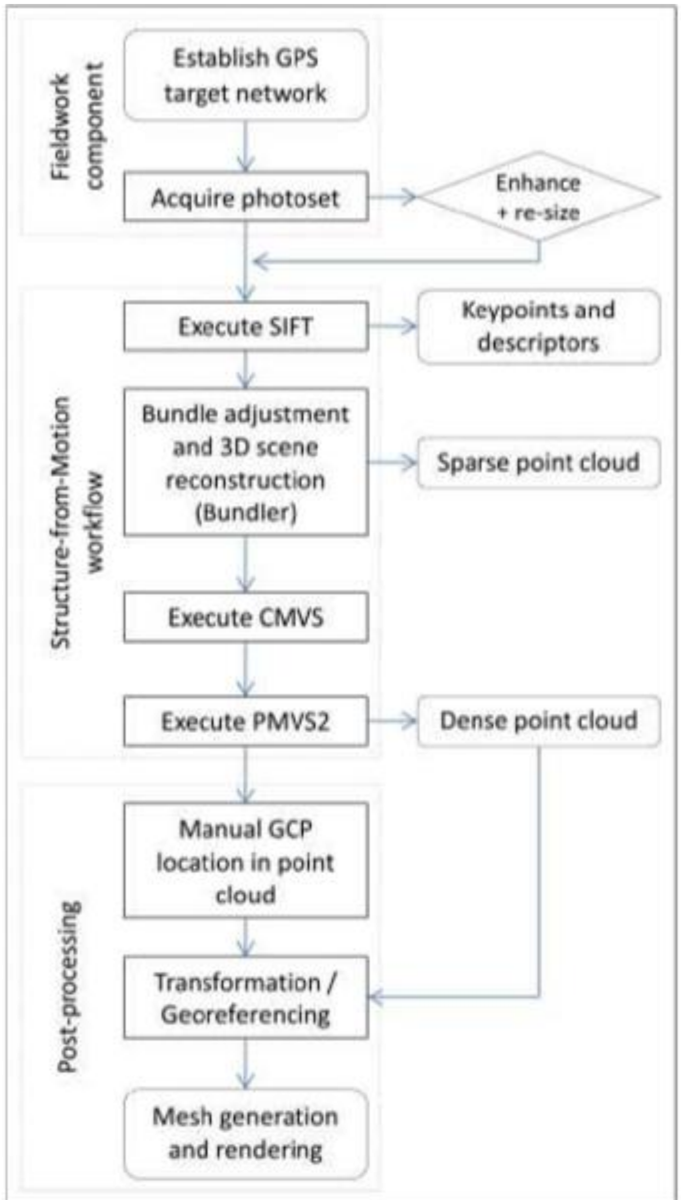

Figure 24- Model Assessment

5.1.9.1 Data Acquisition - Digital images

The sample of pictures with various directions and positions taken utilizing an advanced camera were stacked into the system. The securing of computerized pictures thinks about the ideal inclusion and in a perfect world ought to be taken at various positions and directions to accomplish sound system pictures. The product permits the stacking of handfuls or several photographs. In this demonstration just about numerous photographs were used. The setup of the pictures influences the consequence of the 3D recreation, along these lines the area of taking pictures should be very much orchestrated, if conceivable. When the stacking of the pictures was finished, a brief showed up in the Task Viewer (Log Window) on the privilege of show interface demonstrating culmination.

5.1.9.2 Automatic relative orientation and matching of images:

The images loaded then relatively oriented by use of image matching function along with which computes pairwise matching, for feature detection for each image pair at a time using the SIFT algorithm. To find homologous features the SIFT operator was used.

5.1.9.3 Sparse reconstruction of point cloud model:

In this step, the relevant image matches between the photos were now calculated for their 3D positions in a relative coordinate system. This process is a prerequisite of the dense reconstruction but normally quicker than the dense reconstruction. After acquiring the common features by the SIFT operator, the bundle block adjustment was carried out. This delivered the relative position and orientation of images and sparse surface of mountain peak was obtained. Pairwise Feature-Based Image Matching for GPU (SIFTGPU) was used to obtain the sparse 3D reconstruction in model space. The RANSAC (RANDOM SAMPLE CONSENSUS) robustness estimation was used for filtering mismatches. The general form of the reconstruction was visible and the calculated points, camera positions, and image planes in 3D were viewed.

5.1.9.4 Automatic reconstruction of dense point cloud model:

The sparse surface would need to be adequate to do the 3D displaying as the preparing steps require a thick point haze of the surface. The reproduced (point cloud) information was then spared and given an ideal envelope document, inside a made organizer inside a similar catalog as the picture records area to which the model made was put. When the document name was picked, the Task Viewer (Log Window) started to show the general procedure of thick reproduction. This procedure was tedious as it requires a great deal of preparing memory and an opportunity to finish for the most part shifts from a few seconds or minutes for just a couple of pictures as for this situation to a few hours for enormous datasets, additionally relying upon the equipment (PC) abilities. When finished, the outcome was a thickly remade point cloud. Further, inside the organizer made where the record was spared had a 'models' envelope inside the fundamental registry where all naturally made thick reproductions were set in 'utilize' arrangement and PCL format.

Figure 25- Dense reconstruction of portions

The program attempts to match all the photographs, yet relying upon how the photographs were taken, any regions that can't be coordinated may cause the divided arrangement of models and various models of thick recreations might be made in organizers with successive naming of 01, 02, … and so on.

5.1.10 Evaluation workflow

A. Image acquisition and key point extraction:

The key issue that SfM addresses is the assurance of the 3D area of coordinating highlights in numerous photos, taken from various edges. The underlying handling step in the answer for this issue is the recognizable proof of highlights in singular pictures which might be utilized for picture correspondence. A well-known answer for this, and utilized in the strategies is the SIFT (Scale Invariant Feature Transform) object acknowledgment framework, permitting comparing highlights to be coordinated particularly even with huge varieties in scale, perspective (direction) and understates of fractional impediment and evolving brightening (top) Lines speak to individual key points, relatively scaled by the range of the picture district (pixels) containing the key point.
Focal points, or 'key points', are consequently recognized over all scales and areas in each picture, trailed by the making of a component descriptor, figured by changing neighborhood picture slopes into a portrayal that is to a great extent unfeeling toward varieties in enlightenment and direction. These descriptors are sufficiently exceptional to permit highlights to be coordinated in huge datasets.
The quantity of key points in a picture is reliant basically on picture surfaces and goals, with the end goal that mind-boggling pictures at high (regularly unique) goals will restore the most outcomes. The thickness, sharpness, and goals of the photoset, joined with the

scope of normal scene surfaces will, in the principal occurrence subsequently, decide the nature of the yield point cloud information. So also, diminishing the separation between the camera and highlight of intrigue, in this way expanding the spatial goals of the photo, will upgrade the spatial thickness and goals of the last point cloud. Varieties in the unpredictability, lighting, materials in singular scenes all impact the picture surface so it is difficult to offer unequivocal direction on the base number of photos important for scene recreation. The base necessity is for relating highlights to be noticeable in at least three photos; be that as it may, getting however many pictures for SfM contribution as could be expected under the circumstances, given strategic imperatives, is energetically suggested as this improves a definitive number of key point matches and framework excess.

Specific thought ought to likewise be given to the securing stage. For instance, little scope locales with steep incline points are probably going to be more qualified to an only ground-based methodology, while low elevation aeronautical photography (LAAP) may give better inclusion over bigger destinations and those with increasingly repressed geology. Without a doubt, symbolism consolidated from numerous stages may end up being ideal, giving various degrees of detail in various regions of the scene.

A wide assortment of imaging sensors can be utilized for SfM extending from still recordings to second rate conservative computerized cameras. The center necessity is to accomplish all-around uncovered photos of the feature(s) of intrigue. While picture quality and goals are improved by utilizing progressively costly computerized SLR models, pictures taken at the goals (for example >12 megapixels) will be re-sized (bringing about resulting loss of picture detail) to evade longer occasions for preparing.

B. 3D scene reconstruction

After the key point recognizable proof and descriptor task, an inadequate group change framework Bundler some other SfM arrangement of inclination is applied to gauge the camera present (boundaries and positions) and concentrate a low-thickness or 'scanty' point cloud. Key points in different pictures are coordinated utilizing surmised closest neighbor and Random Sample Consensus (RANSAC calculations, and 'tracks' connecting explicit key focuses in a lot of pictures are set up. Tracks involving at least two key focuses and three pictures are utilized for point-cloud remaking; with those which neglect to meet these rules being naturally disposed of. Utilizing this technique, transient highlights, for example, individuals moving over the region of intrigue are consequently expelled from the dataset before 3D reproduction starts. This additionally applies to non-static articles inadvertently caught in the information photograph set, for example, questions on the sensor, stages, for example, dirigible. Under these occurrences, albeit indistinguishable key focuses referencing such articles will be made, they are not reasonable for scene recreation as their position comparative with other key focuses is continually changing and are naturally sifted utilizing the perceivability and regularization requirements.

Key point correspondences place requirements on-camera present direction, which is remade utilizing a closeness change, while minimization of mistakes is accomplished utilizing a non-straight least-squares arrangement. Robotized camera focal point alignment is likewise led – self-adjustment. At long last, triangulation is utilized to assess the 3D point positions and steadily remake scene geometry, fixed into a relative arrange framework. Full

computerization of this procedure, from key point extraction, to precise reproduction of scene geometry, is an away from of the SfM strategy over conventional advanced photogrammetric approaches.
Inadequate point mists are created whose thickness can be upgraded by executing the Clustering View for Multi-View Stereo (CMVS) and Patch-based Multi-see Stereo (PMVS2) calculations. With camera positions utilized as information, CMVS at that point decays covering input pictures into subsets or bunches of reasonable size, while PMVS2 is utilized to autonomously reproduce 3D information from these individual groups. The aftereffect of this extra handling is a noteworthy increment in point thickness.

C. Post-processing and DEM generation

Since the point cloud is produced in a family member/discretionary 'image space' arrange framework, it must be geo-referenced to a real-world, 'object-space' organize a framework for study applications. Change of SfM picture space directions to a land/object-space arrange framework can be accomplished utilizing a 3D closeness change dependent on a few ground-controls focus (GCPs) with known item space organizes. The 3D, seven boundary change (turns, interpretations, and a scale factor) utilized requires at least three GCPs (coordinating picture and article space arranges). Such GCPs can be inferred post-hoc, distinguishing competitors include unmistakably obvious in both the subsequent point cloud and in the field, and getting their directions by ground study (i.e., by the GPS). The other geo-referencing approach is immediate geo-referencing where picture obtaining areas are gathered with on-board GPS during flight bringing about low spatial precision of a couple of meters.

5.2 Testing:

The testing phase is carried out. There are many test scenarios of SfM which is carried out to check the credibility or the project.

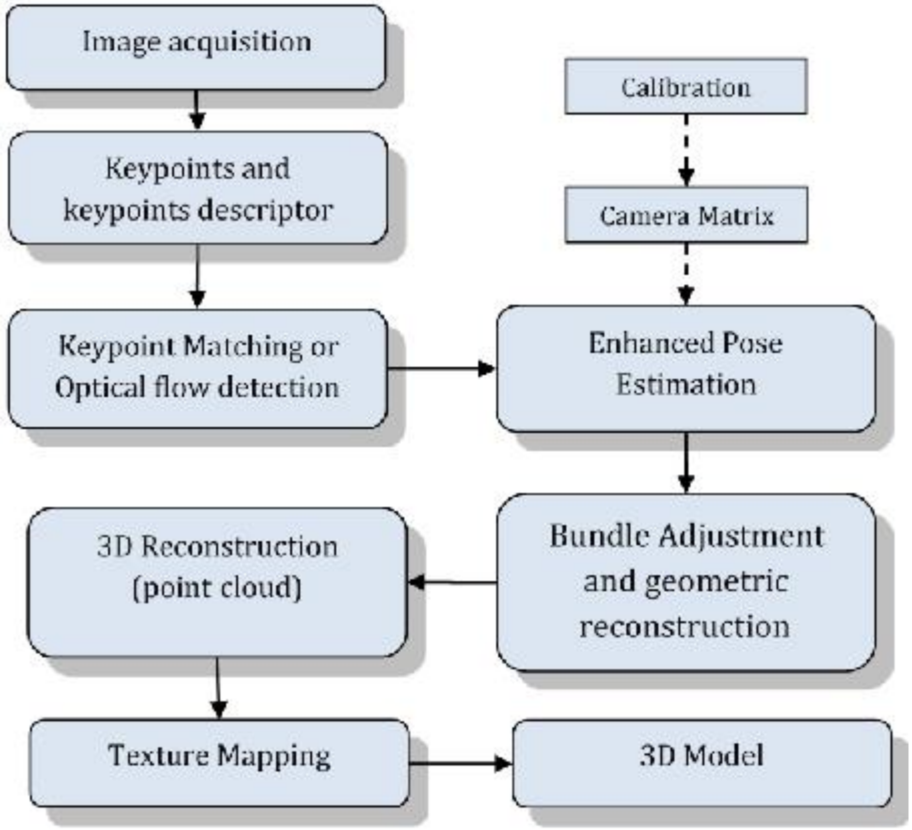

Figure 26-Testing phase of SfM

5.2.1 Test scenario 1:

Check results on inserting valid images at the time of the age acquisition:

A. Test Case 1: Extension of images as PNG or JPG

In this test case, the system will manually check the extension of all the images uploaded by the user whether the extension of the images is JPG or PNG. As because of the system restriction the only images file will be read by the system program that has been further processed for the 3D reconstruction and recreation of a point cloud.

B. Test Case 2: Check the number of images.

The SfM requirement of the images acquisition is 50 to 200 with the possible overlapping and similarities of 40 to 70 percent. So, in this test case, the system will check individually one by one the quantity and the overlapping similarities of the images and find the percentage overlap to check the percentage completely fulfill the condition of SfM.

5.2.2 Test scenario 2:

Check results on key extracted features at the time of feature extraction:

A. Test Case 1:

(check Extracted valid sharp edges and cornered features of images)

By the method of cv2, the system extracts the key feature of images cluster and the test case will check the valid test cases and sharp edges and corners of the images got from the user.

5.2.3 Test scenario 3:

Check the results of key extracted feature matches at the time of key point matches of images:

A. Test Case 1:

(check Match extracted key points of neighbored images)

In this test case, the system matches the key extracted features of all the images overlapping and finds the key feature matches of all the overlapping images. the test case will check the outcome of extracted features and matched features of overlapped images.

B. Test Case 2:

(Check results on key matched points to find matches)

Applying the different algorithms, the test case selects the best algorithm of SIFT to extract the match the key features of the overlapping images.

5.2.4 Test scenario 4:

Check results of sparse matrices at the time of sparse point calculations:

A. Test Case 1:

(Check results on finding triangular components and distance matrices)

In this test case, the system will check the triangular components and distance matrix formed by the system. In the distance matrix the angle of the object along with the triangular components such as length width, hypotenuse, is being stored.

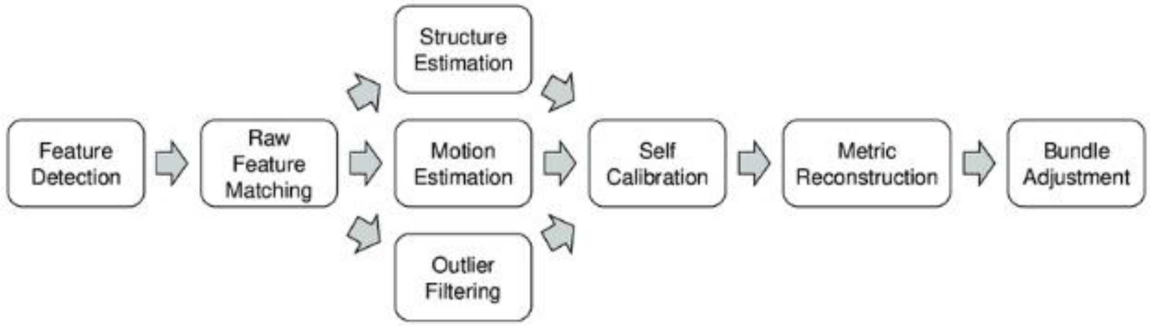

Figure 27-Test Scenario 4

5.2.5 Test scenario 5:

Check results on of merged dense point cloud matrices at the time of 3D reconstruction of sparse point cloud:

A. Test Case 1:
Check the results of overlapped pairs.

B. Test Case 2:
Check results on getting depth maps of the sparse point cloud.

C. Test Case 3:
Check results on merging the 3D point cloud.

In this test scenario, the system generates a depth map from the overlapping pairs and merge the depth maps matrix to form the 3D point cloud. All the test cases of the test scenario of the system it is being checked the depth map and merged 3d point cloud of all the images. On the other hand, the extension of the merged point cloud is made stored in the folder and displayed to the user.

Chapter 6
Future Work

This project can be extended further for future support in online shopping. We will try to add more features to our project by making it useful for online retailers in online shopping. Right now, we are developing a desktop application, but we will try to move it to a website. Cost plays an important role when building a software application. As a software developer, we'll work on reducing the cost of our desktop application by making it effective and easy to use for every individual whole has a smartphone.

6.1 Website

Websites can be utilized in different designs: an individual site, a corporate site for an organization, an administration site, an association site, and so forth. Sites can be crafted by an individual, a business, or other association, and are regularly committed to a specific theme or reason.

We will make our product live by launching its Soloscan.com website where users will be able to upload images from the camera and create a 3D model. We are also planning to help online retailers, so they will upload the images of their clothes, our website will create a 3D model of those clothes. These clothes will be available for everyone, so when the client asks to create a 3D model then he'll be able to able the clothes model to his/her body model, which will help them in online shopping. This mechanism will help not only in online shopping but will also reduce the effort of visiting a shopping mall for shopping because everything will be at hand.

6.2 Background removal

In our application, the background should remain constant which is somehow difficult for a layman to take pictures with a static background. So, in the future, we are planning to add a feature that will help in removing the background of images before the creation of a 3D model. Many online applications are working on removing the background, but we want to retrieve background while generating a 3D model. We need to study more about this subject of removal and retrieval of background.

6.3 Introduction of machine learning

Artificial intelligence forms the premise of all AI and is the fate of all great dynamics. We can improve the generation time i.e. in what capacity may I improve the speed of generation of the 3D model utilizing the arrangement of AI. Different firms take around 48 hours to make a 3D point cloud and deliver it like 3space.com. It takes 48 hours to deliver a 3D model to the customer, so the time of generation is of earlier significance unquestionably.

Accordingly, our framework will be consistently powerful for our clients. Computerized reasoning has taken the entire world. Today everything around us has some fragments of AI. We will utilize the Convolutional Neural Network to improve the generation time of our 3D model. So, we propose a from beginning to end Convolutional Neural Network (CNN) approach for 3D displaying i.e. 3D point cloud generation utilizing a single picture. Showed up contrastingly comparable to those standard frameworks applied to 3D multiplication. Stood out from those standard techniques applied to 3D recreation, CNNs can pick up capability with a raised degree of depiction subsequently with no manual mediations to stretch out a singular picture to the perfect yields.

6.3.1 Input

Our input will be a monocular image that will be used for further processing.

6.3.2 Output

A 3D point cloud

6.3.3 Constraints:

1. Distortion in an image may occur because of the background.
2. Image taken by any angle may involve multiple objects

6.3.4 Solution:

If images are taken from different angles, then the background should remain the same. We will use the convolutional neural system to make a 3D point cloud. Along these lines, the generation of 3D point cloud efficiently involves the following steps:

6.3.4.1 Preprocessing:

In the wake of choosing the key reasons for the picture, we will isolate slices dependent on planes perpendicular to the high tomahawks. Expect that the model space S contains all the reasons for an article whose model is to be made.
The 3D model has three estimations, so we give out:

$(x0,y0,z0) = (x0,y0, \min(Hz + i(dz/1-m))$
Where m is the quantity of cuts and Hz = z.

By and by the missing data will be filled and anomalies will be ousted. By then slice vector will be delivered from the cuts.

The fundamental strategy to calculate a slice circumference is including all detachments of two adjoining focuses. In the ensuing stage, given data will be used for learning reasons.

6.3.4.2 Learning Model

Here we'll use the convolutional neural network with one hidden layer which will bring about objective slice shape. At that point utilizing direct interjection, remaining slices can be found.

6.3.4.3 Activation Function

ReLU

6.3.4.4 Procedure

Using the convolutional layer, the model will be familiar with the nearby slices. By finding the essential slices, other dependent slices will be rectified.

6.3.4.5 Padding:

For the preparation of the model, vertical and symmetric padding will be done to get the exact information from the image ie no missing values.

6.4 ALGORITHM

We will use the convolutional neural network algorithm to detect the features present in an image and for the creation of a 3D point cloud.
The algorithm given below is for getting primary slices from the sample space.

Input: Set of measurements P, Set of 3D shapes in which $S^{(i)} \in D$ is a sample following the slice-structure.
Output: Set of learned parameters W
for $h \in SP$ do
  $\text{loss}_h = 0$
  for each sample $S^{(i)} \in D$ do
    w = length of vector $S_h^{(i)}$
    $r = (P_h^{(i)})/2\pi$
    init X as w-dimensions array with $X_k = r$
    $Y = f(X)$
    $\max = \max Y$
    $\min = \min Y$
    $$\text{loss}_h = \text{loss}_h + (1/w)\|Y - S_h^{(i)}\| + (2\pi\sqrt{(((\max)^2 + (\min)^2)/2) - P_h^{(i)}})$$
  $W^{(h)} = \arg\min(\text{loss}_h)$

For the generation of a 3D point cloud using CNN,

Constructing 3D point cloud.

Input: Set of generated primary slices D' in which $S'^{(i)} \in D'$ comes from Algorithm 1, Set of 3D shapes in which $S^{(i)} \in D$ is a sample following the slice-structure
Output: Set of learned parameters W
$\text{loss} = 0$
for each sample $S^{(i)} \in D$ do
  $n =$ the number of rows of $S^{(i)}$
  $m =$ the number of column of $S^{(i)}$
  init X as $n \times m$ matrix
  for k from 1 to m do
    for h from 1 to n do
      $\text{start} = \max\{x \in PI | x \leq h\}$
      $\text{end} = \min\{x \in PI | x \geq h\}$
        $X_{hk} = S\prime^{(i)}_{\text{start},k} + (S\prime^{(i)}_{\text{start},k} - S\prime^{(i)}_{\text{end},k}) * (h - \text{start})/(\text{end} - \text{start})$
  $Y = g(X)$
  $\text{loss} = \text{loss} + 1/mn\|Y - S^{(i)}\|$
$W = \arg\min(\text{loss})$

Chapter 7
Conclusion

The project "Structure from Motion 3D Scanner" provides a cost-effective solution for the creation of 3D point clouds. 3D modeling will become easier due to the implementation of the SfM technique. We are familiar with today's technological world where almost everything is online and technology-based, so we have developed a system that generates a 3D point cloud of a static object by taking pictures from a different angle using an ordinary mobile camera.

The modern world is working on 3D modeling, but the unique feature of our product is that it is cost-effective. Users prefer products that are cost-effective and do their job efficiently and optimally. We are familiar with the 3D modeling constructors which use stereo vision and LIDAR technology which is way too expensive to be affordable by a layman. Our product is user friendly in this perspective too, as it requires a single camera to do the same job done by those systems that use more than one camera and costly techniques.

This product can be used in many different fields including the medical industry, archaeology, cultural heritage, and geosciences by adding features according to the need for the required job.

Bibliography

[1] Jay, S., Rabatel, G., Hadoux, X., Moura, D., & Gorretta, N. (2014, October 29). In-field crop row phenotyping from 3D modeling performed using Structure from Motion. Retrieved July 1, 2020, from https://www.sciencedirect.com/science/article/pii/S0168169914002609

[2] Fathi, H., & Brilakis, I. (2011, July 06). Automated sparse 3D point cloud generation of infrastructure using its distinctive visual features. Retrieved July 1, 2020, from https://www.sciencedirect.com/science/article/abs/pii/S1474034611000334

[3] Ko, J., & Ho, Y. (n.d.). 3D Point Cloud Generation Using Structure from Motion with Multiple View Images.

[4] Fan, H., Su, H., & Guibas, L. (2017, January 01). A Point Set Generation Network for 3D Object Reconstruction From a Single Image. Retrieved July 01, 2020, from https://openaccess.thecvf.com/content_cvpr_2017/html/Fan_A_Point_Set_CVPR_2017_paper.html

[5] Soule, S., Maurice, K., Walcher, W., & Szabo, J. (2002, September 25). Advanced point cloud generation for photogrammetric modeling of complex 3D objects. Retrieved July 11, 2020, from https://ieeexplore.ieee.org/abstract/document/1039024

[6] Mathews, Adam & Jensen, Jennifer. (2012). Three-Dimensional Building Modeling using Structure from Motion: Improving Model Results with Telescopic Pole Aerial Photography

[7] Yang, G., Huang, X., Hao, Z., Liu, M., Belongie, S., & Hariharan, B. (1970, January 01). PointFlow: 3D Point Cloud Generation With Continuous Normalizing Flows. Retrieved July 01, 2020, from https://openaccess.thecvf.com/content_ICCV_2019/html/Yang_PointFlow_3D_Point_Cloud_Generation_With_Continuous_Normalizing_Flows_ICCV_2019_paper.html

[8] Toda, Y., & Hashimoto, K. (2018). 3D point cloud model reconstruction method by using multi–view 2D images and 3D point

clouds. The Proceedings of JSME Annual Conference on Robotics and Mechatronics (Robomec), 2018(0). doi:10.1299/jsmermd.2018.2a1-i17

[9] Hu, Q., Luo, J., Hu, G., Duan, W., & Zhou, H. (2018). 3D Point Cloud Generation Using Incremental Structure-from-Motion. Journal of Physics: Conference Series, 1087, 062031. doi:10.1088/1742-6596/1087/6/062031

LXIV