Real-Time 3D Environment Reconstruction Using Vision Sensor for Robotics Application

Saumitra Kulkarni , Sumukh Bapat , Rohan Patwardhan , Sahil Sonawane , Manisha Marathe

Department of Computer Science
PVG's COET Pune

***Abstract*- To construct a 3D environment in real-time using a single camera. The camera will be mounted on a ROV. ROV is Remotely Operated Vehicle. The ROV will moved around in the entire room by the user according to his will and the camera will capture the video of the room which will be used for the construction of panoramic view of 3D environment. The transfer of the images will be in real-time, i.e. the captured image will be immediately transmitted and processed, hence the model will be ready when the camera covers the entire room. The images which are captured are analysed for feature extraction, depth is calculated for every pixel using algorithms such as Triangulation which is used for plotting the pixels. Many algorithms are available for mapping the real-world co-ordinates to the image co-ordinates. Image calibration and estimation of the camera pose of an object in the case of computer vision area to do some 3D rendering or in the case of robotics is the most elemental problem nowadays. 3D reconstruction considering of image sequences is an unresolved and active research topic in computer vision.**

Keywords- Camera Calibration, Pose Estimation, Stereo Vision

I. INTRODUCTION

Image processing plays a key role in computer vision applications. Images are processed and key features are extracted from them using image processing for building various applications. The goal of 3D Reconstruction is to develop a 3D model of a scene from a set of 2D images. The knowledge of a 3D model gives a strong clue about the interiors of the room. The process of 3D reconstruction is also called as inverse problem, which can be defined as mapping the 3D world coordinates onto a 2D image plane. The combination of real-time and 3D reconstruction opens up a wide variety of applications. Augmented Reality (AR) is the most important application of 3D reconstruction. Another example is to calculate how exposed each point in a scene is to ambient lighting. To get highly accurate 3D reconstruction, a key requirement is that the construction should be dense. The latter part is achieved in real-time systems using self-localizing systems like Simultaneous Localizing and Mapping (SLAM) and Parallel Tracking and Mapping (PTAM). Other application of 3D reconstruction is the modelling community, where a vision driven reconstruction approach can replace expensive and often complicated laser-based reconstruction systems. Real-time in this context means that, the time between the capture of the image and its corresponding reconstruction should be in the range of seconds, so that the delay does not distract the user. An important requirement of our system is a highly computational hardware in the form of powerful graphics cards, offering high computational power to the end user. After the construction of the entire 3D model, the user can have the complete information about the interiors of the room interacting with the 3D model.

II. RELATED WORK

MediaMill3D

The system is made to serve as a test bed and a software prototype in the Crime Scene Investigation using hand-held cameras project, in summary, its functions are to reconstruct a 3D model of a crime scene from video sequences, add text and voice annotations, and fuse information from deferent video sequences[1]. It will be embedded into theMediaMill Video Search system. The target of our system is the geometric reconstruction and not the photo-metric or image-based reconstruction which directly generates new views of a scene without reconstructing the 3D structure. With the stated purposes and application context, we set the limits as:

Static scenes: There is no moving object or the movement of the objects is relatively small. Uncalibrated cameras: The input data is captured by an uncalibrated camera,i.e. the cameras intrinsic parameters such as focal length is unknown. Varying intrinsic camera parameters: The camera intrinsic parameters can vary freely. Together with the previous, this assumption adds edibility to the system[2].

The following are the steps of 3D reconstruction from video sequences:

1.Intrinsic calibration Calibration of internal parameters of the camera.

2.Image acquisition Capturing images using camera.

3.Pose estimation Calibration of external parameters of the camera.

4.3D reconstruction Map world co-ordinates to image co-ordinates[4]. A list of available resources for all the steps is mentioned below:

1. Intrinsic Calibration:

Intrinsic camera calibration model the projection of world points onto image plane.

$x = KX$

$x = (x, y, z)$ $X = (X, Y, Z)$

camera calibration matrix contains focal length (fx, fy) and principal point (px,py) of the camera.

Methods for internal calibration:

a. Multiview Geometry in Computer Vision - Hartley and Zisserman, 2003 - to find the angle between rays of image points.

b. A versatile Calibration technique for high accuracy 3D machine vision metrology using TV cameras and lenses. - Tsai, 1987 stepwise calibration method.

c. A flexible new technique of Camera Calibration - Zhang, 2000 homographybetween a planar calibration target and the image plane.

d. Self-Calibration and metric reconstruction in spite of varying and unknown intrinsic parameters. - Pollefeys et al., 1999 self calibration method.

e. Camera Calibration with distortion models and accuracy evaluation. - Weng etal., 1992 calibration methods with different distortion methods.

f. Automatic Calibration and removal of Distortion from scenes of structured environments. - Devernay and Faugeras, 2001 property of pinhole camera model.

2. Pose estimation:

Pose estimation is the estimation of external parameters of the camera[4].

$C = [R | t]$

$Xcam = C.Xworld$

$P = K.C = K [R | t]$

$x = PX$

Methods of pose estimation:

a. Multiview Geometry in Computer Vision - Hartley and Zisserman, 2003

b. Linear n-point camera pose determination. Quan and Lan, 1999

c. Random Sample Consensus - Fischler and Bolles, 1981 perspective 3 point

problem or perspective n-point problem.

d. An analytic solution for the perspective 4-point problem. - Horaud et al., 1989

e. Accurate non-iterative solution to pnp problem. - Moreno Noguer et al., 2007

3. Computation of 3D world point:

PVG, Department of Computer Engineering 2018-2019

4a.Navigation using a_ne structure for motion. - P. A. Beardslay and Murray,1994

b. Multiview Geometry in Computer Vision - Hartley and Zisserman, 2003

4. SLAM Simultaneous Localization and Mapping:

Simultaneously estimating the location and building a map of the environment isknown as SLAM[1].

Known methods are:

a. Simultaneous Localization and Map Building. - Csorba, 1997 , Durrany Whyteand Bailey, 2006 joint estimation of robot pose and observed landmark positions.

b. Smith et al., 1990 Kalman Filter

c. Optimization of the simultaneous localization and mapping algorithm for real-time implementation. - Guivant and Nebot, 2001

5. Reconstruction:

5.1. Range Image Generation:

Depth value is calculated for every pixel which corresponds to the position of aworld point in 3D space with respect to the camera frame[5].

a. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. - Scharstein and Szeliski, 2002

$Z = b*f/d$

Z depth value , d disparity , b baseline , f focal length

b. A compact algorithm for recti_cation of stereo pairs. - Fusiello et al 2000 rectication procedure

c. Graph-Cut - Boykov et al 2001, Kolmgorov and Zabih, 20001

d. Belief Propogation - Bonno and Ikeuchi, 2009

e. Bleyer et al, 2011

f. A space-sweep approach to multi-image matching. - Collins, 1996 – PlanesweepMethod

g. Real-time dense geometry for hand held cameras. - Stuehmer et al, 2010 -

Optical Flow, multiple views

5.2. Fusion (Integrate Information in single 3D model)

a. Curless and Levoy, 1996 Incremental updates ability to _ll gaps , robustness.

6. Parallel Tracking and Mapping

Main idea in tracking and mapping is to split tracking and mapping in two threads[3].

a. PTAM uses right hand co-ordinate system

b. Camera model is FOV-Model - Devernay and Faugeras, 2001

c. Five-Point Algorithm estimate the epipolar geometry.

7. Real-time Dense Reconstruction

a.Real-time dense geometry for hand held cameras. - Stuehmer et al, 2010 - Calculate dense depthmaps in realtime

b. Live Dense reconstruction using a single moving camera. - Newcombe andDavison, 2010 - Live dense reconstruction system.

c. Dense Tracking and Mapping in real-time. - Newcombe et al , 2011(a) DTAM.

d. Realtime dense surface mapping and tracking. - Newcombe et al, 2011(b) -KinectFusion.

8. Pros and cons of using video sequences are described below:

8.1 Advantages of using video sequences:

The most important advantage of using input of video sequences is the higher quality one can obtain. Both geometric accuracy and visual quality can be improved by exploiting the redundancy of data. Intuitively, more back-projecting rays of a points projections limits the possible 3D coordinates of the point. Other advantages are the automaticity and exibility. Instead of manually selecting some images from a video, it is better to have a system that can do everything automatically[5].

8.2 Problems in using video sequences:

1. Frame Selection.
2. Sequence Segmentation.
3. Structure fusion.
4. Bundle adjustment.

Methods exist for every step. The quality and robustness of a larger process, especially of the structure and motion recovery step, is the main concern. Though the process looks like a sequential one, practical solutions often require loops andfeedbacks at different levels. This makes the quality analysis and management ofthe process difficult.

III. PROPOSED METHOD

3D modelling is one way the construction industry can benefit from 3D technology greatly. It involves the process of creating a mathematical representation of any object or surface in three-dimensional spaces using specialized software. It basically brings 2D, flat ideas to life. We use a stereo camera to capture the video of the environment. This is given as input to the source code.

To model a simple scene in 3D out of 2D images using a stereo camera there are three possible approaches :

I. If intrinsic parameters of the camera are known then using two images that are loaded from a XML for instance loadXMLFromFile() =>stereoRectify() => reprojectImageTo3D().

II. If the camera parameters are not known then calibration can be performed camera=>stereoCalibrate()=>stereoRectify() =>reprojectImageTo3D().

III. If stereo camera is not available then there is need to find pair key points on both images with SURF, SIFT for instance, then compute descriptors of these key points, then matching key points from image right and image left according to their descriptors, and then find the fundamental mat from them. The processing is much harder in this approach.

We use the second step in which camera parameters are not known. Camera calibration is done using chess board images which are given as input feed to the stereo camera. stereoCalibrate() function is used to find the intrinsic and extrinsic parameters of the camera from several views of a calibration pattern. We give objectPoints, imagePoint1 and imagePoint2 as input arguments to stereoCalibrate() function. The outputs of this function are cameraMatrix1, distCoeffs1, cameraMatrix2, distCoeffs2, Output rotation matrix between the 1st and the 2nd camera coordinate systems (R), Output translation vector between the coordinate systems of the cameras (T), Output essential matrix (E), Output fundamental matrix (F). In this step we give output from previous step as input to stereoRectify() function from that we get 3x3 rectification transform (rotation matrix) for the first camera (R1), Output 3x3 rectification transform (rotation matrix) for the second camera (R2), Output 3x4 projection matrix in the new (rectified) coordinate systems for the first camera (P1), Output 3x4 projection matrix in the new (rectified) coordinate systems for the second camera (P2), Output 4x4 disparity-to-depth mapping matrix (Q).

we calculate disparity map further disparity map is used to plot 3D point cloud. We use stereoSGBM () function and WLS filtered to create disparity map . Open3D is a open source library that provides convenient visualization functions which use geometric objects like PointCloud, TriangleMesh, or Image and renders them together.

IV. REFERENCES

[1] KienDang Trung. A review of 3D Reconstruction from videoSequences. Amsterdam, Netherlands,Intelligent Sensory Information Systems

[2] Theo Moons, Luc Van Gool, and Maarten Vergauwen, 3D reconstruction from multiple images. Foundation and Trends in Computer graphics and vision, 2009.

[3] Diego Thomas and Akihiro Sugimoto, A two-stage strategy for real-time dense 3D reconstruction of large-scale scenes. Tokyo, Japan: National Institute of Informatics.

[4] Soulaiman El Hazzat, AbderrahimSaaidi and Khalid Satori, Euclidean 3D Reconstruction of Unknown Objects from Multiple Images. Fes, morocco: Journalof Emergingtechnologies in web intelligence, Feb 2014.

[5] Gottfried Graber, Realtime 3D Reconstruction. Graz University of technology, January 2012.