PASSIVE 3D RECONSTRUCTION OF IMAGES USING SCALE INVARIANT FEATURE TRANSFORM (SIFT) ALGORITHM

K Sreenivasa Reddy[1], Dr. Pramod Pandurang Jadhav[2]

Article History: Received: 06.04.2023 **Revised:** 26.05.2023 **Accepted:** 21.06.2023

Abstract

A key challenge in computer vision is to reconstruct the three-dimensional geometry of an item or scene from two dimensional pictures taken with various sensors. The geometric structure of real-world objects may be recovered using a variety of approaches that are used in gaming, medical imaging, mechanical component inspection, object identification, and other fields. The precision of the reconstruction, the complexity of the calculation, and the speed all affect how well 3D reconstruction algorithms function. Passive and active 3D reconstruction techniques are examined in the quest to create an effective approach, and a few well-known techniques are selected based on the knowledge found in the literature at the time. To enhance 3D reconstruction performance, these techniques are updated. The goal of this project is to provide approaches for reconstructing the 3D structure of real-world objects or scenes from photographs taken with various sensors and to assess how well each approach performs in terms of accuracy, complexity, and speed. It is challenging to get precise depth information from a single picture, hence two or more photos must be used to complete 3D reconstruction. In the suggested study, stereo vision is used to execute a passive approach of 3D reconstruction. To determine the depth of a scene or an item, two photos are taken. Scale Invariant Feature Transform (SIFT) is used to extract and match key points. Since SIFT produces more key points than other feature extraction techniques, it is used because it produces more features overall. The camera calibration toolkit is used to calibrate cameras. Epipolar rectification is carried out after calculating the basic matrix. The region-based stereo method with global error energy reduction is used to create the disparity map. The disparity map is then used to create the depth map. The depth information is not retrieved in certain locations when the findings are compared to ground truth values. The 3D reconstruction is carried out using numerous photos to get over this limitation. Newline

Keywords: 3D reconstruction, Camera, Multiple Views, 2D images, depth perception.

[1]Research Scholar [2]Department of Computer Science & Engineering, Dr. A.P.J. Abdul Kalam University, Indore (M.P.) - 452010
[2]Associate Professor, Department of Computer Science & Engineering, Dr. A.P.J. Abdul Kalam University, Indore (M.P.) - 452010

DOI: 10.31838/ecb/2023.12.s3.523

1. INTRODUCTION

One of the core problems in image-based modelling and computer vision is producing a 3D model from 2D photos that is as realistic as feasible. The ideal option is to recreate the scene automatically with minimal to no human input. There are various ways for creating 3D models from 2D photos at the moment; each algorithm has its own execution requirements, advantages, and disadvantages. In this article, we define 3D reconstruction from 2D photos and outline its application fields [1]. From these 2D photos, we suggest a number of 3D reconstruction strategies and compare them.

When compared to passive approaches, active methods of 3D reconstruction have become more popular recently. Active reconstruction techniques that use carefully regulated lighting are used to address the correspondence issue that results from passive reconstruction techniques. The major use of active approaches is to increase the precision of 3D reconstruction. Active reconstruction primarily employs two methods: structured light and time of flight [2]. A well-known methodology for obtaining precise and trustworthy outcomes in reconstruction is structured light technique. Low-cost cameras and projectors are used to create the structured light technology. In the literature, many configurations for applying this approach for 3D reconstruction have been suggested.

Using digital image correlation, the difference between the matching spots was calculated. The converging optical axis model was found to recreate the surfaces more accurately based on the experimental findings. It was also covered how to record data, create organized binary and grey code patterns, and turn 3D points into surface meshes while building a 3D scanner with one or more cameras and a projector. In this paper, we provide a thorough and organized analysis of current developments in deep learning-based 3D object reconstruction[4]. We initially concentrate on general forms before talking about particular examples, including reconstructing the contours of human bodies and faces and parsing 3D scenes. We provide a novel approach for formulating features that are taken from pictures captured with varying tilt degrees using the SIFT technique.

2. RELATED WORK

All application areas heavily rely on 3D object reconstruction. Many approaches are put forward for reassembling the depth or 3D information from 2D photos captured with various sensors. These techniques may be roughly divided between passive and active techniques. Both passive and active approaches of 3D reconstruction have been investigated in this study; a select number of the methods have been adjusted to produce an effective algorithm based on their performance as described in the literature. a review of 3D reconstruction strategies, including surface integral reduction, surface cutting, visual Hull, multi-view stereo reconstruction, and laser scanning technology [5]. A review of semi-automated and completely automated approaches to 3D modelling of terrestrial pictures and an explanation of the various techniques for image capture, reconstruction, and display were provided [6].

2.1 Passive 3d Reconstruction Using Stereo Vision (3dsv)

To extract keypoints that are resistant to changes in rotation, scale, and illumination, the author suggested the Scale Invariant Feature Transform (SIFT). In comparison to previous feature extraction techniques like SURF and FAST, this approach is employed for identifying a greater number of key points. By measuring the intrinsic and extrinsic camera characteristics, camera calibration is crucial to 3D reconstruction. Photogrammetric and self-calibration are two types of calibration that may be carried out with or without a calibration object, respectively [7]. Flexible camera calibration method in which the camera is made aware of a planar pattern at various angles. The feature points were found in the photos, and for every image, homography was retrieved. Estimating the highest probability was used for refinement. The homographies were used to retrieve the camera settings.[8].

2.2 Passive 3d Reconstruction Using Multi View Images (3dmv)

The prior study uses two photos to generate a 3D reconstruction. Passive reconstruction is done utilising numerous photos to increase the correctness of the 3D outcome. picture quality's impact on the 3D reconstruction process. The structure from motion method, which involves gathering several pictures, feature point recognition using SIFT, and triangulation, is used for 3D reconstruction. By reducing the quality of the photographs by introducing noise, changing the sharpness, and adjusting the JPEG compression ratio, the algorithm's performance was evaluated [9].

Active methods of 3D reconstruction are renowned for more accurately reproducing the scene and the objects in it. The Time-of-Flight method and structured light are used to conduct active 3D reconstruction. By using an experimental set-up made up of line scan cameras, the team adjusted phase shifting and time correlation structured light approaches to produce 3D surface reconstruction. Applications of 3D imaging using structured light and algorithms for camera and projector calibration are also covered. By comparing the scanning system's performance with a commercial scanner, the accuracy, orthogonality, sampling rate, and linearity of the system were assessed. The encoding of patterns in the temporal domain, calibration, decoding, and reconstruction using linear triangulation were also examined [10].

3. PROPOSED WORK

Scanning electron microscopy (SEM) is used in the proposed 3D reconstruction technique to take a number of pictures while varying the tilt angles. Scale Invariant is used to identify key points for the obtained pictures. Feature For all potential picture pairings, transform and matching is done. The two picture pairings with the most matches are selected for further processing. Using the information of the image coordinates of the matched points in both left and right photos, tilt angles used to acquire the images, and scale factor, photogrammetry processing is used to compute the 3D points. The Iterative Closest Point and SFIT algorithms are then used to integrate the 3D points. In terms of computing time and iterations, the performance of the KD tree-based SFIT method is compared to that of the brute force SFIT technique.

3.1 Photogrammetry Processing

Measurements and dimension determination from 2D photos are done via the method of photogrammetry. Results from reconstruction using photogrammetry are accurate when paired with SEM images. SEM pictures are used in photogrammetry techniques to determine the topography of friction surfaces. Using both parallel and central projective imaging geometries, a photogrammetry programme is created, and it is discovered that, at greater scales, the parallel projection approach is better suitable for photogrammetry processing [11]. The suggested 3D reconstruction approach uses SEM to capture several pictures by adjusting the tilt angles. Depending on the magnification range, one may choose a parallel or central projection model. The parallel projection model is selected for the proposed study under the assumption of a wide magnification range.

3.2 Proposed 3d Reconstruction Algorithm

Similar to how humans utilise two eyes to improve depth perception, this study recovers depth from two photos of an item or scene. Scale Invariant Feature Transform is used to extract features that are noise, rotation, and scale invariant (SIFT). Additionally, these characteristics are mostly unaffected by variations in 3D perspective and lighting. Finding the features in both photos with the smallest Euclidean distance is the next step in matching. Using the camera calibration toolkit, the intrinsic and extrinsic camera parameters are calculated. Standard The Fundamental matrix is calculated using an eight-point technique, and epipolar rectification is employed to align the epipolar lines parallel to the picture plane. Using a region-based stereo growing approach, a disparity map is produced, and a depth map is acquired. Finally, a 3D reconstruction is produced utilising the depth map. The flow diagram below shows the many steps in the passive 3D reconstruction process employing stereo vision.

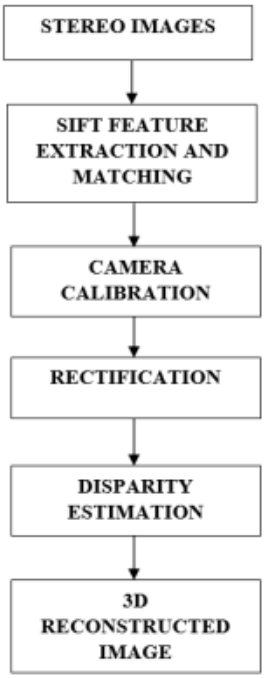

Figure 1 : 3D Reconstruction using Stereo Vision

The stereo pictures obtained with a single camera positioned at various places are used in this passive approach of 3D reconstruction. Utilizing the Scale Invariant Feature Transform, key points are recovered from these stereo pictures.

3.3 Scale Invariant Feature Transform (Sift)

Finding the relationship between the key points in two or more pictures is a common task in computer vision applications including object identification and image alignment. To extract the key points, use the Scale Invariant Feature Transform (SIFT) technique. When compared to other feature extraction methods like SURF and FAST, the approach finds more key points. Using a multi-scale technique, Scale Invariant Feature Transform (SIFT) is utilised to find meaningful and stable features. The discovered features are resistant to adjustments in scale, rotation, noise, and perspective. The technique employs the following phases to extract descriptors from a grey scale image: Extrema detection, Key point Localization, Orientation assignment, and Key point descriptor.

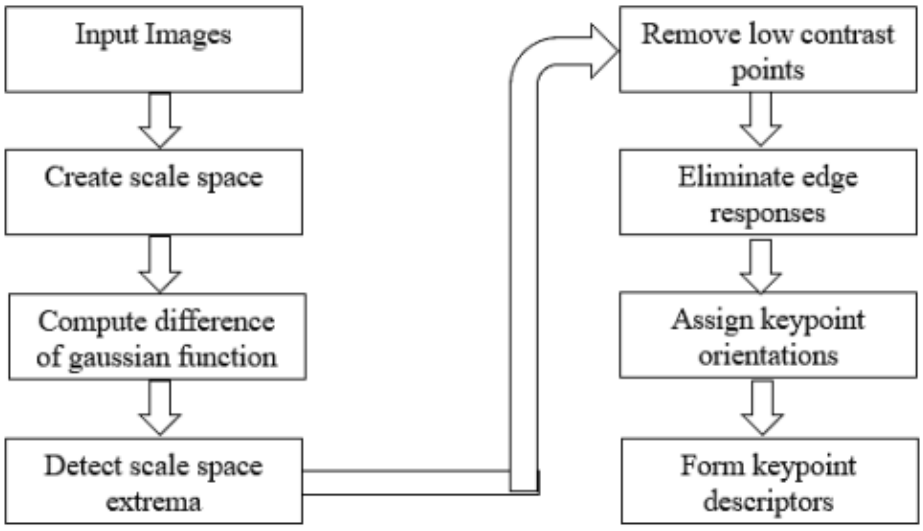

Figure 2 : Scale Invariant Feature Transform

3.4 Extrema Detection

Using the Difference of Gaussian function (DoG) Witkin approach, the local extrema that correspond to the zero crossings of the second order derivative of scaled copies of the picture are found. The local extrema point often corresponds to the edge and corner points. To create a series of scale space blurred pictures, the image is first convolved using Gaussian functions of various values [12]. The collection of scale space pictures creates an octave where the neighbouring scales differ by σ. Images in the scale space that follow the initial image have k times the value of the prior image.

If an octave's number of pictures is thought as as s+1, then k=21/s.
The Gaussian pictures are down sampled by a factor of 2 to create the second octave.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y),$$

Where

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}.$$

To acquire the Difference of Gaussian picture, the difference of nearby scale space images is computed.

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned}$$

The Laplacian of Gaussian filtered picture and the DOG image both use zero crossing detectors to emphasise areas with quick intensity changes, such as corners and edges.

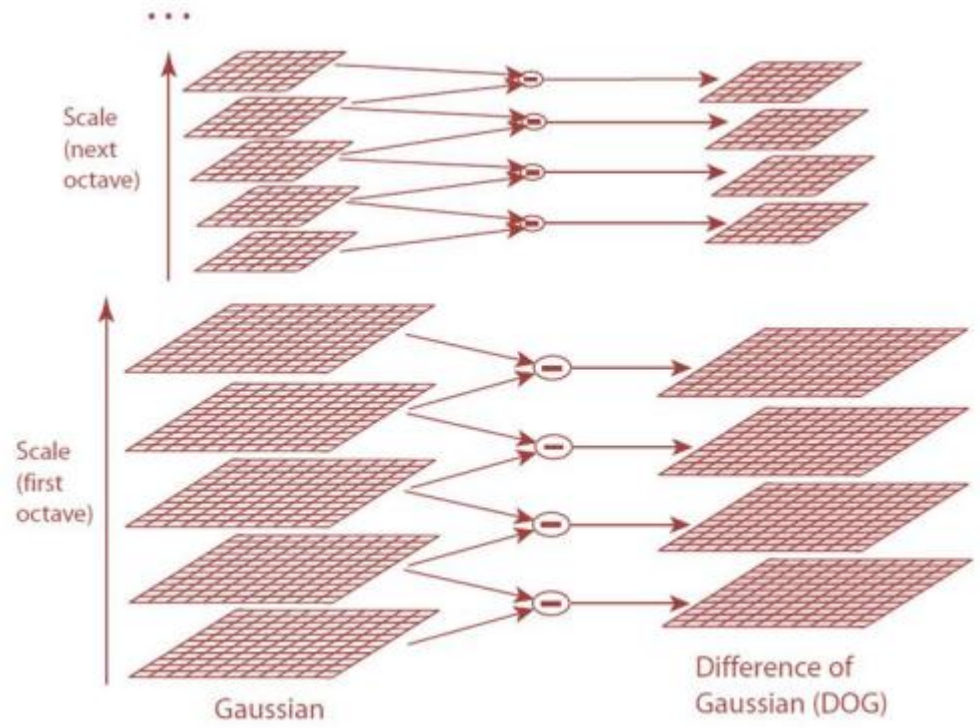

Figure 3 : Scale Space and Difference of Gaussian Images

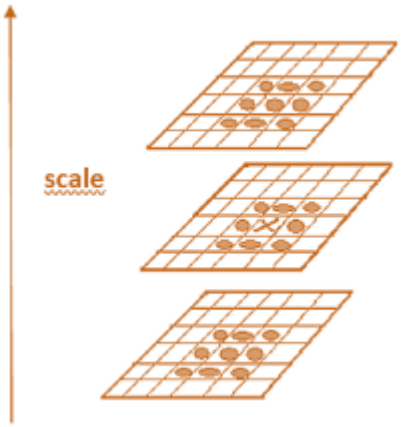

Figure 4 : Detection of Maxima and Minima of DOG Images

Each pixel is compared to its 26 neighbours in neighbouring scales as well as its eight neighbours in the current scale. The key point is regarded as the maximum or minima point.

3.5 Localization of Key Points

The position of the key point is discrete and may be more precisely interpolated. According to their stability, key points are localised. The placement of key points is further refined using a parabolic fit. Some of the key points, which are used to describe the picture, have low contrast values [12]. Through Taylor extension of the scale space, low contrast and edge-located points are removed D(x, y, σ).

$$D(x) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X$$

Low contrast key points are insufficient for stability elimination. The DoG pictures show a significant primary curvature around their edges. There are many key spots along a straight edge. By calculating the point's major curvatures using the Hessian matrix, these points may be removed.

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$$

where H stands for the 2x2 Hessian matrix and r is the ratio of major curvatures.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Based on the ratio of the Eigen values in the Hessian matrix, the edge keypoints are deleted.

3.6 Key-Point Descriptor

The local gradient histograms are used to build the key point descriptors. The gradient histogram is created by computing the magnitude and direction of each pixel in a sample space that is chosen around each of the key points [14]. To boost the magnitude of the pixels around the key point and lessen the impacts of noise, a Gaussian weighting function with a standard deviation of half the window size is applied to each pixel. The histogram with 8 orientation bins is made up of the pixels from each 4 × 4 section. In each sample space, the gradient magnitudes are added for each direction. The SIFT descriptor is a 128 vector in the real implementation since 4 x 4 histograms with 8 orientations are used.

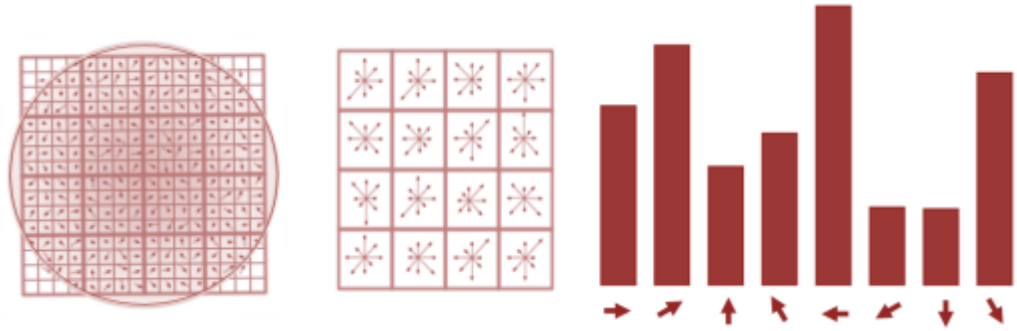

Figure 4: Building a Keypoint Descriptor

3.7 Matrix Computation Basics

The epipolar geometry is encoded using the basic matrix. The epipolar line that will be utilised to find the correspondence points is identified using a fundamental matrix. The associated picture points are used to generate the 3 x 3 fundamental matrix. The construction of this matrix does not need knowledge of the camera specifications [15]. The Eight Point Algorithm is used to calculate the basic matrix in the following steps:

Step 1: Construct the homogeneous system Ax=0. A is n x 9 matrix where n in the number of correspondence. $A = UDV^T$ is the singular value decomposition.

Step 2: Compute SVD of F by $F = U_F D_F V^T_F$

Step 3: Smallest singular value is set to 0; The corrected matrix is represented by D'_F.

Step 4: $F' = U_F D'_F V^T_F$ is the corrected F.

4. EXPERIMENTAL VALIDATION

Python software has been used to accomplish the suggested 3D reconstruction technique employing stereo vision for various stereo pictures. The middle bury website's stereo datasets are used to test the suggested approach. SIFT feature detector is used to find and match features in the stereo pictures. The matching points are used to create the 3 x 3 basic matrix. The left and right photos show the epipolar axes and epipoles. The epipolar lines are then rectified to align them parallel to the baseline. Using a region-based stereo matching approach, the disparity map is calculated, and the depth is extracted from the disparity.

Table 1 Comparison of Feature Detection Methods

S.No	IMAGES	PAIR	SURF	SIFT	FAST
1	Money	A B	64 52	238 242	26 28
2	Keyboard	A B	4028 4143	7785 7652	885 1045
3	Angle0	A B	48 52	68 68	25 21
4	Ballmouse	A B	572 570	1458 1287	449 475
5	Hall0	A B	156 248	478 441	99 143

Table 1 show the comparison of the number of key points found using SIFT, SURF, and FAST feature detection techniques. These techniques are used on both the common stereo photos seen on the Middlebury website as well as photographs taken with a digital camera. When compared to other approaches, the SIFT feature detector detects a greater number of features.

Figure 5: Input Left Imageand Input Right Image

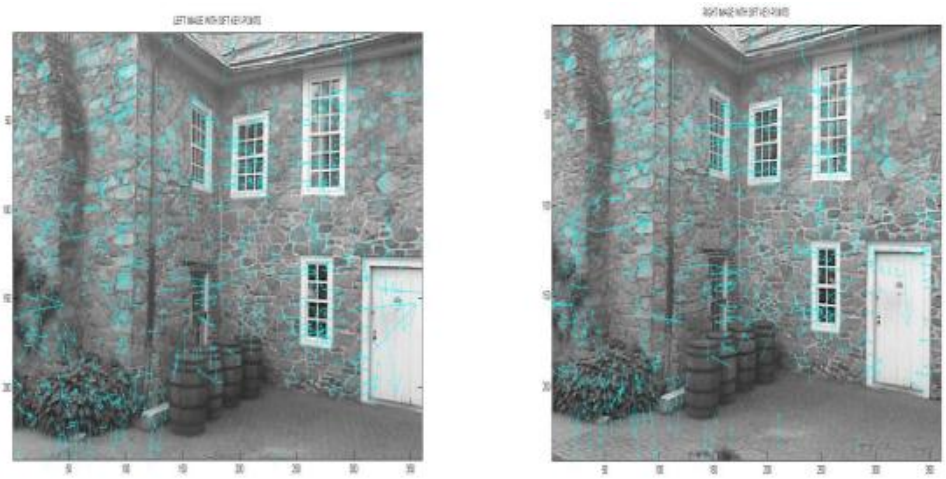

Figure 6 : Left Image with SIFT Key points, Right Image with SIFT Keypoints

Figures 5 and 6 show the input stereo left and right pictures and, correspondingly, Figures 6 shows the key points found in the left and right images.

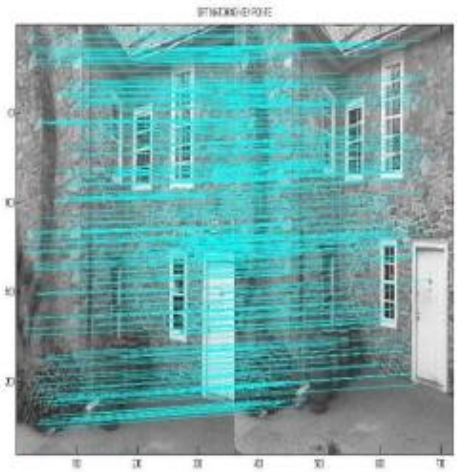

Figure 7 : SIFT- Key-Points Matching

The 3 x 3 fundamental matrix returned in the command window is shown in Figure 7 after the key points matched the photos.

Figure 8 : Left Rectified Image, Right Rectified Image

Along with the dataset, the calibration values for the middle bury website's standard photos are supplied. The camera calibration toolbox is used to determine the calibration values for actual photographs. The left and right corrected pictures are shown in Figures 8.

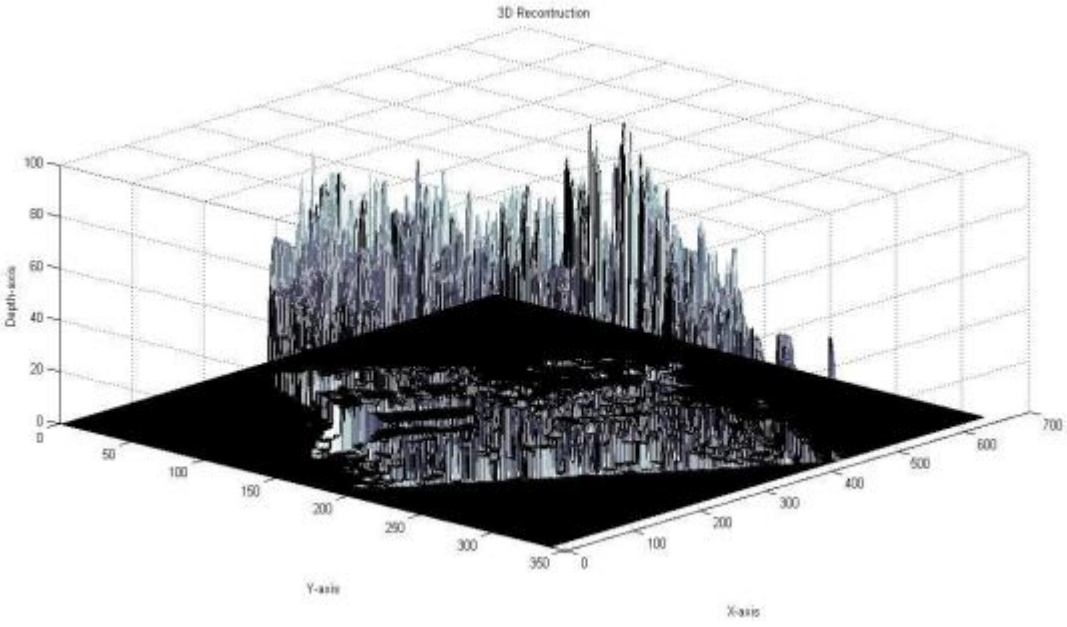

Figure 9 : 3D Reconstruction

The disparity map picture was created using a stereo matching technique based on region-based stereo matching and global energy reduction.

5. CONCLUSION

Thus, the three-dimensional structure of an item or scene was recreated utilising the suggested passive approach of 3D reconstruction employing stereo vision. The three-dimensional data was acquired using binocular vision, which simulates the human visual system. Using the SIFT feature identification

technique, the suggested approach identified a greater number of key points. In order to increase the accuracy of 3D reconstruction, camera calibration was done. Utilizing a region-based stereo matching approach based on global energy reduction, disparity was determined from the corrected pictures. Comparing the produced Disparity map's accuracy to ground truth values allowed for evaluation. Middlebury's database had the ground truth values disparity map for a few common stereo pictures. In a few places, the depth information could not be acquired.

6. REFERENCES

1. H. Herrmann and E. Pastorelli, Virtual reality visualization for photogrammetric 3D reconstructions of cultural heritage, Springer International Publishing, Berlin, 2014.
2. J. Esparza, M. Helmle, and B. Jähne, "Wide base stereo with fisheye optics: a robust approach for 3D reconstruction in driving assistance," in German Conference on Pattern Recognition, pp. 342–353, Springer, Cham, 2014.
3. C. Y. Yu and Y. Y. Yin,"RRSH: an architecture for rapid reconstruction of service composition in Smart Home," Applied Mechanics and Materials, vol. 719, pp. 661–664, 2015.
4. J. Durgin, D. Zhang, M. Weldon, and N. Gupta, "SU-E-T-559: quality assurance utilizing 3D dose reconstruction for stereotactic lung body radiotherapy[J]," Medical Physics, vol. 38, no. 6Part19, pp. 3618–3618, 2011.
5. Kordelas, Georgios & Diego Perez-Moneo Agapito, Juan & Vegas, Jesús & Daras, Petros., 2019, „State-of-the-art Algorithms for Complete 3D Model Reconstruction". Summer School ENGAGE-Immersive & Engaging Interaction with VH on Internet.
6. Fabio Remondino, Sabry El-Hakim, „Image-Based 3D Modelling: A Review", The Photogrammetric Record 21(115): 269–291
7. David G.Lowe, 2004, „Distinctive Image Features from Scale-Invariant key points", International Journal of Computer Vision 60(2), Kluwer Academic Publishers, Netherlands.
8. Lin Yang, Longyu zhang,Haiwei Dong, Abdulhameed Alelaiwi, Abdulmotaleb El Saddik, 2015, „Evaluating and improving the depth accuracy of Kinect for Windows v2," Sensors Journal, IEEE, vol. 15, no. 8, pp. 4275-4285.
9. Fredrick Hollsten, 2013, „The Effect of Image quality on the Reconstruction of 3D geometry from Photographs", MS Thesis, Aaltoo university, pp: 49-52.
10. G. Yu and J. M. Morel, "ASIFT: an algorithm for fully affine invariant Comparison," Image Processing On Line, vol. 1, pp. 11–38, 2011.
11. E. Rosten and T. Drummond, "Machine learning for highspeed corner detection," in Conference on Computer Vision, pp. 430–443, Springer, Berlin, Heidelberg, 2006.
12. M. Calonder, V. Lepetit, C. Strecha, and P. Fua,"BRIEF: binary robust independent elementary features European Conference on Computer Vision, pp. 778–792, Springer, Berlin, Heidelberg, 2010.
13. E. Rublee, V. Rabaud, K. Konolige, G. Bradski, and ORB: An efficient alternative to SIFT or SURF, "International conference on computer vision," Computer vision and image understanding, vol. 81, no. 1, pp. 2564–2571, 2001.
14. X. J. Yan, C. X. Zhao, and X. Yuan, "A robust local feature description method based on image contrast," Journal of Electronics and Information Technology, vol. 36, no. 4, pp. 882–887, 2014.
15. J. B. Li, X. Xiao, and X. P. Li, "A new feature descriptor combined with the division of annular regions," Journal of Xidian University, vol. 43, no. 2, pp. 64–69, 2016.
16. N. Sai-Nan, Z. Ming, S. Hong-Hai, and X. Fang,"Realization of improved Sobel adaptive edge detection algorithm based on FPGA," Chinese Journal of Liquid Crystals & Displays, vol. 29, no. 3, pp. 395–402, 2014.
17. L. Tiegang, Z. Ya, and L. Shizhong, "RANSAC exact matching method combined with improved particle swarm algorithm," Mechanics and Electronics, vol. 35, no. 7, pp. 20–24, 2017.
18. P. Speciale, D. Pani Paudel, M. R. Oswald, T. Kroeger, L. Van Gool, and M. Pollefeys, "Consensus maximization with linear matrix inequality constraints," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017.
19. P. Xiao, Z. Wang, and S. Rajasekaran, "Novel speedup techniques for parallel singular value decomposition," in IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, Exeter, UK, 2018.
20. P. T. Zhang, Research on feature point matching algorithm in 3D reconstruction based on portable camera images, Zhejiang University of Technology, Hangzhou, 2018.