

## 3D MODEL RECONSTRUCTION USING MULTIPLE PLANES FROM SINGLE 2D IMAGE

Spurthi V<sup>1</sup>, Deepu R<sup>2</sup>, Murali S<sup>3</sup>

<sup>1</sup>M. Tech, <sup>2</sup>Asso prof. & Head

Department of Computer Science & Engineering  
Maharaja Institute of Technology Mysore, India

**Abstract:** *The perception of 3D scene with stereovision is the capability of human vision but it is a challenge to computer systems. The challenge is to obtain 3D geometric shape information from planar images. This is termed as 3D reconstruction. It is usually done in a piece-wise fashion by identifying various planes of the scene and from those planes constructing a representation of the whole. The image is captured from a calibrated camera. The captured image is prospectively distorted. These distortions are removed by corner point estimation. The selected points give the true dimensions in the input image using view metrology. Then a 3D geometrical model is constructed in VRML according to true dimensions from metrology. In rendering, a texture map is created to each corresponding surface of a polygon in VRML model. VRML supports for walkthrough to a rendered model, through which different views are generated.*

### I. INTRODUCTION

As a popular research topic, image-based 3D scene modeling has attracted much attention in recent years. It has wide range of applications in civil engineering where civil engineers need to generate 3D views to conceive their product to customer. It is used in games and entertainment, to create virtual reality and also for robot navigation inside the building. In some cases, buildings that have disappeared can be modeled from as little as a single image, for example, an old photograph or a painting. To generate 3D views, one solution is to use specialized devices for acquiring 3D information of the image. Since these devices are cost expensive, it is not possible to use at all time. The other solution is to manual generation. The user requires prior knowledge about the object and also needs engineering skills. To generate thousands of models it is a time consuming process. Hence, there is a need for generating 3D views with less user interaction.

### II. PROBLEM STATEMENT

The main objective of 3D modeling is to computationally understand and model the 3D scene from the captured images, and provide human-like visual system for machines. The approach begins by capturing the 2D image from calibrated camera. Captured images are usually prospectively distorted. These perspective distortions are eliminated to get true dimensions of the image. 3D model is constructed in Virtual reality modeling language (VRML) using the true dimensions. VRML is a tool that supports visual

representation of the 3D model. Then the prospectively corrected images are mapped to corresponding planes in VRML model to create realistic effect. VRML supports walk-through simulation in a 3D space, so that the user can navigate the model according to the requirement.

### III. EXISTING METHODOLOGY

Most of the existing methods paid attention in generating 3D views from stereo images, sequence of monocular images and combining both these images to get 3D views. Each of their approach differs in cost, amount of time and user interaction for generating 3D models. Existing approach presented a 3D surface reconstruction from sequence of images. Structure from motion (SFM) method is used to perform automatic calibration and depth map is obtained by applying multi-view stereoscopic depth estimation for each calibrated image. For rendering two different texture based rendering techniques view dependent geometry and texture (VDGT) and multiple local methods (MLM) are used. Another approach used Potemkin model to support reconstruction of the 3D shapes of object instance. They stored different 3D oriented shape primitives of fixed 3D positions in a class model. They label the each images of a class for learning. A 2D view-specific recognition system returns the bounding box for the detected object from an image. Then a model based segmentation method is used to obtain object contour, using that object outline individual parts in the model is obtained. Shape context algorithm match and deform the boundaries of the stored part-labeled image for detected instance, thus it generated 3D model of the class using detected object. Also there are many approaches which are inconsistent and require more user intervention.

### IV. ARCHITECTURE OF THE PROPOSED METHODOLOGY

The overall steps to generate 3D model is shown in figure 1. Steps are as follows

#### A. View Metrology

Measurement of object's dimensions from their images acquired from an imaging device is called as view metrology. In a picture taken by the camera the depth information about the scene is lost. By Observing an image a human being can feel the shape of the object but not metric like height or width of object. So view metrology is used to measure the dimension of the objects in the image.

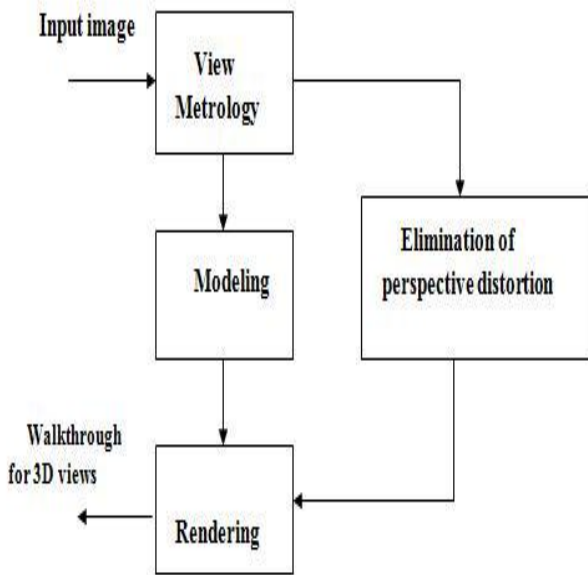


Figure 1: Steps to generate 3D model

**B. Perspective Elimination Process**

Perspective projection is the representation of an object on a plane surface, called the picture plane, as it would appear to the eye, when viewed from a fixed position. Thus the image obtained from any camera is nothing but the form of perspective view of the actual scene. Perspective distortions occur during the image acquisition stage. In perspective imaging, the shape is distorted because parallel lines in the image tend to converge to a finite point. Figure 2 below shows the example of perspective view of the object. The perspective distortions are eliminated by vanishing point and corner point selection to get real dimensions of the image which is explained in the section 6 and section 7.

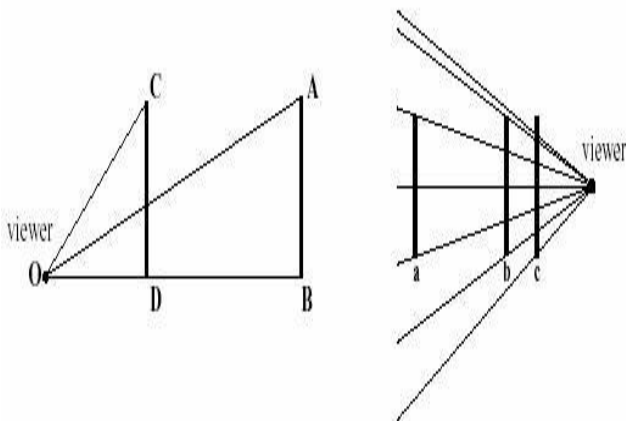


Figure 2: Perspective view of the object

**C. Modelling**

Virtual reality modelling language (VRML) is used for visual representation of model. VRML defines a file format that integrates 3D graphics and multimedia. Conceptually, each VRML file is a 3D time-based space that contains graphic.

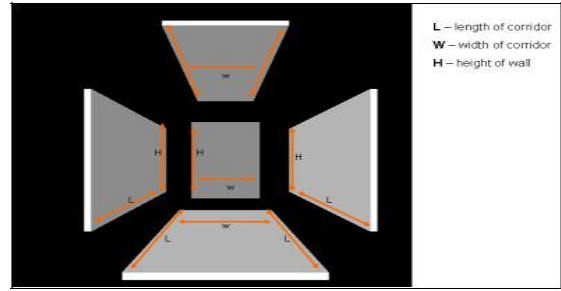


Figure 3: Model

Objects that can be dynamically modified through a variety of mechanism. The model of sample image is built using dimensions in VRML as shown in figure 3. Model can be simulated according to the user requirements.

**D. Rendering**

In rendering, the perspective corrected textures are used for mapping. The texture is mapped to corresponding polygon surfaces as shown in figure 4. Rendering gives the realistic effect to the model.

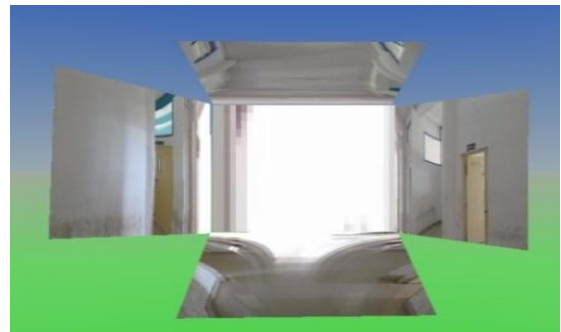


Figure 4: Model with texture

**V. SYSTEM DESIGN**

System design aims to identify the modules that should be in the system, the specifications of these modules will interact with each other to produce the desired results. The purpose of the design is to plan the solution of a problem specified by the requirements document.

**A. Dataset Collection**

The first stage in image analysis is data collection. Different set of 2D images at fixed distance and for fixed focal length are taken. The images are scaled to 512X512 resolutions. Different test images are shown in below figure.



Figure 5: Test image 1

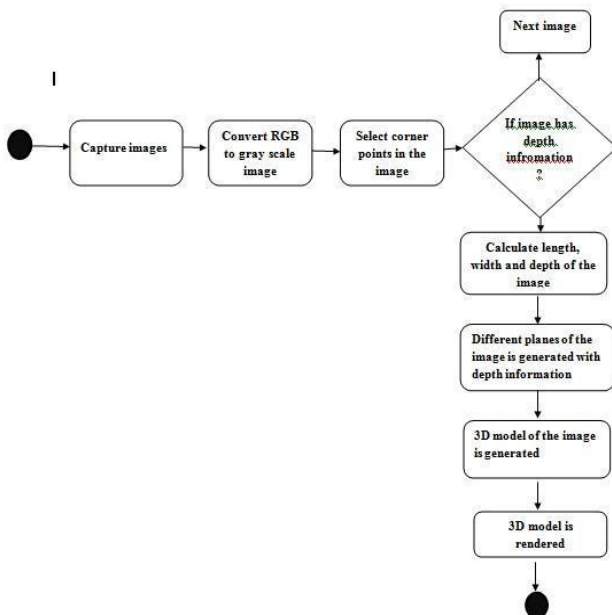


Figure 6: Test image 2



Figure 7: Test image 3

B. Flow diagram of the overall procedure



VI. CORNER ESTIMATION

Harris corner detector is a mathematical operator that finds a variation in the gradient of a image. It is rotation, scale and illumination variation independent. It sweeps a window  $w(x,y)$  with displacement  $u$  in the  $x$  direction and  $v$  in the  $y$  direction and calculates the variation of intensity. A window with a score  $R$  greater than the threshold is selected and points of local maxima of  $R$  are considered as a “corner” and are shown in Figure 9(a). Further it is reduced to four using geometric based calculations as desired by the plane homography method.

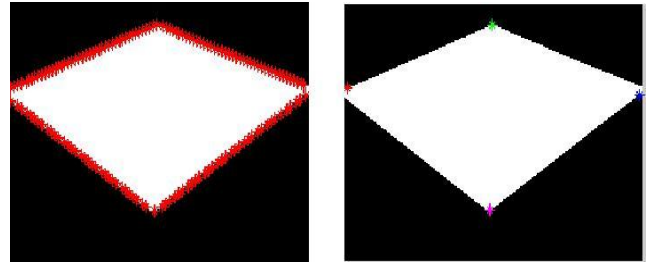


Figure 9: (a) All corners are detected (b) Desired corners obtained

VII. PERSPECTIVE ESTIMATION

Images captured in pinhole cameras are perspective in nature. In a perspective image, objects of similar size appear bigger when it is closer to view point and smaller when it is farther. Hence the number of pixels to represent farther objects in the image requires lesser pixels as compared to the number of pixels required to represent the closer objects in the image. Example: when objects are inclined to the viewer, it goes farther away from the viewpoint. Hence, the lesser the pixels means less information for image processing. Perspective images lead to many ambiguous results when we tend to measure the size of the objects in the picture. The rectangular plane ABCD in the Figure 10(a) is seen as  $A_1B_1C_1D_1$  in a prospectively distorted case. The perspective view of the rectangle is depicted as in the Figure 10(b). This is how the images of the objects are formed in the image acquisition process.

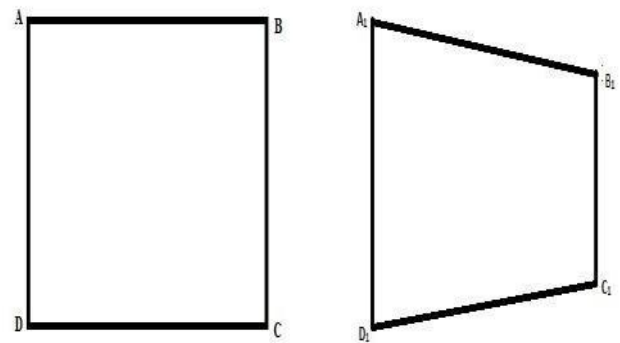


Figure 10: (a) Rectangular Plane (b) Perspective distortion

VIII. IMAGE REPRESENTATION

Firstly small homogeneous regions in the image, called “Superpixels,” are identified and used them as a basic unit of representation.(Fig.11) Such regions can be reliably found using over-segmentation, and represent a coherent region in the image with all the pixels having similar properties. In most images, a superpixel is a small part of a structure, such as part of a wall, and therefore represents a plane. In this experiment, algorithm is used to obtain the superpixels. Typically, over-segment an image into many superpixels, representing regions which have similar colour and texture. The goal is to infer the location and orientation of each of these superpixels.

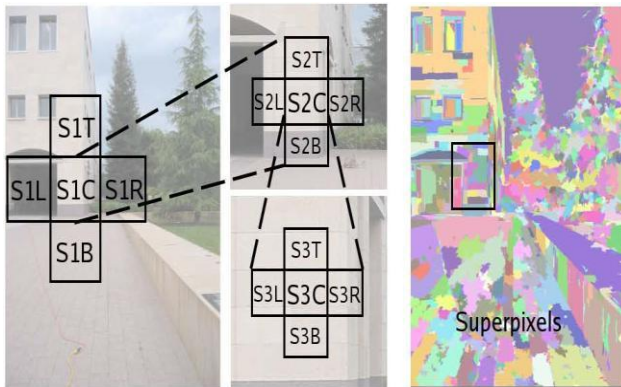


Figure 11: The feature vector for a super pixel, which includes immediate and distant neighbours in multiple scales

### IX. PROBABILISTIC MODEL

It is difficult to infer 3D information of a region from local cues alone and one needs to infer the 3D information of a region in relation to the 3D information of other region.

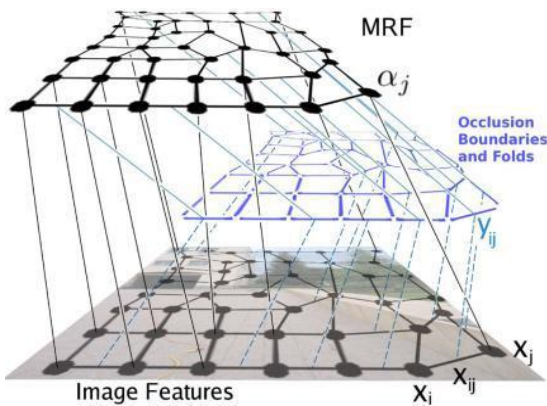


Figure 12: An illustration of the Markov Random Field (MRF) for inferring 3-d structure.

### X. TRANSFORMATIONS

Suppose the two images I1 and I2 to be registered having both translation and rotation with angle of rotation being 'theta' between them. When I2 is rotated by theta, there will be only translation left between the images. So by rotating I2 by one degree each time and computing the correlation peak for that angle, a stage where there is only translation left between the images is reached. That angle becomes the angle of rotation.

- Translation Transform  $x' = x + b$
- Affine Transform  $x' = ax + b$
- Bilinear  $x' = p1xy + p2x + p3y + p4$   
 $y' = p5xy + p6x + p7y + p8$
- 4. Projective Transform  $x' = (ax+b)/cx+1$

### XI. IMPLEMENTATION

The implementation phase of any project development is the most important phase as it yields the final solution, solves the problem at hand. The implementation phase involves the actual materialization of the ideas, which are expressed in the analysis document and developed in the design phase.

#### A. Image acquisition

During this phase, calibrated camera is used to capture the images as discussed above.

#### B. View metrology

The captured image from calibrated camera is shown in figure. The image is measured to get different dimensions like length,width ,breadth etc.. The image given to the system is converted into grayscale.



Figure 13: Input image

#### C. Corner point selection

In order to find the true dimension of the image, corner points and also one vanishing point are selected as shown in figure 14. The corner points and vanishing point helps to measure the dimensions in the image. Square symbol shows the corner points selected. Triangle shows reference points generated with respect to corner points. The lines generated in the figure 15 shows the distance between the pixels in the various planes. These points separated the image into different planes like front, top, bottom, side plane etc. Depth information of the each plane is generated as shown in below figure 16. These planes are used to generate 3D model in VRML. 3D model is generated according to the true dimensions from metrology.



Figure 14: Corner points selection



Figure 15: Measurement of dimension



(a): Front plane (b): Right plane



(c): Left plane (d): Top plane



(e): Bottom plane

Figure 16: 3D planes of the input image

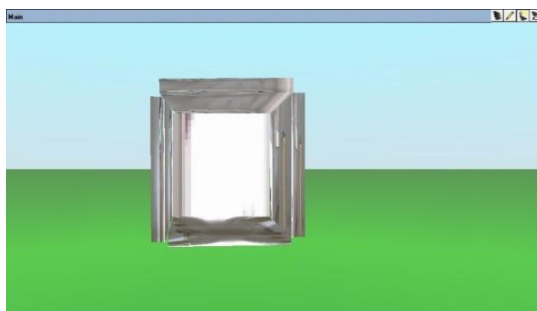


Figure 17: 3D rendered model in VRML

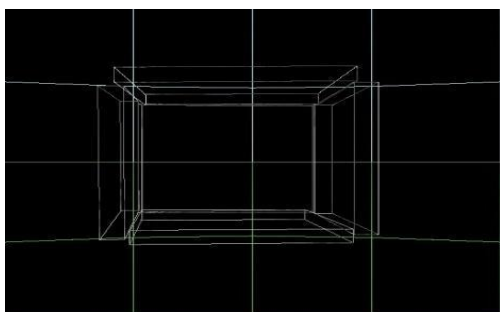


Figure 18: Wireframe model

In rendering a texture mapping is performed to each corresponding surface of a polygon in VRML model as shown in figure 17. Also wireframe model of the generated 3D model is shown in figure 18. VRML supports for walkthrough to a rendered model, through which different views are generated.

## XII. RESULTS

This section discusses the results of the proposed method. The method is implemented in matlab and output is viewed on VRML viewer. The method is tested for several 2D images. As compared to the existing approach as shown in the figure 19. This approach gives more information on the depth from the input image as compared visually as shown in figure 20. So that 3D model can be built exactly as a real world using these dimensions from the proposed method.

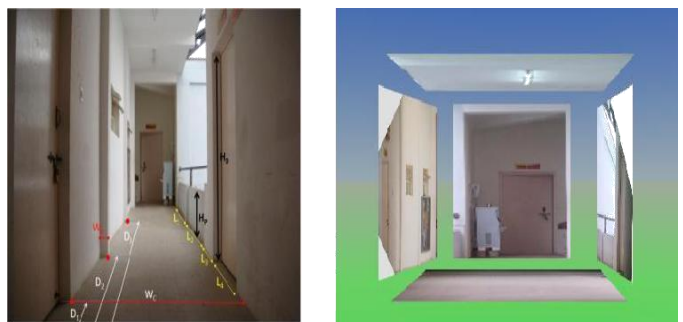


Figure 19 (a): Input image Figure (b): Output 3D planes



Figure 20 (a): Input image Figure (b): Output 3D plane

## XIII. CONCLUSION AND FUTURE WORK

In this approach a new method is presented for generating 3D model using multiple planes of the 2D image. The generated 3D model preserves the actual dimension of the object in the image. View metrology constructs orthographic views from the prospectively distorted image. Prospectively distorted image is corrected by vanishing point and corner point selection and the image is stretched according to its true dimension obtained from view metrology. It is an easy way to generate model with less user interaction. Finally the 3D model is rendered to give the realistic effect. Any image containing depth information can be applied to this method to calculate the true dimensions. This method can be used in robot path planning inside the building, artificial intelligence, animation, gaming, interior designing and architecture.

#### XIV. FUTURE WORK

In this approach, images are used for 3D modeling. Since this approach determines the depth of the images which has at least one vanishing point and corner points. The Future work includes enhancement of approach to accept any image and also to work on video processing. For example goal line technology is used football match to detect whether the ball is crossed the goal line.

#### REFERENCES

- [1] Akash Kushal And Jean Ponce “Modeling 3d Objects From Stereo Views And Recognizing Them In Photographs “ University of Illinois at urbana-champaign,2005
- [2] Ashutosh Saxena, Min Sun And Andrew Y. Ng “Learning 3-D Scene Structure From A Single Still Image “IEEE transactions on pattern analysis and machine intelligence (PAMI),2008.
- [3] S Murali N Avinash “Estimation Of Depth Information From A Single View In An Image “3DTV-conference: The true vision-capture, Transmission and Display of 3D video(3DTV-CON),2010
- [4] Heewon Lee, Alper Yilmaz “3d Reconstruction Using Photo Consistency from Uncalibrated Multiple Views” ohio state university, 2010.
- [5] K Susheel kumar, Vijay Bhaskar Semwal, Shitala Prasad And R.C Tripathi “Generating 3d Model Using 2d Images Of An Object “International Journal of Engineering Science and Technology (IJEST),2011.
- [6] Geetha Kiran A1 And Murali S “Automatic Rectification Of Perspective Distortion From A Single Image Using Plane Homography” International Journal on Computational Sciences & Applications (IJCSA) Vol.3, No.5, October 2013.