School

of

Electronics and Communication Engineering

Mini Project Report

on

Novel View Generation Using Sparse Views

By:

1. **Praveen C.** 01FE18BEC261
2. **Nikhil A.** 01FE18BEC088
3. **Jagadish B.** 01FE18BEC062
4. **Vijayalaxmi P.** 01FE18BCS253

Semester: V, 2020-2021

Under the Guidance of

Ramesh Ashok Tabib

K.L.E SOCIETY'S
K.L.E Technological University,
HUBBALLI-580031
2020-2021

SCHOOL OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled **"Novel View Generation Using Sparse Views"** is a bonafide work carried out by the student team of **"Praveen C(01FE18BEC261),Nikhil A(01FE18BEC088), Jagadish B(01FE18BEC062), Vijayalaxmi P(01FE18BCS253)"**. 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 Communication Engineering of KLE Technological University for the academic year 2020-2021.

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

External Viva:

Name of Examiners **Signature with date**

1. Shamshuddin K
2. Ujwala P 17/3

ACKNOWLEDGMENT

The feeling of fulfillment that lies within the effective completion of Novel View Generation using Sparse Views, would have been unfinished without citing the names of the individuals who helped us generously enough in the process of achieving the goals of the project, with their consistent directions, encouragement brought about in it's realization. We are thankful to our esteemed institute KLE Technological University, Hubballi, which has provided us this opportunity to fulfil the most cherished desire to reach our goal. We also express our deep sense of appreciation and gratitude to the School of Electronics and Communication, Nalini Iyer, the Head of Department, and the faculty at Computer Vision and Graphics Lab,Uma Mudenagudi, Ramesh Tabib, Chaitra Desai and, Ujwala Patil and Rohan Raju Dhanakshirur for guiding us and, encouraging us throughout the completion of this project. Finally we would like to thank all those who either specifically or indirectly helped us, in achieving the goals and objectives in regards with the project.

-The project team

ABSTRACT

The rapid advancements in the field of Augmented Reality and Virtual Reality and the vastly available resources make it conceivable to join hands in innovation. Smart gadgets nowadays such as mobiles, tablets, AR devices etc, as we even speak of it, are highly capable of rendering various things in augmented reality and virtual reality as well. This has led to the development and advancements in 3D reconstruction and Novel View Generations of any given object. With this motivation, we propose a design, to reconstruct or generate a new view, for a given object. In this project we mainly aim to generate a new view, for any given image/object, employing computer vision methods and algorithms. To generate a novel view, we take an input sparse view, calibrate the camera, and then render the subject to a 3-dimensional Point Cloud, and select a suitable view to project to an image. We then employ masking technique, and fill the holes, generated using Navier Stokes and Telea algorithms. We then present the results, of using the described method by generating a new view of the given input image of a man.

Contents

List of Figures

Chapter 1

Introduction to Novel View Synthesis

Novel view synthesis is the task of generating new images that render the objects from a different view point than the given. Meaning our result is the same image that would have resulted if we really put a camera and took a picture from that location or view point. To achieve this, we use camera perspective warping method. In this method, we calibrate a camera, with the checkerboard a.k.a., calibration board. To calibrate a camera, pin-hole model camera to be specific, means to extract the intrinsic matrix for the given camera, which contains information such as focal length, principal point offset, and axis skew. The focal length (fx, fy) describes the distance between the pin-hole and the image plane.Note that the focal length, needs to be same for all the images taken, however, it isn't terribly intuitive. The principal point offset(x0,y0) defines the location of the principal point relative to the origin of the image plane. Principal point is the point where principal axis of camera which is perpendicular to the image plane that passes through the pinhole, intersects. And axis skew causes shear distortion in the projected image [3]. With all the information such as intrinsic matrix of camera and the calibration data for the images, we now generate a disparity map, using which we then render a point cloud for the input image. The point cloud is transformed and desired view point is set, and the next step is project the corresponding transformed view to an image. This new image generated has some missing data or holes or voids as we call it.To fill these generated holes, we use inpainting methods or filling of holes, to make the image look like its real and natural. To achieve this, we use 2 state-of-the-art algorithms. NavierStokes method, and Telea method proposed by Alexandru Telea. The NS method uses the information from the neighbouring pixels where the hole is present, and subsequently fills the holes. While the Telea method uses a technique called as Fast Marching Method(FMM) [5].
With inpainting, we are left with the results, i.e., the inpainted image, that portrays how the subject would look from the given point of view.

1.1 Motivation

- Today's computer graphics(CGI) and artificial intelligence world generating a new virtual world from the available data is in a lot of demand, but the CPU's and monitors used for this kind of work are too high-end and cost inefficient.

- Our aim is to propose a method to generate a new perspective(novel view) of a scene or an object from the available data and can be processed for further application.

1.2 Objectives

- Generate the 3D structure of the given scene or an object for a given source pose.
- For a given input image and a given source pose of a scene or an object, a new view(novel view) must be generated given the target pose(new pose) of that same scene or an object.
- The sparse image or 3D sparse structure must be rendered to fill all the missing data present in it.

1.3 Literature survey

- **SynSin: End-to-end View Synthesis from a Single Image [6], by Olivia Wiles et. al.,** encorporates mainly two GANs in order to acheive novel view generation. The author mainly speaks of the Spatial Feature network f, which at its core uses a BigGAN architecture and consists of 8 ResNet blocks which maintain the resolution and the final block predicts a 64-dimensional feature for each pixel of the input image. The Depth network d which employs a UNet and has 8 upsampling and downsampling layers that give a final prediction of the same spatial resolution as the input., which is followed by a sigmoid layer and a renormalisation step so that the predicted depths fall within the per-dataset min and max values. The author also proposes a neural point cloud renderer in order to overcome the problems of small neighborhood, and the hard-z buffer causes by the Naive renderer to render the point cloud. In this, a differentiable renderer for meshes by similarly softening the hard rasterisation decisions and that renders point clouds by splatting points to a region and accumulating. The flow can be known in the block diagram shown in fig 1.1

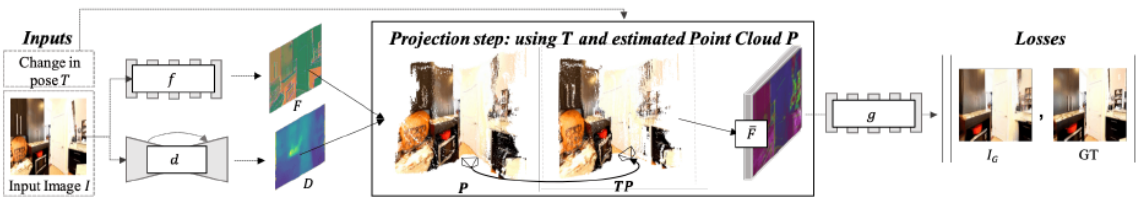

Figure 1.1: End-to-end system proposed in SynSin

- **Paul Debevec et. al., in Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping** [1], proposed classical vision-based methods of obtaining novel views. This project builds on the polygon graphics mainly. VDTM is a technique used for generating novel views of a scene whose geometry is approximately known, and by making maximal use of a sparse set of original views. The novel view is generated by querying the view map for each polygon, to determine a set of no more than three original images to be blended to render the polygon with projective texture-mapping. Holes are usually generated in the form of triangles or invisible triangles as the author calls it, are filled using an object-space hole-filling method. But the resultant novel view image was found to be of lower resolution than the input image. Results are shown below in fig 1.2.

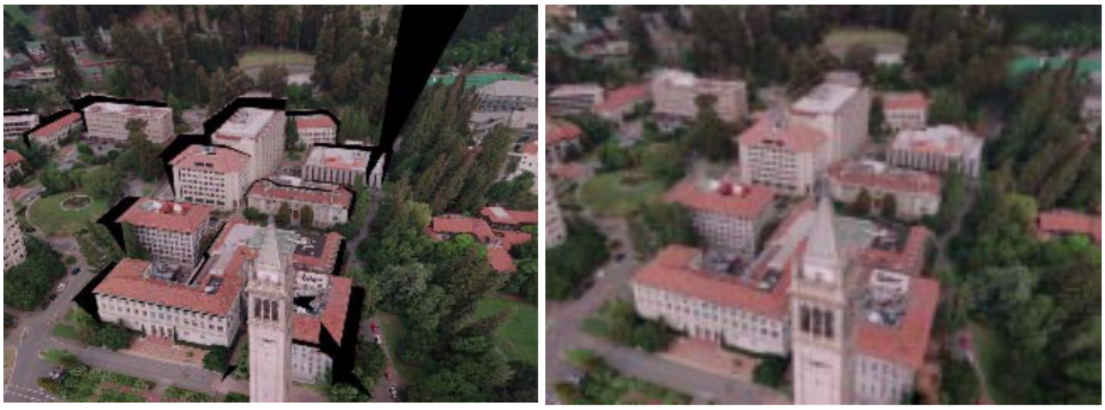

Figure 1.2: Results of Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping

- **Composition of novel views through an efficient image warping** [7], by Xin Zheng et. al., propose a method of novel view generation, using image warping techniques. In this method, reference images are taken from multiple viewpoints, and inverse-warping is performed to get the novel view. This consists of three main steps, preprocess the edge-pixel extraction, inverse warping from the primary image and hole filling from the remaining reference images. The concept of epipolar geometry is also incorporated for efficient inverse warping on the primary image.

1.4 Problem statement

We focus on a problem where one aims at generating samples corresponding to some objects/scene under various views, Design and synthesize the target image with an arbitrary target camera pose from a given source image and their camera pose.

1.5 Project Planning

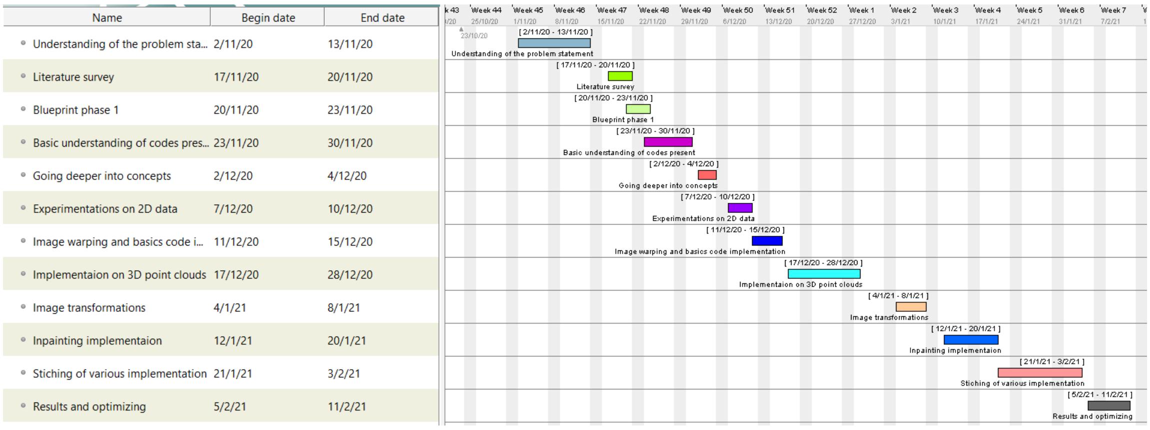

Figure 1.3: Gantt chart depicting the timeline of the project

1.6 Organization of the report

In Chapter 1, we discuss about a brief overview of the introduction to Novel View Synthesis, motivation, objectives of achieving a Novel View, literature survey, and problem statement. In the next chapter, Chapter 2, we describe the proposed system design, a block diagram of the system, alternate solutions available for the generation of novel views, and the implemented solution. In Chapter 3, we describe the implementation details, algorithm of the proposed model, flow of the designed system along with specifications and a detailed explanation of all the stages involved in the efficient generation of novel views. In Chapter 4, we talk about the optimizations for efficient novel view synthesis, and various techniques available for it, and the optimization techniques used in the algorithm. In Chapter 5, we discuss on the results of novel view synthesis performed. Finally we conclude the project in Chapter 6, with conclusions and the future scope of the project.

Chapter 2

System design: novel view synthesis

System design elaborates the functions,that the system should perform and the various means of achieving the generation of novel views. Finally, the best design among the various design alternatives is selected.

2.1 Functional block diagram for novel view synthesis

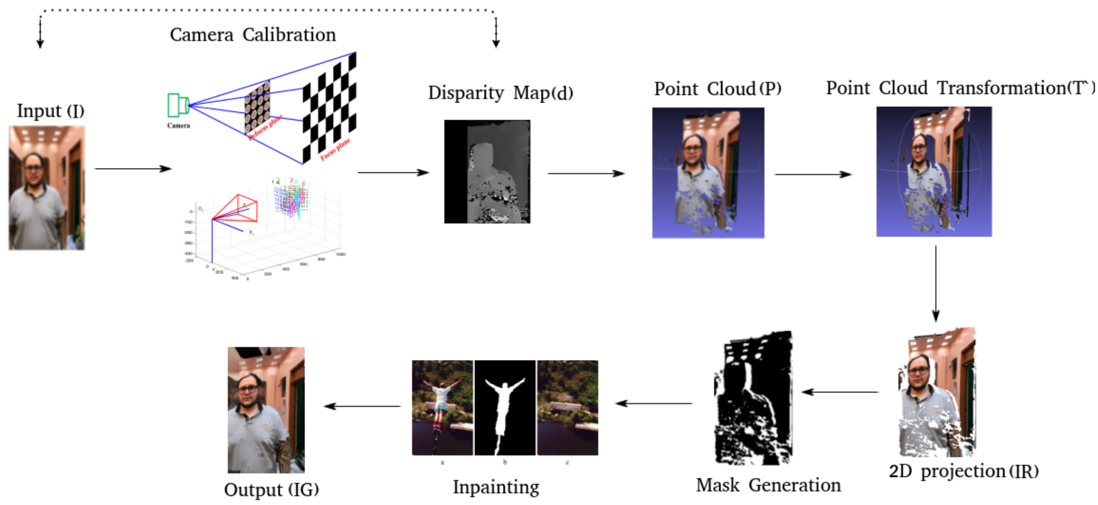

Figure 2.1: Block diagram for the design proposed to synthesize novel views

The initial step is the calibration of the camera, the system takes stereo images as the input images(I) of the given camera pose(T) and generate the disparity map(d), then project the images(2D data) into 3D point cloud(P),take the new camera pose(T') and apply transformation on the 3D point cloud, re-project back the 3D point cloud into a new 2D image, the rendered sparse image(IR) is passed through a classical inpainting algorithm to generate the final image(IG).

2.2 Alternative designs for synthesizing novel views

Here we discuss about the various design alternatives available to achieve the generation of novel views.

- The state of the art method: SynSin: End-to-end View Synthesis from a Single Image: which was published by Olivia Wiles et. al., This method specifically uses two GAN-based architectures, one for spatial feature prediction and the other for predicting the depth feature for the corresponding spatial features. It also uses ResNets and UNets, for maintaining the image resolution and to predict the depth respectively. However, this might be a little expensive on the hardware side of the implementation to train the generators and the discriminators. But the results, are fairly acceptable for novel views. As for the results, the SynSin has a PSNR of about 20.9 to 21.14 for Matterport 3D, RealEstate10K, and Replica datasets.

- The classical method: Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping: This method uses the traditional old polygon graphics-based algorithms. Meaning the given scene is divided into polygons, and an estimated view map is generated. The novel view is generated by querying each polygon to determine a set of no more than three original images to blend to render the polygon with projective texture-mapping. Hole-filling techniques used in this method are fairly acceptable as well. However, the result has a lesser resolution than the input image, meaning the resolution of the input and the output is not maintained uniformly throughout the process of generation of the novel views for the given scene.

- The classical warping method: Composition of novel views through an efficient image warping: this method uses a technique of image warping called inverse image warping, wherein, the reference image is raster-scanned for the corresponding sample/destination. This adversely affects its time complexity due to the unavailability of depth information in the destination pixels. This also uses the concepts of epipolar geometry. Hole filling is done based on the epipolar lines. The searching time can be reduced by filling pixels in the same region of the hole along one epipolar line simultaneously and by segments. Raster scanning could be avoided, by storing the constructed bi-directional list for the epipolar line at the destination and the intervals that have been processed on the corresponding epipolar line at the source/reference image.

2.3 Final design chosen for synthesizing novel views

In this section we define our proposed design for generation of novel views.

2.3.1 Proposed Design

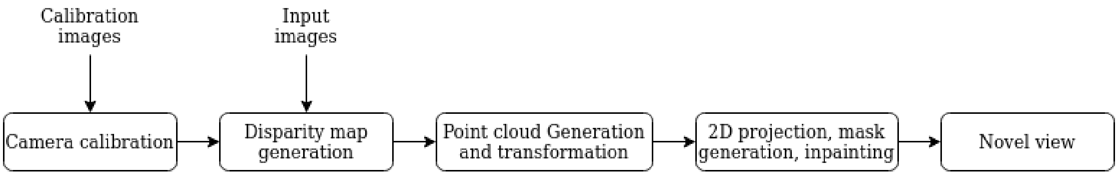

Figure 2.2: Proposed design for novel view generation

We propose a design, wherein we use the classical image warping based on camera parameters, we mainly extract the camera's intrinsic parameters after the process of calibrating the camera. With the extracted intrinsic matrix of the given camera, we then generate a disparity map for the given scene whose novel view is to be generated. Once the disparity map is generated, we then render it to a point cloud, transform it, and then project the transformed point cloud to a 2D image. This projected image will of course contain holes, which we then fill after generating a custom mask. To fill the holes, we use inpainting techniques. Navier-Stokes, and Telea methods to be specific.

Chapter 3

Implementation of novel view synthesis

3.1 Specifications and final system architecture

There are several concepts required to achieve the efficient generation of novel views. In this section we shall discuss the methodology, the concepts used, the algorithm developed, and the various other aspects of the proposed system design.

3.1.1 Camera parameters and Calibration of the camera

Camera parameters are the parameters used in a camera model to describe the mathematical relationship between the 3D coordinates of a point in the scene from which the light comes from and the 2D coordinates of its projection onto the image plane.

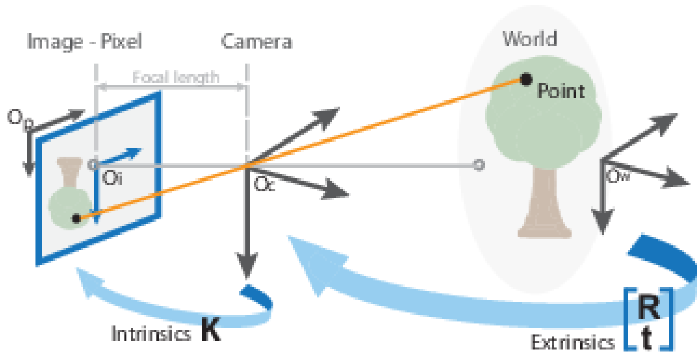

Figure 3.1: Camera Parameters

Types of Camera Parameters

- Intrinsic camera parameters It comprises of parameters such as focal length, distortion, aspect ratio of the pixel. The intrinsic camera parameters vary from each camera to camera.
- Extrinsic camera parameters It comprises of the parameters such as translation(T) and rotational(R) component of the camera co-ordinates. The relation between camera matrix and camera parameters is shown in Fig. 3.3

$$
\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X_w - T_x \\ Y_w - T_y \\ Z_w - T_z \end{bmatrix}
$$

Figure 3.2: Extrinsic parameters of camera

$$
\underbrace{w}_{\text{Scale factor}}\,[\underbrace{x\ y\ 1}_{\text{Image points}}] = [\underbrace{X\ Y\ Z\ 1}_{\text{World points}}]\,P
$$

$$
\underbrace{P}_{\text{Camera matrix}} = \underbrace{\begin{bmatrix} R \\ t \end{bmatrix}}_{\substack{\text{Extrinsics} \\ \text{Rotation and translation}}} \underbrace{K}_{\text{Intrinsic matrix}}
$$

Figure 3.3: Relation between camera matrix and camera parameters

Camera calibration is the recovery of the intrinsic parameters of a camera. Cameras, especially metric cameras, are calibrated in laboratories in a well-controlled environment. To calibrate, we need a calibration board, as shown in Fig. 3.5 and, the flow of process of calibrating the camera is shown in Fig. 3.6 Although the camera parameters can be obtained with high precision, the calibration is rather intensive and costly. In addition, the principal distance of the calibrated cameras needs to be fixed during application.

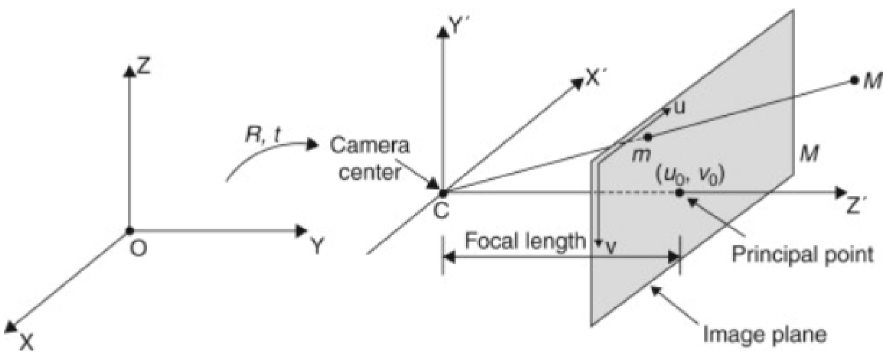

Figure 3.4: Camera Calibration Geometry

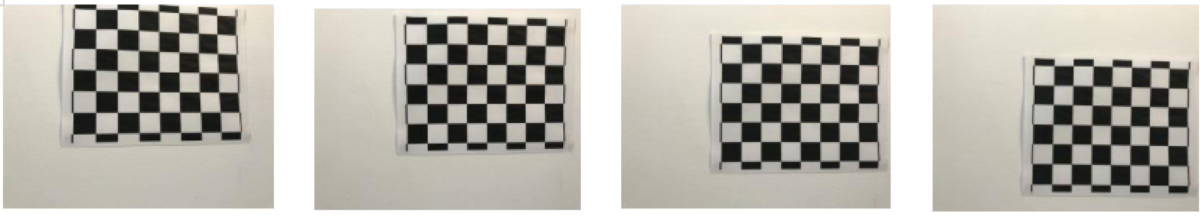

Figure 3.5: Calibration board images for camera calibration[4]

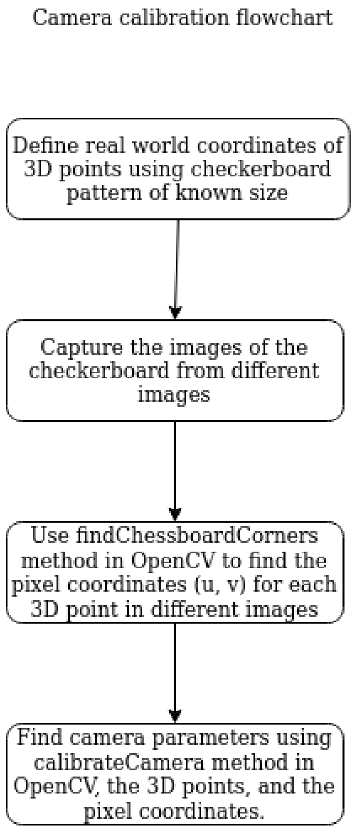

Figure 3.6: Camera calibration flowchart

3.1.2 Generation of disparity map

To generate a disparity map d(x,y) for each pixel in one of the input images, which is typically stored in the form of greyscale images, two or more rectified greyscale images as fed as input to the stereo correspondence algorithms. The disparity is the 2D vector between the positions of corresponding features seen by a human's left and right eyes, which is inversely proportional to the depth and is possible to define a mapping from an (x,y,z) triple to a 3D position. The disparity map generated is shown in Fig. 3.8

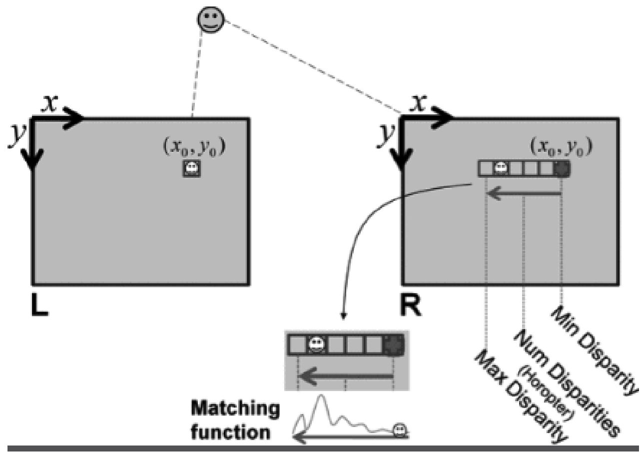

Figure 3.7: minimum and maximum disparity[2]

Figure 3.8: Generated disparity map

The algorithm has 3 steps:

- Pre-filter image to normalize brightness and enhance texture
- Correspondence search along horizontal epipolar lines using a SAD window
- Post-filtering to eliminate bad correspondence matches

3.1.3 Generation of point cloud

The disparity map is calculated and next step is to get an array of colors used in the image. Since the downsampling of the image is done we need to get the height and width. The transformation matrix is calculated. This matrix is responsible for to reproject the depth and colors into 3D space. The generated point cloud is shown in Fig. 3.10 below.

Each pixel of the image with the help of disparity map is projected onto x,y and z plane generating a point cloud given perspective transformation matrix.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f' & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z/f' \end{bmatrix} \Rightarrow \underbrace{(f'\frac{x}{z}, f'\frac{y}{z})}_{\text{divide by the third coordinate}}$$

Figure 3.9: Perspective transform matrix

Figure 3.10: Sample point cloud

3.1.4 Rigid transformation on the point cloud

A rigid transformation is one that moves the point cloud while preserving the distances between points in the cloud. Rigid transformations can be rotations, translations, and combinations of the two, but not reflections. A rigid transformation is parametrized by a transformation matrix in homogeneous coordinates. For a 3D transformations the matrix size is 4x4. Since the information is missing in the different view point of the point cloud, holes will be present in the point cloud which is termed as sparse point cloud as shown in Fig. 3.11

Figure 3.11: Transformed point cloud

3.1.5 2D projection

After the transformation on the 3-D point cloud generated, the new camera view or the new camera pose is taken as a 2-D plane. Each pixel present in the 3-D space is projected onto the 2-D plane. Since it is a sparse point cloud due to missing information in the different viewing angles the projected image will be sparse.

3.1.6 Mask generation

The holes generated in the sparse image needed to be filled with an algorithm. But before that, we need to detect where are the holes present in the 2D sparse image. For that, a mask is generated this mask has the information about where are the holes present in the image. Wherever the holes are present the pixel intensity of the mask will be 255 i.e, white and wherever image information is already present will have the pixel intensity has 0 i.e, black. This process is illustrated well in Fig. 3.12

Figure 3.12: Image with randomly generated holes and its corresponding mask

3.1.7 Inpainting with the mask on the 2D image

Image inpainting is a form of image conservation and image restoration. once we have the sparse 2-D image and mask generated by the algorithm. We need to fill those holes with an inpainting algorithm. In classical methods, there are two efficient methods inpaint telea and inpaint NS(Navier-stokes).

- **Inpainting using Telea algorithm:** Telea is based on the Fast Marching method: Let's say we have an image to be inpainted. Inpainting starts with the boundary of this region in the image and moves around inside the region gradually filling everything in the boundary region first. A small neighborhood around the area or pixel to be inpainted is considered and is replaced with the normalized weighted sum of all known pixels in the neighborhood. The pixels lying near to the point, near to the boundary normal, and those lying on the contours of the boundary are given more weightage. The fast Marching method is used to move around the next nearest pixel, after inpainting a pixel.

- **Inpainting using Navier-Stokes algorithm:** the Navier-Stokes algorithm builds on fluid dynamics and utilizes partial differential equations. The heuristic is used as the basic principle. The algorithm first travels alongside the edges from known regions of the image to the unknown regions of the image because edges are meant to be continuous. While gradient vectors at the boundary of the inpainting region are being matched, it continues isophotes. Some methods from fluid dynamics are used and as soon as they are obtained, color is filled to reduce the minimum variance in that area.

  Here we have used both the methods on the generated sparse image with the help of mask generated the holes present are filled and results are shown in Fig. 3.13 This generates the new view with a given new camera pose(T').

Figure 3.13: Inpainted image using Telea and Navier-Stokes algorithm respectively

3.2 Algorithm for the generation of novel views

Algorithm 1: Camera Calibration

```
Result: Camera's intrinsic matrix extracted
initialization;
while end of calibration images do
    detect checkerboard(claibration board);
    if checkerboard detected then
        extract camera parameters;
    else
        analyse next image for checkerboard;
    end
end
```

Algorithm 2: Disparity map generation

```
Result: Generate disparity map
if intrinsic matrix exists then
    generate disparity map;
else
    No intrinsic matrix found;
end
```

Algorithm 3: Point cloud generation

```
Result: Point cloud
if disparity map exits then
    generate point cloud;
else
    No disparity map found;
end
```

Algorithm 4: Point cloud transformation

```
Result: Point cloud transformation
if point cloud exits then
    transform point cloud;
else
    No point cloud found;
end
```

Algorithm 5: 2D projection

```
Result: 2D projected image
if transformed point cloud exits then
    project to 2D;
else
    No transformed point cloud found;
end
```

Algorithm 6: Inpainting

```
Result: novel view
if 2D image exists then
    do inpainting;
else
    No 2D image found;
end
```

3.3 Flow diagram for novel view synthesis

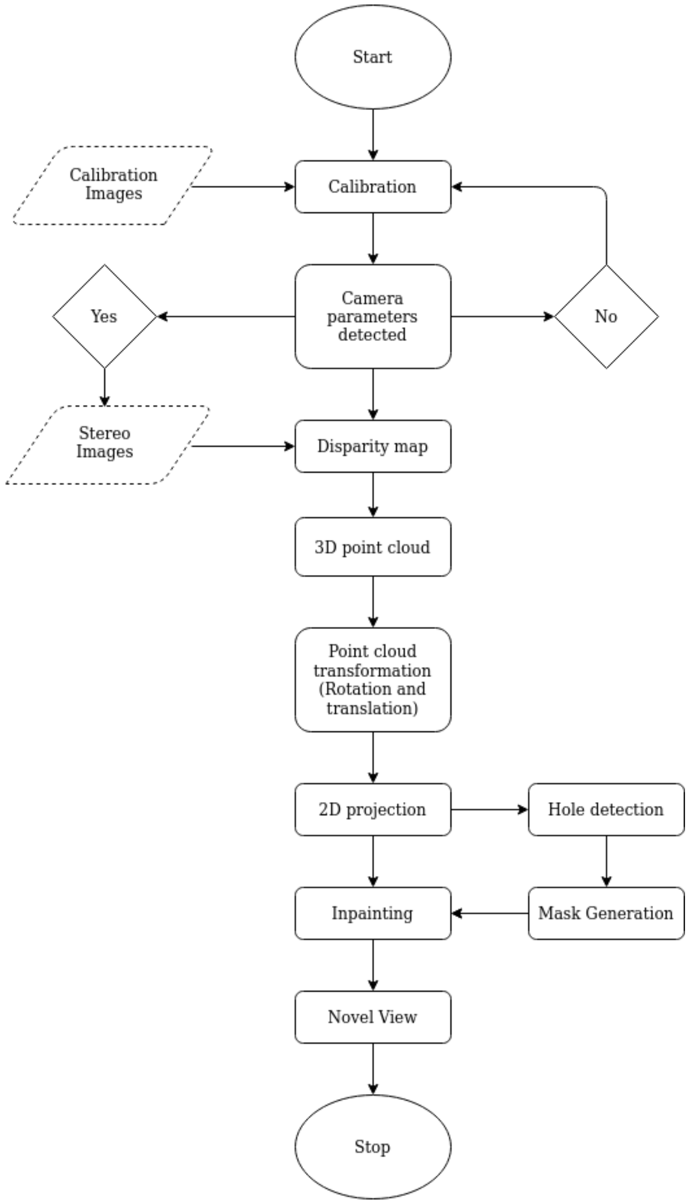

Figure 3.14: Working flowchart for the proposed system

Chapter 4

Optimizing the generation of novel views

4.1 Introduction to optimization

Our non-learning-based algorithm resembles the algorithm from the paper composition of novel views through an efficient image warping. However many optimizations can be done in the algorithm to get better results efficiently.

4.2 Types of Optimization

- Data optimization(Reducing the number of reference images)
- Data processing optimization(Construction of 3D point cloud or data)
- Using inpainting rather than primitive hole filling methods

4.3 Selection and justification of optimization method

4.3.1 Reduced number of reference images taken

In the reference paper the number of images taken as the reference for generation of novel view are 4-5, whereas in our proposed methodology, the number of reference images taken are 2(stereo images). Since our methodology uses the concept of point cloud data and rendering of 3D data the flexibility and usage of given reference data will be more compared to concept of image to image transformations.

4.3.2 Construction of 3D point cloud

Generation of 3D point cloud and data processing on 3D point cloud would give more flexibility for each data pixel present in the reference images since each pixel can be translated or rotated independent of other pixels or data points, unlike the image to image translation where each data point needs a reference epipolar line in the desired image.

4.3.3 Inpainting rather than primitive hole filling

Hole filling is a primitive algorithm of filling up the missing data in the 2D image or the 3D point cloud, but the algorithm presented in the paper uses 2 additional reference images to gather or collect the missing data points and does image to image translation through epipolar geometry which is computationally expensive. Hence our methodology uses the inpainting algorithm which needs a mask and the sparse images only as of the data, the inpainting algorithm averages the missing pixel value with the use of data points present surrounding it. Which is suitable in this architecture and has better results.

Chapter 5

Results of the proposed system to synthesize novel views

With the transformed point cloud projected to a 2D image, we perform inpainting on these resulting images to get a complete novel view of the given scene. In this chapter we shall discuss on the results obtained.

5.1 Result Analysis

5.1.1 Inpainting results

As discussed, we mainly use two algorithms to perform this. Navier-Stokes and Telea algorithms. Below are the results

Figure 5.1: 2D projected image, mask image, Inpaint NS, Inpaint Telea Results of inpainting

The results shown in fig 5.1 represents the 2D projected image, the respective mask, the results obtained off Navier-Stokes algortihm and the later image shows results obtained from inpainting off Telea algorithm. Both the results look about the same with some minor differences, indicating the generation of novel view, for the given scene.

With the proposed method, we were able to achieve:

- Peak Signal To Noise Ratio (PSNR) of about 28.8dB
- Structural Similarity Index Measure (SSIM) of about 0.58

Chapter 6

Conclusions and future scope

6.1 Conclusion

The algorithm developed was able to generate a new viewpoint for the given image. However, as mentioned still many refinements are always needed for the algorithm.The understanding of various concepts of computer vision and various algorithms like generating point clouds, warping, interpolation, inpainting, etc has been done.Novel view synthesis has been in many applications of today's AI and CGI world, This algorithm can be deployed into many other applications with some refinements

6.2 Future scope

6.2.1 Refinement of the algorithm

Instead of using only two stereo images as the input we can have multiple reference images, by this we will be having a dense point cloud, and flexibility and threshold to apply the transformation on the point cloud will increase

6.2.2 Replacing the inpainting part with a deep learning algorithm

Since we are using the Classical computer vision algorithm we would be reducing the computational cost of various GANs in a single algorithm. However with only one GAN in the algorithm for inpainting is affordable.By this we would have a efficient and better result.

Bibliography

[1] Paul Debevec, Yizhou Yu, and George Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. In George Drettakis and Nelson Max, editors, *Rendering Techniques '98*, pages 105–116, Vienna, 1998. Springer Vienna.

[2] Adrian Kaebler Gary Bradski. Learning opencv. https://www.eecs.yorku.ca/course_archive/2010-11/W/4421/doc/LearningOpenCV_1_2.pdf.

[3] Ksimek. Dissecting the camera matrix, part 3: The intrinsic matrix. http://ksimek.github.io/2013/08/13/intrinsic/, May 2013.

[4] Omar Padierna. Stereo 3d reconstruction with opencv. https://medium.com/@omar.ps16/stereo-3d-reconstruction-with-opencv-using-an-iphone-camera-part-ii-77754b58bfe0, May 2019.

[5] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, 9, 01 2004.

[6] Olivia Wiles, Georgia Gkioxari, Richard Szeliski, and Justin Johnson. Synsin: End-to-end view synthesis from a single image, 2019.

[7] Enhua Wu and Xin Zheng. Composition of novel views through an efficient image warping. *The Visual Computer*, 19:319–328, 08 2003.