An Approach to Calculate Depth of an Object in a 2-D Image and Map it into 3-D Space

Ashis Pradhan
Dept. of Computer Science &Engg.
Sikkim Manipal Institute of Tech.
Majhitar, East-Sikkim, India.

Ashit Kr. Singh
Dept. of Computer Science &Engg.
Sikkim Manipal Institute of Tech.
Majhitar, East-Sikkim, India.

Shubhra Singh
Dept. of Computer Science &Engg.
Sikkim Manipal Institute of Tech.
Majhitar, East-Sikkim, India.

ABSTRACT

The essence of an image is a projection from a 3-D scene onto a 2-D plane, during which the depth information is lost. The 3-D point corresponding to a specific image point is constrained to be on the line of sight. From a single image, it is very difficult to determine the depth information of various object points in an image. If two or more 2-D images are used, then the relative depth point of the image points can be calculated which can be further used to reconstruct the 3-D image by projecting the image points which includes the depth information as well. This paper presents two techniques namely binocular disparity and photometric stereo for depth calculation and 3-D reconstruction of an object in an image as it requires minimum user intervention.

Binocular disparity method requires a pair of stereo images to compute disparity and depth to generate the desired 3-D view whereas the photometric stereo method requires multiple images under different light directions.

Keywords: Feature point, Binocular disparity, Edge detection, Depth, Photometric stereo, Normal map, Highlight.

1. INTRODUCTION

These days there is an increasing need for 3-D reconstruction in the areas as diverse as medical imaging and artistic applications, movie industry, product design, games and virtual environments. The manual creation of 3-D models is time consuming and therefore cost expensive. The previous works for reconstruction of a real world scene from images is heavily based on projective geometry which are not fully satisfying because they require user driven model and 3-D scanners which are long, error prone and costly.

Usually, only 2-D representations of our 3-D world exist. Images of a scene provide very valuable information, but they simply lack the spatial information that we are commonly used for navigating around our environment. A partial reconstruction of the scene will prove useful in providing a better visualization of that scene for closely inspecting the environment and as a result give the user a sense of presence without actually ever being in the scene. Commonly used technique for such conversion is to develop a depth map for each frame of 2-D image. When observing the world, the human brain usually integrates the depth cues for the generation of depth perception. The major depth perceptions are binocular depth cues from two eyes and monocular depth cues from a single eye.

The common approach is to find the relationship between the different images by calculating disparity. Another method is photometric stereo which utilize reflection models for estimating surface properties from transformations of image intensities that arise from illumination changes. In addition photometric stereo methods are simple and elegant for Lambertian diffuse models. In this project report a Methodology for a similar 3-D reconstruction has been proposed which is less expensive and less mathematically complicated.

Since, the realistic view of an object cannot be seen in a 2-D image because its depth information cannot be recorded. The objective of this report is to create a 3-D view of an object from two or more images for better understanding. The work here describes an efficient 2-D-to-3-D conversion method based on the region of interest and depth information.

2. RELATED WORK

The mostly used technique for reconstructing the 3-D image is either from a single image or from multiple images. However the complexity for designing such systems depend on the number of images involved, the type of images used and the method used to calculate depth.

Methods for single image reconstruction commonly use cues such as shading, texture and vanishing points [1, 2]. These methods pose some restrictions in reconstruction by placing some constraints on the properties of reconstructed objects. (E.g. reflectance properties, viewing conditions and symmetry.) To solve this problem two or more images can be used to reconstruct a 3-D scene as introduced by Ted Shultz and Luis A. Rodriguez [3]. They have used a method to partially reconstruct a real scene using two static images taken from an un-calibrated camera. An advantage of this algorithm is that if a quick 3-D reconstruction is required not much computation work is necessary and depth accuracy of the reconstructed scene can be improved. There can be different techniques to calculate this depth as stated by Sebastian Roy [4] who has used the Maximum Flow Algorithm. The authors Lee Sang-Hyun, Park Dae-Won, Jeong Je-Pyong and Moon Kyung-II has proposed an algorithm which is composed of the estimation of depth levels by using Modulation Transfer Function (MTF) squeeze model and determination of gradient map related to each depth level [5]. Another method for depth estimation is using fundamental matrix [6] as presented by Arne Henrichsen but this results in complex calculations.

A different technique named photometric stereo is used in [7] by Woodham who found that this method is good if surface gradient is to be determined and works best when on smooth surface with few discontinuities. Similar approach is presented by Carlos Hernandez, George Vogiatzis and Roberto Cipolla in [8] to address the problem of obtaining a complete and detailed reconstruction of an object. The binocular disparity and photometric stereo methods used here are motivated by the method used in [3, 7, 11].

3. METHODOLOGY

3.1 Binocular Disparity

Assume that we have two eyes/cameras and the optical axes of the two cameras are parallel, i.e. the eyes have the same Z direction, but one eye is displaced to the right (positive X direction) by a distance T_x (called as the baseline of the eye pair). Based on these assumptions, a 3-D point with coordinates (X_0, $Y_{0,}$ Z_0) in the left eye's coordinate system would have coordinates ($X_0 - T_x$, Y_0, Z_0) in the right eye's coordinate system. As such, this 3-D point would project to a different x value in the left and right images. The difference in x position is called the binocular disparity.

$d = X_l - X_r$ ------------------------ Equation (1)

Where,

d is the disparity. X_l is the x-coordinate of a point on left image, & X_r is the x-coordinate of the same point on right image.

3.1.1 Facts/Definitions about disparity

Consider that both the eyes are fixated on a point (i.e. fixation point) in space. Then there are following conditions:

i. If the object is located further than the fixation point then the disparity will hold a positive value.

ii. If the object is located closer than the fixation point then the disparity will hold a negative value.

iii. If the object is located at the fixation point then disparity will be zero.

3.1.2 Calculation of Disparity

With two images of the same object captured from slightly different view points, the binocular disparity can be utilised to calculate the depth of an object.

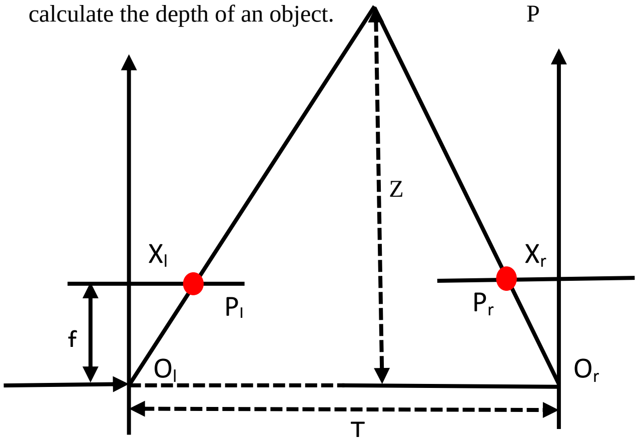

Fig 1: Disparity [7]

In the figure above, assume that P_l and P_r are the projections of the 3-D point P on the left and right image; O_l and O_r are the origin of camera coordinate systems of the left and right cameras; f is the focal length, T is the baseline (i.e. is the distance between two cameras) and Z is the depth. Based on these parameters, depth can be calculated as described in the next section.

3.1.3 Calculation of depth

The formula given below has been used to calculate depth from a pair of images.

Z = f * (T/d) ------------------------ Equation (2)

Where,

Z is the depth to be calculated, f is the focal length, T is the baseline & d is the disparity (as discussed in the previous section)

3.1.4 Feature point identification of an object

In this paper, SURF matching algorithm has been used for collecting interest points from each image as it gives better matching results. SURF (Speeded Up Robust Feature) is a robust local feature detection technique which was first presented by Herbert Bay el al. ECCV 9th in May 2006 in an international Conference on Computer Vision held in Austria. This technique can be used in computer vision tasks like object recognition or 3-D reconstruction. It is inspired by SIFT descriptor but several times faster than SIFT.

3.1.4.1 Matching of feature points

There are two ways to match feature points in two images.

- Get the characteristic points of the first image and its descriptor and to do the same with the second image. So it becomes easier to compare the two images descriptor correspondences between points and establish some kind of measure.
- Get the characteristic points of the first image with the descriptor. Now compare this descriptor with the points of the second image which is believed to be the partner concerned.

3.1.5 Approach

The following steps have been followed to project 2-D feature points in 3-D space.

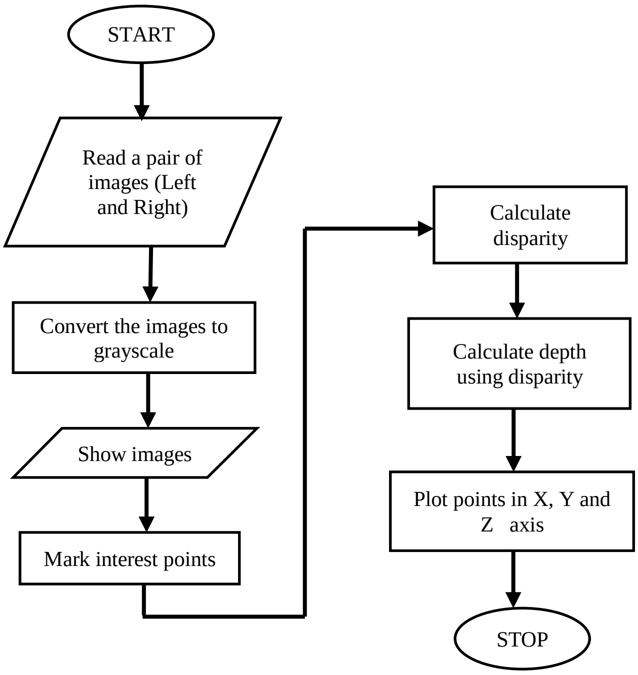
START
Read a pair of images (Left and Right)
Convert the images to grayscale
Show images
Mark interest points
Calculate disparity
Calculate depth using disparity
Plot points in X, Y and Z axis
STOP

Fig 2: Flowchart to plot X, Y and Z coordinates of 2-D feature points

The detailed steps carried are as follows:

i. **Identification and selection of feature points from 2-D image:** First identify which points have to be marked for plotting in 3-D space so that it will give a realistic view of an object. One has to be very careful while marking these points as the wrong selection of points will give erroneous result. To solve this problem Canny edge detection algorithm [8] has been used which will clearly identify the edges of an object and mark the points automatically.

ii. **Disparity calculation:** This step involves the calculation of disparity (using Eq. 1) from the points marked on the two images as illustrated in step 1.

iii. **Depth estimation:** This step involves the calculation of depth (using Eq. 2). Depth estimation has been based on the fact that camera moves in the X direction.

iv. **Mapping of 2-D feature points in 3-D space:** After all the feature points has been marked, the last step involves the mapping of these points in 3-D space i.e. X, Y and Z axes so that it will give a clear view of an object which helps in better understanding.

3.2 Photometric Stereo

Photometric Stereo is an approach to reconstruct a 3-D surface from a series of images of a diffuse object under different light sources. These light sources are ideally point sources some distance away in different directions, so that in each case there is a well-defined light source direction from which to measure surface orientation. Therefore, the change of the intensities in the image depends on both local surface orientation and illumination direction.

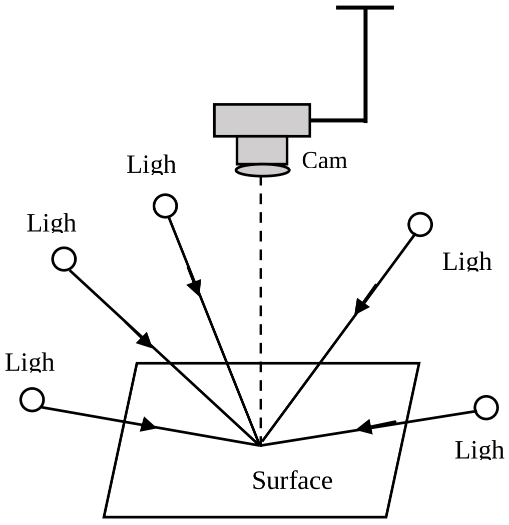

Fig 3: Illustration of photometric stereo geometry

Photometric stereo uses several images of the same surface under different illumination directions. The advantages of photometric stereo are:

- Unlike single image shape from shading algorithms, photometric stereo makes no assumption of the smoothness of the surface.
- It requires only additional lighting and can be easily implemented in at a reasonable computational cost.
- Each image brings along its own unique reflectance map, therefore each image will define a unique set of possible orientations for each point.
- Photometric stereo can recover not only surface orientation but also surface albedo.

3.2.1 Algorithm for photometric stereo

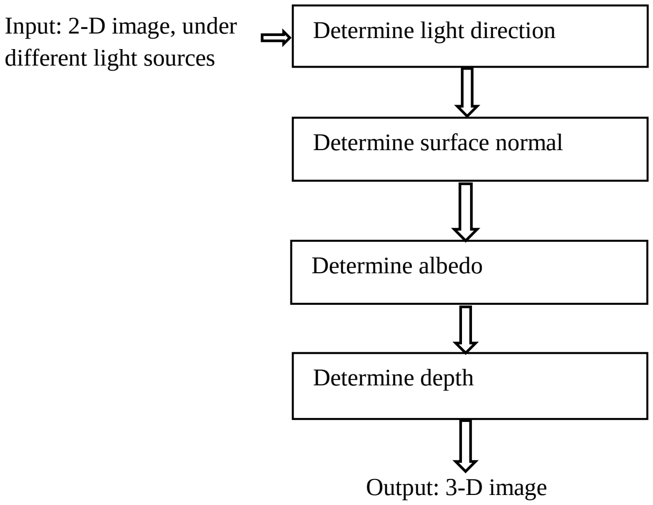
Input: 2-D image, under different light sources
Determine light direction
Determine surface normal
Determine albedo
Determine depth
Output: 3-D image

Fig 4: Block diagram for photometric stereo algorithm

*3.2.1.1 Determine light direction***:** Light Direction is essential for this method to work, so how we can get the light direction? The trick is placing a Chrome ball beside the object. When light falls on this chrome ball it will shine in the direction of light from which we can get the direction.

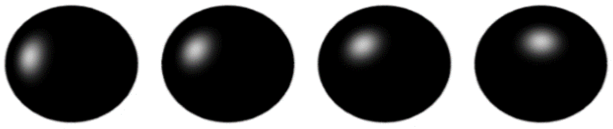

Fig 5: The highlight in the chrome ball gives the direction of the source of light

The highlight on the chrome ball can be seen when V=R

Then, $L = 2(N.R)\,N – R$ --------------- Equation (3)

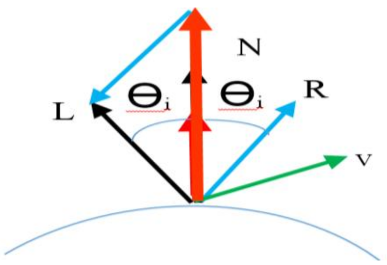

Fig 6: Reflection of light from a chrome ball

3.2.1.2 Determine surface normal: After knowing the light direction we can easily calculate the normal at each pixel. For the lambertian surface three images of an object are enough to construct the 3-D view. Intensity at any point on the surface can be given as

$I = K_d L n^T$ ----------------- Equation (4)

Where I is the pixel intensity, K_d is the albedo (reflection coefficient) or length of the vector, L is the reflected light direction (a unit vector) and n is the unit surface normal.

$Q = \sum w_i (I_i - L_i g^T)^2$ ------------------ Equation (5)

Where,

$$w_i = \begin{cases} I_i \,, & I_i \le 0.5 \\ 1 - I_i \,, & I_i > 0.5 \end{cases}$$ ----------------- Equation (6)

w_i is the weighting scheme and

$g = K_d n$ ----------------- Equation (7)

3.2.1.3 Determine albedo: After calculating the surface normal, albedo at each pixel can be calculated using the equations given below:

$$K_d = \frac{\sum_i I_i L_i.n}{\sum_i (L_i .n)^2}$$ ------------------------------Equation (8)

3.2.1.4 Determine depth:

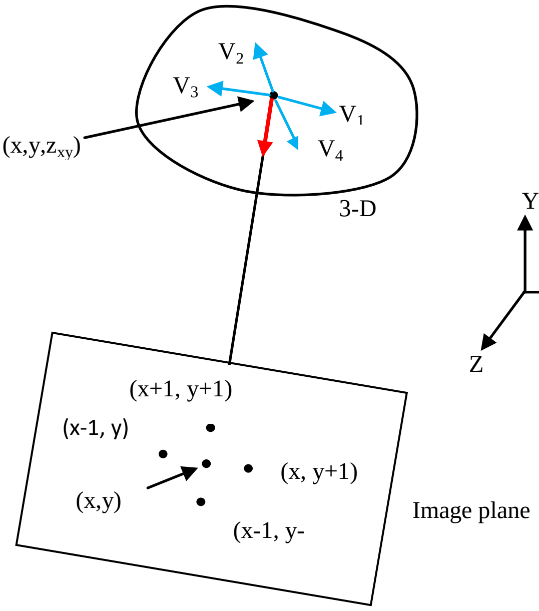

Fig 7: Projecting 2-D points in 3-D space

For V_1 we have,

$V_1 = (x+1, y, z_{x+1, y}) - (x, y, z_{x, y}) \quad = (1, 0, z_{x+1, y} - z_{x, y})$

$O = N.V_1 \quad = (n_x, n_y, n_z) . (1, 0, z_{x+1, y} - z_{x, y})$

Now we have $N.V_1 = 0$

$(n_x, n_y, n_z) . (1,0, z_{x+1, y} - z_{x, y}) = 0$

$n_x + n_z (z_{x+1,y} - Z_{x,y}) = 0$ ------------------ Equation (9)

For V_2

$V_2 = (x, y+1, z_{x, y+1} - (x, y, Z_{x, y})$

$= (0, 1, z_{x, y+1} - z_{x, y})$

$O = N.V_2$

$= (n_x, n_y, n_z) . (0, 1, z_{x, y+1} - z_{x, y})$

Now we have $N.V_2 = 0$

$(n_x, n_y, n_z) . (0, 1, z_{x, y+1} - z_{x, y}) = 0$

$n_y + n_z (z_{x,y+1} - Z_{x, y}) = 0$ ------------------ Equation (10)

Similarly for vector V_3 and V_4 we have the following equations:

$- n_x + n_z (z_{x-1, y}, z_{x, y}) = 0$ ------------------- Equation (11)

$- n_y + n_z (z_{x, y-1}, z_{x,y}) = 0$ ------------------- Equation (12)

This will generate a sparse matrix of (2*npixel, npixel)

Where, npixel is the number of pixels in the masked region and not the entire region. These system of equations are solved by least square method.

M z = v

Now, z = M\v where, '\' stands for mldivide

4. RESULT AND DISCUSSIONS

4.1 For binocular disparity

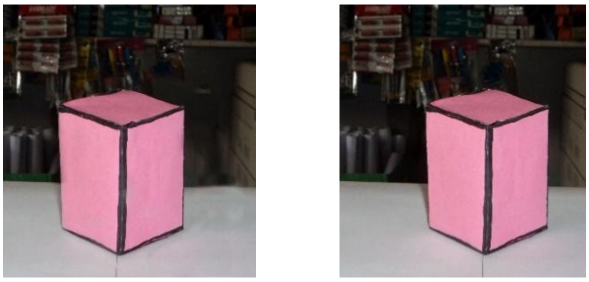

Fig 8.1: Left image of an object

Fig 8.2: Right image of an object

Fig 8: Two images of a cube in a scene

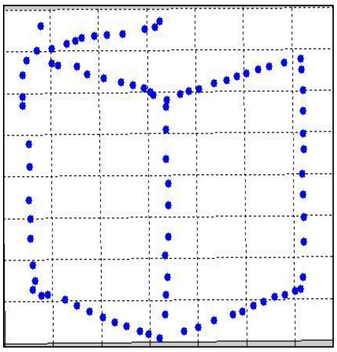

Fig 9: Plotted points of the cube in 3-D space

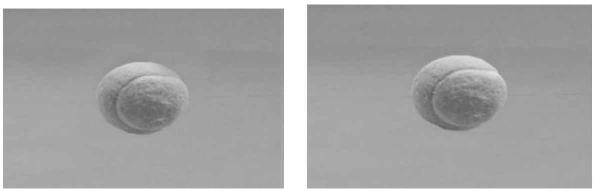

Fig 10.1: Left image of a ball Fig 10.2: Right image of a ball

Fig 10: Two images of a ball in a scene

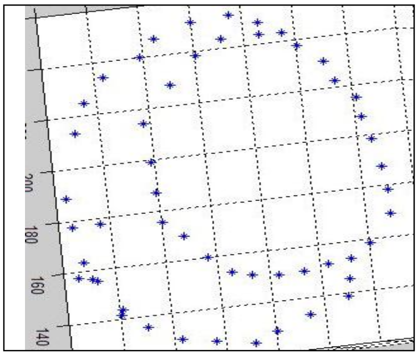

Fig 11: Plotted points of the ball in 3-D space

Figure 8.1 and 8.2 show the left and right (stereo) images of a cube in a scene. These images are taken as the input images. Figure 9 shows the plotted points of the cube in 3-D space which is the output of the images shown in figure 8.

Similarly, figure 10.1 and 10.2 show the left and right images of a ball and figure 11 is the output of the image shown in figure 10.

As it is clear from the above output that Binocular disparity method does not give a very clear view of an object so for better results the photometric stereo method has been used as described in the next section.

4.2 For photometric stereo

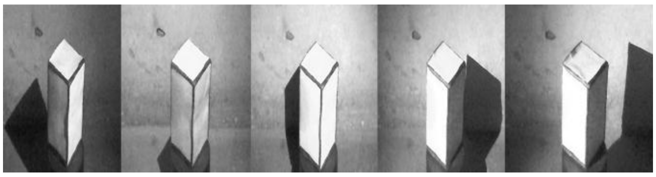

Fig 12: Input images of cube having different intensities

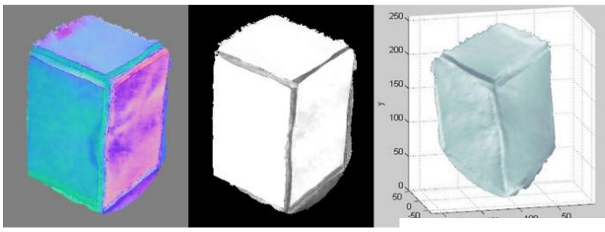

Fig 13.1: Normal map Fig 13.2: Albedo Fig 13.3: 3-D view

Fig 13: Output images

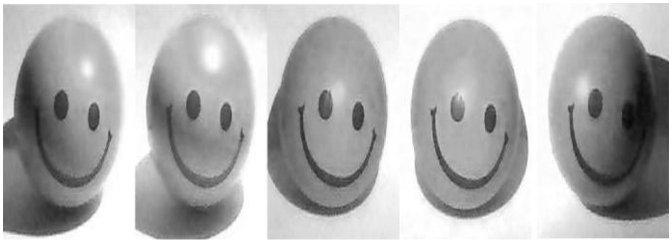

Fig 14: Input images of a smiley ball having different intensities

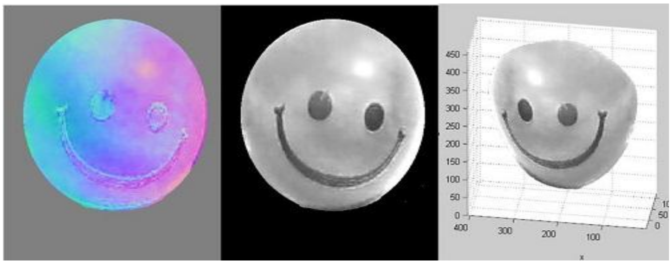

Fig 15.1: Normal map Fig 15.2: Albedo Fig 15.3: 3-D view

Fig 15: Output images

Figure 12 shows input images of a cube showing different intensities of the same pixel in all the images. These images are then used to compute normal map, albedo and the final 3-D view of the object as shown in figure 13.

Figure 14 shows input images of a smiley ball clicked such that the object and camera positions remains fixed and only the light direction has to be changed. Again, figure 15 shows the normal map, albedo and the final 3-D view of the object.

5. CONCLUSION

This paper proposes two techniques which are efficient for converting a 2-D image into 3-D for better understanding and visualization of a real scene. The method proposed first takes two stereo images as an input and uses a method named binocular disparity to plot the same in 3-D space. An advantage of this method is that if a quick 3-D reconstruction is desired then not much mathematical computation is required, no camera calibration is required and no intrinsic or extrinsic camera parameters are to be determined.

The second method used here is photometric stereo, which takes multiple images. [13] has suggested to use 12 images of an object to construct 3-D view but this work uses only 5 images. The output obtained here shows that it does not give as clear view as the one with 12 images as an input. This concludes that distant light sources and more number of images will give better result. Future work is to build an algorithm which requires minimum number of images to produce good quality results.

5.1 Comparative Analysis

Binocular Disparity	**Photometric Stereo**
• Number of images required is two.	• Minimum number of images required is 3.

• Easy mathematical calculations.	• Complex mathematical calculations.
• Does not give a very clear view in 3-D space.	• Gives better result as compared to Binocular Disparity.

6. ACKNOWLEDGEMENT

This project required great amount of work, research and dedication. Still, implementation would not have been possible if we did not have a support from Sikkim Manipal Institute of Technology (SMIT) for the provision of expertise and technical support in the implementation.

7. REFERENCES

[1] Han, F. and Zhu, S.C., "Bayesian Reconstruction of 3-D shapes and scenes from a single image". Proceeding of the first IEEE International Workshop on Higher Level knowledge in 3-D Modeling and Motion Analysis. 2003

[2] Rother, D., Patwardhan, K., Aganj, I. and Sapiro, G., "3-D Priors for Scene Learning from a Single View." S3-D Workshop (at Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition). 2008.

[3] Ted Shultz and Luis A. Rodriguez, "3-D Reconstruction from two 2-D images", ECE 533 Fall 2003.

[4] Sebastian Roy and Ingemar Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In IEEE Proc. of Int. Conference on Computer Vision, pages 492–499, 1998.

[5] Lee Sang-Hyun, Park Dae-Won, Jeong Je-Pyong and Moon Kyung-Il, "Conversion 2-D Image to 3-D Based on Squeeze function and Gradient Map", International Journal of Software Engineering and Its Applications Vol.8, No.2, 2014.

[6] Arne Henrichsen, "3-D Reconstruction and Camera Calibration from 2-D Images", Department of Electrical Engineering, University of Cape Town, December 2000.

[7] Assoc. Prof. Dr. Ir. E. A. Hendriks Dr. Ir. P. A. Redert, "Converting 2-D to 3-D: A Survey", Information and Communication Theory Group (ICT) Faculty of Electrical Engineering, Mathematics and Computer Science Delft University of Technology, the Netherlands, December 2005.

[8] Rashmi, Mukesh Kumar, Rohit Saxena, "Algorithm and Technique on Various Edge Detection: A Survey", Signal & Image Processing: An International Journal (SIPIJ) Vol.4, No.3, June 2013.

[9] Robert J. Woodham, "Photometric method for determining surface orientation from multiple images", Department of Computer Science, University of British Columbia, Optical Engineering, 1980.

[10] Carlos Hernandez, George Vogiatzis, Roberto Cipolla, "Multi-View photometric stereo", Pattern Analysis and Machine Intelligence, IEEE Transactions, March 2008.

[11] Aaron Hertzmann, Steven M. Seitz, "Example-Based photometric stereo: Shape reconstruction with general, varying BRDFs", Pattern Analysis and Machine Intelligence, IEEE Transactions, Vol 27, August 2005.

[12] S. Barsky, "Surface Shape and Colour Reconstruction using Photometric Stereo", PhD thesis, School of Electronics and Physical Science, University of Surrey, UK, 2003.

[13] C. Hernandez, F. Schmitt, and R. Cipolla, "Silhouette coherence for camera calibration under circular motion," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 2, pp. 343-349, Feb 2007.

[14] Hayakawa, Hideki. "Photometric stereo under a light source with arbitrary motion." Journal of the Optical Society of America, 1994. Available online at http://pages.cs.wise.edu%7Ecs7661/projects/phs/hayakawa.pdf

[15] Zhang, R.; Tsai, P.; Cryer, J.E.; Shah, M. (1999), "Shape from Shading: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence archive, Vol. 21, Issue 8, Pages: 690-706.

[16] Ning Qian, " Binocular Disparity and perception of Depth.", Center for Neurobiology and Behavior Columbia University, Neuron, Vol. 18, 359-368,March 1997.

[17] Seitz, Steven. "Computer Vision (CSEP 576), Winter 2005 - Project 3: Photometric Stereo." Available online at: http://www.cs.washington.edu/education/courses/csep576/05wi/projects/project3/project3.htm

[18] Adam Bechle. "Computer Vision (CS 766),2008 Project 3: Photometric Stereo" Available online at http://pages.cs.wisc.edu/~lizhang/courses/cs766-2008f/projects/phs/students/bechle/website.html

[19] Dr. SukhenduDas. "Computer Vision (CS 635), Shape from Shading." Available online at http://www.cse.iitm.ac.in/~vplab/courses/CV_DIP/PDF/ShapeFromShading.pdf