

# Design and Calibration of an Experimental Setup for 3-D Reconstruction of A Scene from Stereo Images

Rachna Verma

Dept. of CSE, Faculty of Engineering  
J.N.V. University, Jodhpur  
Rajasthan, India

A. K. Verma

Dept. of P&I Engineering, Faculty of Engineering  
J.N.V. University, Jodhpur  
Rajasthan, India

**Abstract**—The ultimate goal of a stereo vision system is to reconstruct 3D geometrical models of scenes. Once a 3D model is available, it can be used in many real life applications such as autonomous navigation of robots, metrology, artificial eyes, etc. This paper reports a design of an experimental stereo vision setup for generating 3D geometrical models of scenes from stereo pair images. It further reports the results of stereo camera calibration, intrinsic and extrinsic parameters, of the stereo system along with the calibration procedure. These parameters are used to rectify captured stereo pair images for generating the disparity map. The paper further describes the basic steps of 3D reconstruction from a pair of stereo image. Finally, the paper presents a reconstructed scene from the stereo image captured by the stereo system.

**Keywords**— Stereo vision, 3D reconstruction, camera calibration, disparity map.

## I. INTRODUCTION

The stereo vision is based on optics of a set of two pin-hole cameras where three-dimensional real world scene is projected as two-dimensional images and a set of intelligent algorithms to interpret these images. In this, two images of a scene are simultaneously captured by our two eyes. These images are further processed by our brain to recreate the three-dimensional model for visualization, depth perception and many other applications. It has been theoretically established that a set of two projections of a scene captured by two cameras from two slightly different viewpoints are enough to reconstruct three-dimensional model of the scene. Physics of binocular vision system is simple and very powerful. As per pinhole camera image formation concept, it is found that the projection of a scene point in one image captured by one eye is slightly displaced in the image captured by the other eye. The displacement between the locations of a point in the two images is commonly referred as disparity and it is inversely proportional to distance between scene point and eyes. This is the fundamental mechanism used by the stereo vision system to reconstruct the scene and depth perception. The perception of depth which is so intuitive in humans and other animals is eluding researchers in computer vision from past few decades to develop visual perception capabilities in machine.

The ability of a machine to reconstruct a 3D scene from 2D images is extremely useful for many applications in science and industry. One such application is robotics/machine vision where proper distance estimation is

important for obstacle avoidance. Other applications include automatic navigation of mobile robots in an environment where human intervention is dangerous or is unreachable and automatic driving and navigation of road vehicles. The stereo vision is also useful for accurate human face recognition system, reconstructing 3D environment for path planning, retrieving a 3D object, creation of 3D maps and many other applications in engineering and medicine where human vision like capability is required.

Currently, active sensing technologies such as SONAR (Sound Navigation and Ranging), LIDAR (Light Detection and ranging), structured light etc. are used for automatic navigation and other vision applications. These methods are based on emitting energy into the environment and analyzing the reflected pattern. Unfortunately, such techniques are invasive and have limited range, and thus have a restricted application domain [1]. Further, they require special purpose hardware (laser projector) that is bulky, expensive and power consuming. Besides, these methods are sensitive to the reflection properties of the elements in the scene. Passive sensing approaches, such as multi view stereo vision are robust and very cheap alternatives because only cameras (two in case of binocular stereo) and a computer are required and no energy emission is involved. However, this technology is in its infancy and requires extensive research to make it a commercially and technically viable alternative to above mentioned methods. Specific issues of concern are computational efficiency and improper reconstruction near object boundaries and textureless areas [2].

In this paper, the design of an experimental stereo camera setup for acquiring stereo pair images is presented. It further reports the results of stereo camera calibration in the form of intrinsic and extrinsic parameters of the stereo system along with the calibration procedure. These parameters are used to rectify the captured stereo pair images for generating the disparity map. The paper also describes the basic steps of 3D reconstruction from stereo images. Finally, the paper presents a reconstructed scene from stereo images captured by the stereo system using the disparity map generated by the algorithms developed by the authors in their earlier work [3-4].

## II. RELATED WORK

The major challenge to obtain 3D reconstruction of a scene is to generate an accurate disparity map in real time.

The high computation time and inaccuracy in disparity map is still the main hurdle for the real time application of stereo vision. Hence, there is a need to develop a robust and accurate method for disparity map computation. The stereo correspondence methods used to generate disparity map can be broadly divided into local and global based approaches[5]. The local methods have potential for real time implementation but do not produce good result whereas global methods produce better results but are computationally expensive.

The local methods are further classified into window based method and feature based method. In a window based methods pixels in a small window surrounding the pixel of interest are used to find a corresponding point of maximum similarity. Finding the correct window size and shape is a critical issue which represents a trade-off between obtaining good disparities in low textured regions versus precisely outlining depth boundaries. Various other window based approaches work on the basis of bilateral filtering and assigning and adjusting adaptive weight to each pixel of the window by using photometric and geometric relationship. Bilateral filter output at each pixel is a weighted average of its neighbours, whereas, adaptive weight technique uses color similarity and spatial proximity to the center pixel. These methods have proven to be very effective and produced excellent results but they are computationally demanding.

The feature based methods generally extract features and then perform stereo matching on these extracted features. This results in generation of a sparse disparity map. Values have to be interpolated by some heuristic to generate dense disparity map. This method holds the potential of real time stereo vision application where dense disparity map is not required. Currently, very few papers have been reported on this method.

The global methods are described as energy minimization problem which can be optimized by various methods like dynamic programming, graph cut and belief propagation. Energy based methods attempt to model some global image properties that cannot be captured by local correlation techniques. Dynamic programming finds optimal solutions for each scanline at a time. The objective is to find the path with the smallest accumulated cost in the disparity space image using the ordering constraint. This is computationally efficient but cannot be applied to optimize an objective function having a two-dimensional smoothness term. Main strength of the method is its real-time capability and to explicitly identify occlusions in both views. On the other hand, graph cut and belief propagation based approaches represent an efficient means to optimize a two dimensional cost function, consisting of a data term that measures the pixel dissimilarity and a smoothness term that penalizes neighbouring pixels assigned to different disparities. The optimization of this cost function is shown to be NP-complete. Graph cuts are a powerful optimization method but, they are only applicable to a limited set of energy functions and are computationally expensive. A belief propagation algorithm produces good results but is often too slow for practical stereo vision use because of large number of iterations and is known to be memory intensive. Efforts have been made to speed up the belief propagation method.

### III. STEREO CAMERA EXPERIMENTAL SETUP

A stereo camera system consists of two cameras looking at the same object from different positions. Fig.1(a) shows the experimental setup for taking stereo images. It consists of two identical color digital cameras. The two cameras are encased in a stable box to avoid relative motion during operation. It is operated by a computer program to simultaneously capture stereo pair images to avoid photometric variations. The captured images are in RGB format of size 480 x 640 in pixel unit and stored in the computer.

#### A. Stereo Camera Calibration

Before using the stereo camera for capturing images it must be calibrated. The stereo camera calibration is a process of establishing interrelationship between the projections of the two cameras of the stereo system using a checkerboard image. The checkerboard image consists of patterns of fixed size black and white squares as shown in Fig.1(b). The checkerboard is selected because corners of square patterns are very easy to identify and its geometry is very simple. Calibration is carried out by acquiring a number of stereo image pairs of the checkerboard. To capture the sequence of images, a computer program was developed in Visual C++ using OpenCV. The OpenCV is an open source library developed to promote research in computer vision. The checkerboard with 25mm square size is used to generate calibration image sequences. These images and MATLAB stereo camera calibration toolbox were used for the calibration of the experimental stereo setup.

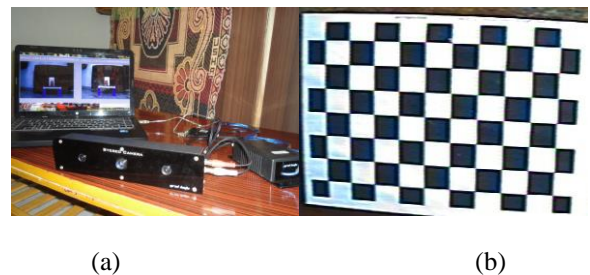


Fig.1. (a): A stereo Camera System. (b): A standard checkerboard pattern

The MATLAB stereo camera calibration toolbox can be downloaded from the website [http://www.vision.caltech.edu/bouguetj/calib\\_doc/download/index.html](http://www.vision.caltech.edu/bouguetj/calib_doc/download/index.html). The steps to calibrate the stereo camera using MATLAB are explained as follows:

1. Take images of the checkerboard in different positions and orientations by the stereo camera. Fig. 2(a) and Fig.2(b) show the images in different positions and orientations captured by the left and the right camera respectively.
2. Run *calib\_gui\_normal.m* from the camera calibration toolbox. Calibrate the left images. Save the result for the left images which contains the internal parameters for the left camera.

3. Repeat step 2 to calibrate the right image sequence. Save the result which contains the internal parameters for the right camera.
4. To get the internal and the external parameters of the stereo system run *stereo\_gui.m*. This would load the internal parameters of both the left and the right cameras computed in steps 2 and 3 and compute the external parameters of the stereo system.

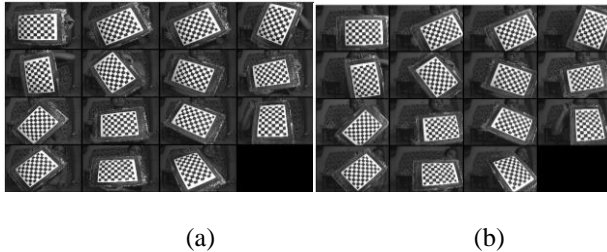


Fig. 2 (a).A sequence of fifteen images captured by the left camera. (b) A sequence of fifteen images captured by the right camera

Table 1 shows the internal and the external parameters of the stereo camera setup computed by the calibration. Once the stereo camera system is calibrated (internal and external parameters are known) any stereo images captured by the stereo camera system can be rectified. After rectification of stereo images, conjugate points lie on the same row in both rectified images. The rectification process can be described as projecting the original images onto a common rectified image plane.

**B. Rules for Taking a Stereo Image Pair**

After calibration, stereo setup is ready for capturing images of a scene. Following rules should be considered during the capturing of the stereo image pairs.

1. Ensure proper illumination of the scene and avoid reflections.
2. If the object shows no texture, project texture onto it, because corresponding points cannot be determined correctly in the areas without sufficient texture.
3. Place the object such that the repetitive patterns are not aligned with the rows of the rectified images.

Fig. 3(a) shows a stereo image pair captured by the experimental stereo camera. Fig.3(b) shows images after rectification. Fig.3(c) shows cropped rectified stereo images used for correspondence and 3-D reconstruction.

TABLE I. THE INTERNAL AND THE EXTERNAL PARAMETERS OF THE STEREO CAMERA SETUP

Intrinsic parameters of left camera:	
Focal Length:	[ 532.49296 581.43605 ] ± [ 8.33144 9.00429 ]
Principal point:	[ 340.31456 253.85841 ] ± [ 3.26328 3.96463 ]
Intrinsic parameters of right camera:	
Focal Length:	[ 545.20558 594.65483 ] ± [ 8.05261 8.60343 ]
Principal point:	[ 346.95355 263.55599 ] ± [ 3.56602 5.03466 ]

Extrinsic parameters (position of the right camera with respect to the left camera):	
Rotation vector:	= [ -0.00400 -0.02502 -0.00693 ]
Translation vector:	= [ 146.15630 2.06938 14.50156 ]

**C. 3D Reconstruction from Stereo Image Pairs**

Following are the basic steps used for 3D reconstruction of a scene from a pair of stereo image. Fig.4 is the block diagram of stereo vision based 3D reconstruction process.

*Step 1: Stereo Camera Calibration:* The camera calibration is essential for any stereo vision application. It is onetime procedure to find out intrinsic and extrinsic parameters of the stereo system.

*Step 2:Image Rectification:* It is a pre-processing step for efficiently solving the correspondence problem. The parameters estimated by camera calibration step (focal lengths, rotation matrix and translation vector) are used for image rectification. The image rectification removes lens distortion and transforms the raw stereo images pair into the aligned stereo images pair, i.e., it projects images onto a common image plane in such a way that the epipolar lines are aligned horizontally. Once the epipolar lines are aligned horizontally, the corresponding points lie in the same scanline of the two images. This reduces the correspondence problem from a two dimensional problem to one dimensional problem.

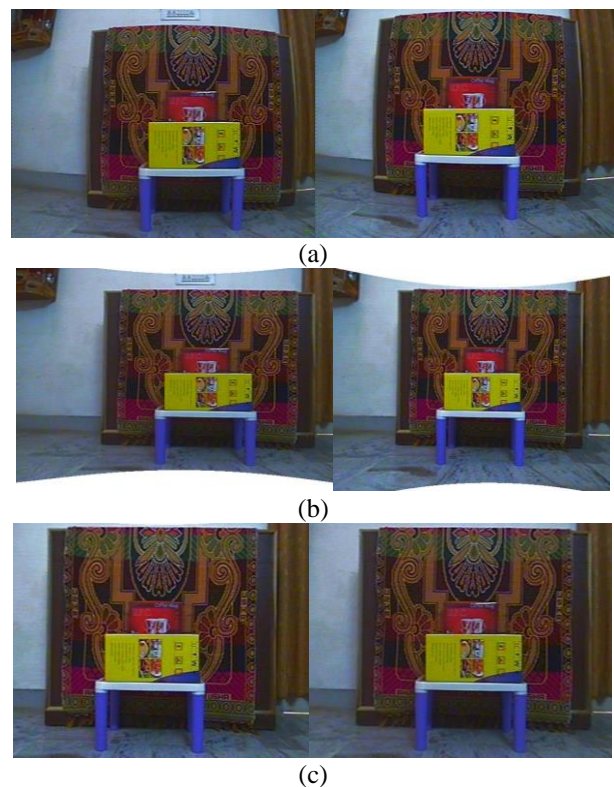


Fig. 3.(a): Raw stereo images(b): Rectified stereo images (c): Cropped rectified stereo images used for stereo correspondence.

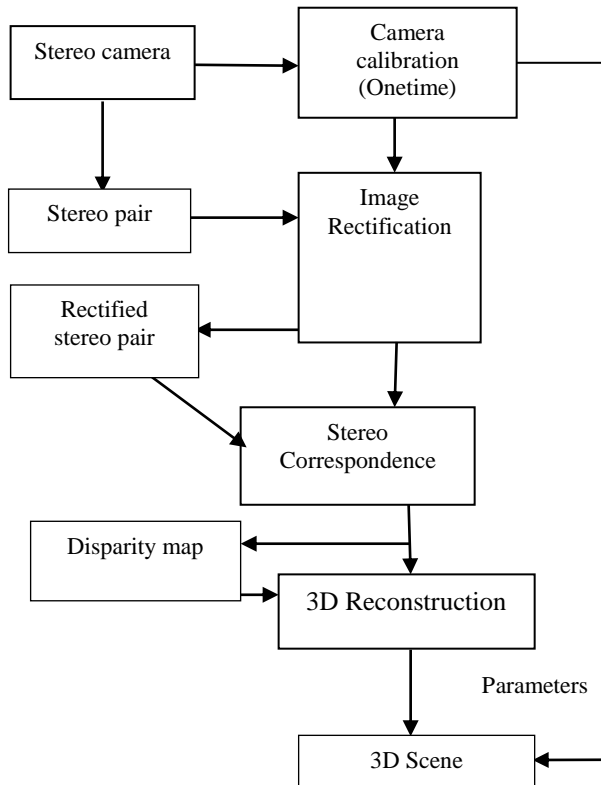


Fig. 4 Block diagram of 3-D reconstruction using a stereo vision system

**Step3: Stereo Correspondence:** The rectified images generated by the previous step are used to generate disparity map by using various correspondence matching algorithms. This step is the backbone of a stereo vision system.

**Step 4: 3D reconstruction:** The disparity map generated in the previous step is used to reconstruct the 3D model of the scene using the stereo system parameters calculated during calibration step. It is the process in which the computed disparity map and the intrinsic and extrinsic parameters are used to compute the 3D coordinates of all points in the scene using triangulation method.

#### IV. RESULTS

The algorithms for computing disparity map discussed in [3] is used on the stereo images pair captured by the experimental stereo setup. The rectified stereo images pair shown in Fig. 3(c) is used for 3D reconstruction. The pair of image is then segmented using method discussed in [6]. The segmentation parameters used are *spatialradius = 3*, *range radius = 3*, *minimum region size = 35*. Fig.5(a) shows the disparity map computed by the dynamic programming based correspondence algorithm [3].

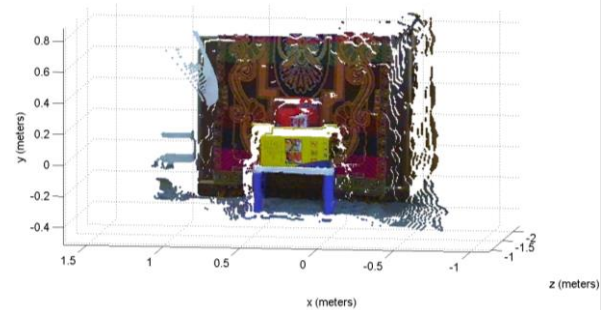
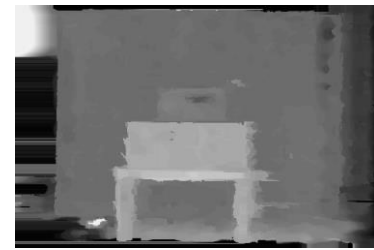


Fig. 5(a) The disparity map. (b)The reconstructed 3-D scene

Disparity map contains the depth information of the scene points. Points in the disparity map which appear brighter are nearer to the camera and which appear darker are farther from it. The 3D reconstruction of the scene from the disparity map using the intrinsic and the extrinsic parameters is computed and is shown in Fig.5(b).

#### CONCLUSIONS

The 3D reconstruction of a scene, using a stereo vision system, requires parameters of the stereo system and the disparity map of the scene. In this paper, an experimental setup of a stereo vision system for 3D reconstruction is discussed. The calibration of the system is carried out using a checkerboard pattern. After camera calibration, the image rectification of captured scenes is done using stereo parameters. The disparity maps of scenes are generated by dynamic programming based method. Further, 3D models are created in MATLAB using the disparity map and the stereo parameters, using principle of triangulation method.

#### REFERENCES

- [1] Lin M.H. and Tomasi C., "Surfaces with Occlusion from Layered Stereo", PhD Dissertation, Stanford University, 2003.
- [2] Sizintsev M. and Wildes R. P., "Coarse-to-Fine Stereo Vision with Accurate 3-D Boundaries", Technical Report CS-2006-07.
- [3] Rachna, H. S. Singh and A. K. Verma "A Stereo Matching Using Dynamic Programming with Segment-Support", *ISST Journal of Mathematics & Computing System*, vol. 2 (2), pp. 37-42, 2012.
- [4] Rachna, "3-D Object Reconstruction from Stereo Image Pairs", PhD thesis, JNV University, 2013.
- [5] Nalpantidis L., Georgios C.S. and Gasteratos A., "Review of Stereo Vision Algorithms: from Software to Hardware", *Int. Journal of Optomechatronics*, Vol. 2, pp. 435 - 462, 2008.
- [6] Comaniciu D. and Meer P., "Mean Shift: A Robust Approach Toward Feature Space Analysis", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, , 2002.