# A Study Of Structure From Motion Photogrammetry For Generating 3D Model From 2D Images

Yug Shah, Suyog Raut, Sagar Wadle, Smita Patil
*Student, BE-Information Technology, Atharva College of Engineering, Mumbai, India*
*Assistant Professor, Information Technology, Atharva College of Engineering, Mumbai, India*

***Abstract:*** *With the advancement in technology and the rapidly increasing usage of internet applications for online shopping, there is an increase in the need and expectation of the online consumers.The online commerce that are used now for shopping mostly provide a 2D view of the products . Therefore buyers have only partial view of the product they intend to purchase. Therefore We aim to develop a 3D grocery review application that would provide a 3D object view of the products online using reconstruction.3D modeling is a process of mapping an object into the real-world coordinates in 3 dimensions representing the height, width and also the depth.. Traditional and recent technology for building 3D models from images include photogrammetry, structure from motion and other 3D scanning approaches such as laser, lidar, etc have been discussed in this study. Photogrammetry is a process of extracting features from images that measurement of an image(distance,height etc) and then structure from motion , to build a 3D model from those images.Later,using three.js generated model will be displayed .THREE.JS is a WEB API for 3d modelling.*

***Keywords:*** *3D modelling, photogrammetry, structure from motion, 3D from 2D images, 3D reconstruction*

## I Introduction

The design and reconstruction of 3D models has now become a significant part of computer graphics and of computer vision applications and also in areas such as architecture, visual and special effects in entertainment, augmented and virtual environments, games, engineering and education. 3D reconstruction can be achieved by two approaches called as active and passive[1,2,3]. 3D Max and Blender are the examples of the traditional modeling system that uses 3D meshes for the construction of 3D models. However, reconstruction from these 3D meshes was challenging and therefore, resulted in need for a better approach. A major progress has been seen in the last decade to resolve this issue. Photogrammetry, Structure from motion(SFM) and Image-Based Modelling(IBM) are the widely used modern approaches for building 3D models. There are few alternative technologies such as laser scanner, lidar, structured light, camera calibration, etc. used for the same purpose. Every technology and approach has its own significance and performance factors and can be preferred based on the requirements and the resources.

Photogrammetry, the term is a combination of ‘Photo’ which means light, ‘gram’ that means drawing and ‘metry’ that means measurements[4]. So, it is the technology of extracting information from images. It measures and processes features such as length, depth, angular position, etc in images(photographs). This technology has been successfully employed in a wide range of industries. The main application of photogrammetry is to generate 3D models out of the images taken from an object, drawings, measurements and also used for many different purposes such as the maps that we use are also created using photogrammetry. Processing complicated shapes of small objects, uneven surfaces, reflective surfaces and objects, extracting features of transparent object surfaces are the challenges required to be improved[3]. This technology is further then classified on the basis of the location of the camera as Aerial Photogrammetry and Terrestrial or Close Range Photogrammetry. The camera is located in an aircraft in Aerial photogrammetry and is typically used for creating topographic maps, for surveying the terrain-surfaces, soil, land,etc, whereas in Terrestrial type, the camera is usually on ground and can be hand-held or mounted on any support. The Close Range Photogrammetry is the technology that is preferred for creating 3D models mainly[5].

Structure from motion(SFM) is the science of creating 3D model and it uses the same approach as in photogrammetry, but it creates the structure from the movement of the camera or the object, i.e. translation or rotation[4]. The image sequence in this approach, is taken either by moving the camera to different angles or by moving the object itself. This is achieved by bundle adjustment[5]. The extracted features from 2D images that include pixel positions, edges and corners of objects, lines and curves across the corners and edges, angular positions of object with respect to 3D coordinates, light intensity and other measurements such as length, depth, resolution is given as input to the SFM algorithm.

The Image Based Modelling is also a type of photogrammetry in which cameras are used to design, model and measure buildings, archaeological artifacts, structures, film sets, etc. Laser scanners uses lasers for scanning the object known as 3D scanning. Lidar is a term combined from Light and radar, or an acronym for

Laser Imaging Detection And Ranging. Lidar is based on radar and it creates continuous points between an abject and the laser by using the lasers. Laser are considerably more costly and required a desired system for it .yet it provides high accuracy i.e 90-95% but for inexpensive grocery product IBM may be not most usable.

The sfm photogrammetry starts with processing of images and removing redundancy from overlapping images. The feature extraction involves edge and corner detection of the object from the images. Harris and canny algorithm has been proposed for edge detection with different approaches. Then comes the feature matching process that is carried out by SIFT[8] and BRISK[9]. The image features are quantized using K-Means algorithm[7] and Brute-Force algorithm to compare features of images. The RANSAC[10, 11] is used for the estimation of the fundamental or projection matrix between two images and validation of the putative matches from the images(closest descriptor of feature from two images).

To determine the correlation between the world and image coordinates, camera calibration is the method used[12]. The next step is to minimize the reprojection errors and perform techniques to recover the structure. Bundle Adjustment (BA)[6, 13] and Direct linear transformation(DLT) is that step which performs refinement of all elements after processing of images and matching the features and different approaches. Soulaiman El hazzat[1] have proposed a method for 3D object reconstruction using structure from motion approach, by global and local bundle adjustment.

In this paper, Section II gives the idea about the technology used and a brief introduction about the terms and intermediary processes involved in structure from motion. Section III shows the application areas of SFM and Section IV holds the proposed work for improving the user's experience in viewing the online products by providing a complete 3D view of product, and further the complete process of 3D modelling has been explained.

## II Structure From Motion

The purpose of designing this tool is to prepare a 3D model of an object using series of 2D images taken from various angle. 3D reconstruction or 3D model of an object can be created using various technology like scanner, laser, etc., but photogrammetry is less expensive of all. Photogrammetry uses different images of an object taken from different angles(possibly all angles) and generate a 3d model of that object. This entire process requires a digital camera for capturing images, and the presented tool as a processor and would result in finally generating a 3D model of the object. The tool comprises of techniques as edge detection, epipolar geometry, feature matching, 3D Camera position and coordinates estimation and, sparse and dense reconstruction.

The presented tool performs 3D reconstruction from images that includes the following phases:

### A. Edge and Corner Detection

Edge detection is an image processing technique that identifies all local content like corners, edges, features etc of objects within images using Canny or Harris or sobel algorithm[7,14].

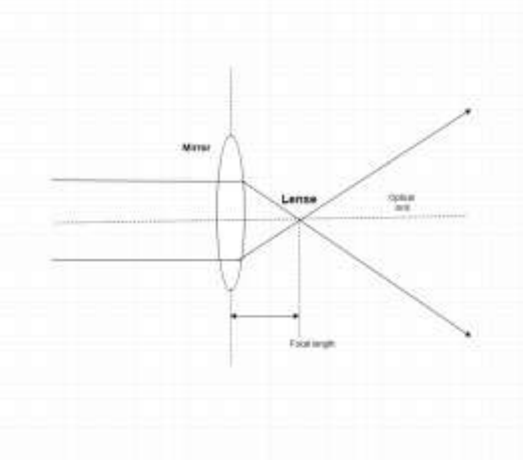

**Fig. 1 Focal length**

To perform edge detection it is necessary to know focal length and sensor width of image. Focal length is the distance between the lense and mirror of projection. sensor width is the size i.e width of the sensor.In canny edge detection we smoothen the image to reduce the noise in the image.after reducing the noise we calculate the gradient of all pixel in X and Y direction [Gx & Gy], where we calculate edge gradient as follow :

$$(G)=\sqrt{\boldsymbol{G_a^2}+\boldsymbol{G_b^2}} \qquad \text{eqn.1}$$

where $\boldsymbol{G_a}$is for horizontal direction
$\boldsymbol{G_b}$is for vertical direction

At the other side harris corner detection uses mathematical operator that search features in images. Features are some matching points between images. for example, When performing image matching, we need a few corresponding points between the two views(i.e left and right ). Once we found those, we can triangulate almost all points on the image.while sobel forms a 3x3 matrix from pixels presented in image.

The below table Table 1. gives analysis of edge detector algorithms. It shows that Canny is more complex in terms of computation compared to Sobel and Harris. Sobel is more efficient as Canny but has higher rate of fall edges. Harris algorithm does not keep removing outliers before finding edges whereas Canny keeps removing outliers, thus Canny has more time and space complexity.[15]

**Table 1. Analysis of Edge Detector Algorithms**

| S No. | Algorithm for Edge Detection | Pixel Comparison | Photo Smoothening | Time complexity | Space Complexity | Noise Sensitivity | False Edges |
|---|---|---|---|---|---|---|---|
| 1. | Sobel | No | No | Lower | High | Less | More |
| 2. | Canny | Yes | Yes | High | High | Least | Least |
| 3. | Harris | Yes | No | Lower | Lower | Least | Least |

**B. Camera Coordinates Estimation**

Camera position of an image is obtained by comparing the position with the center image i.e. all motion of the camera is considered and change in angle is saved. To generate a matrix it is necessary to have camera position(coordinates). Parameters such as rotation, motion, angle and translation of camera is in a matrix form estimated using SIFT[13].

SIFT algorithm has the ability to find strong matching points and it process the matching problem with the help of translation, rotation from different images.its is more stable and capture strong point at certain extent for images which are taken from various directions Pair of images with more matching feature and corner are then used to get 3D coordinates for such image. Camera position and image edge matching will generate the projection matrix. The projection matrix is the 3x3 matrix of 3D coordinates that is used to display the 3d object. Essential matrix is motion or translation between two images. This all information will be later used to form a structure from motion.

There are various co-ordinates from taken images,compared using Rotation and Translation.

$$Z`=RZ+T \qquad \text{eqn.2}$$

where Z'=image co-ordinates of image1
Z=image co-ordinates of image2
R=rotation matrix
T=translation matrix

and the resulted essential matrix will be
E=[T]xR.[13]

**C. Epipolar Geometry**

It finds the putative match between two images, i.e. the closest feature match from the other image while comparing a pair of images, and finds the geometry of these putative matches. It uses 5 point relative pose or 8 point linear method. It generates matrixes such as fundamental and essential matrix using this matrix 3d position (coordinate) i.e. X,Y and Z axis of two images are found i.e. projection matrix, which is also called as multi-view geometry. [12]

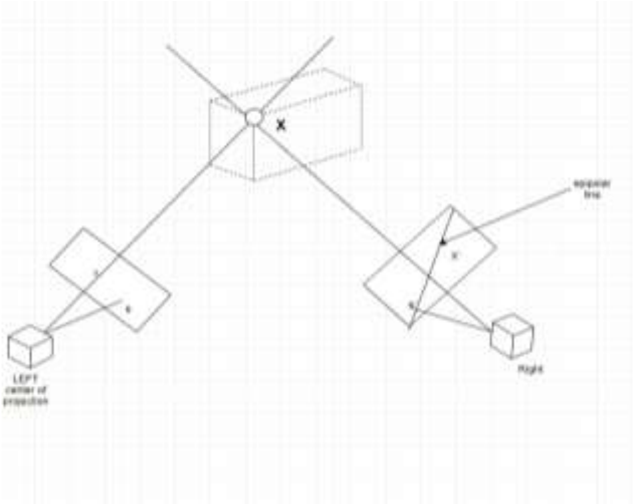

**Fig. 2 Epipolar Geometry**

Mostly to perform epipolar geometry brute Force algorithm is used.As all the point has to be traversed (checked). Different methods (random check, bottom up) but that may be efficient for some Condition while take more time for some Condition. By comparing pixels and co-ordinate an matrix is formed that states a pixel to be mapped for 3d modelling

**D. Dense Reconstruction**

Dense reconstruction is mapping the 3D coordinates on a 3D point cloud. 3D point cloud is a cloud like structure with coordinates mainly using to map coordinates that can be used to generate a 3d model. Reconstruction is mainly performed using bundler[6]. Bundler is structure from motion tool that perform reconstruction by take several image as input and there local content data i.e. essential and projection matrix, triangulation ray and 5 or 8 point output. The next step is to compare such matrices and data, and map them on point cloud which is also termed as sparse cloud using bundler[19]. Triangulation ray is forming a triangular structure around 5 point so that surface can be formed using color and Pixel detail of point chosen. Direct linear transformation (DLT) compares image coordinate with variables and form vector . DLT algorithm is comparatively more detailed than bundle yet bundler is more popular than DLT. 3d sparse point cloud is an environment on which matrix that were created while epipolar geometry and camera coordinate estimation is mapped. It is a 360 degree viewable environment mainly used for 3d operation i.e augmented reality, virtual reality, modelling etc.

**Table 2. Analysis of Techniques to find coordinates for dense reconstruction**

| Techniques | Time Complexity | Space Complexity | Accuracy | | Triangulation |
|---|---|---|---|---|---|
| | | | Low Quality | High Quality | |
| Bundler | Moderate | High | High | High | Yes |
| DLT | HIgh | High | High | Low | No |

**E. 3D Modelling**

The data generated on sparse cloud will be used to generate a 3d model that can be used on on web browser for presenting the 3D view of the object. To present or display three.js is required.while generating 3d model is only be possible on windows platform.

Three.js is a api that is generated to display 3d dimension data,object , environment etc on web platform that were not possible few years before.

## III Conclusion

Mainly 3D reconstruction is used for artifacts and sculpture. Laser and Scanner gives very minute and detailed result (3d model) .Accuracy of laser and scanner is more than that of photogrammetry due to it cost and system photogrammetry is more efficient for 3d modelling of grocery product.

In photogrammetry ,Harris is used for edge detection. Those coordinates are processed using SIFT algorithm which is most efficient for generating fundamental,relational and projection matrix. Brute force is used for epipolar geometry as it gives an average time notation for all conditions later by performing triangulation and surface reconstruction using poisson surface reconstruction an 3d model get generated . Generated 3d model is displayed using three.js i.e visible on web browser.

## References

**Examples follow**:

[1]. Soulaiman El hazzat, Abderrahim Saaidi, Khalid Satori, "Structure from Motion for 3D Object Reconstruction Based on Local and Global Bundle Adjustment", IEEE(2015).

[2]. Luo Jianxin, Qiu Hangping, Wu Bo," Survey of Structure from Motion", International Conference on Cloud Computing and Internet of Things (CCTOT 2014).

[3]. Human Esmaeili, Harold Thwaites. "Virtual Photogrammetry", Centre for Research-Creation in Digital Media (CRCDM).

[4]. Natan Micheletti, Jim H Chandler, Stuart N Lane, "Structure from Motion(SfM) Photogrammetry", Geomorphological Techniques, Chap. 2, Sec. 2.2, British Society for Geomorphology(2015)

[5]. Agarwal,Snavely,Simon et al.Building Rome in a Day[C]. The Twelfth IEEE International Conference on Computer Vision,Kyoto,2009.

[6]. Bill Triggs, Philip McLauchlan, Richard Hartley et al. Bundle Adjustment-A Modern Synthesis[J]. Vision Algorithms: Theory and Practice,2000,13(5):47-71.

[7]. Harris C, Stephens M. A combined corner and edge Detector[C].Proceedings of the Fourth Alvey Vision Conference. [S.I.]: [s.n.], 1988: 147-151.

[8]. Liu Ran, Zhang Hua, Liu Manlu, Xia Xianfeng, Hu Tianlian "Stereo Cameras Self-calibration Based on SIFT "Robotics Laboratory, School of Information Engineering, Southwest University of Science and Technology ,2009

[9]. Leutenegger S, Chli M, Siegwart R. BRISK: Binary Robust Invariant Scalable Keypoints[C].Proceedings of the 13th European Conference on Computer Vision.Spain, 2011 :2548-2555.

[10]. J.-M. Frahm and M. Pollefeys. RANSAC for (quasi-) degenerate data (QDEGSAC). CVPR, 2006.

[11]. M. A. Fischler and R. C. Bolles. Random sample consensus: Aparadigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381–395, 1981.

[12]. Richard Hartley, Andrew Zisserman.Multiple View Geometry in Computer Vision [M]. 2002

[13]. Agarwal, S., Snavely, N., Seitz, S.M., Szeliski, R.: Bundle adjustment in the large. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6312, pp. 29–42. Springer, Heidelberg (2010).

[14]. https://docs.opencv.org/3.3.1/da/d22/tutorial_py_canny.html

[15]. Lowe D. Object recognition from local scale-invariant features[C]. Proceedings of the 7th IEEE International Conference on Computer Vision. Greece, 1999: 1150-1157.

[16]. S.K. Katiyar, P. V. Arun, "Comparative analysis of common edge detection techniques in context of object extraction", IEEE TGRS Vol.50 no.11b. pg no:68-79

[17]. Richard Szeliski.Computer Vision:Algorithms and Application [MI/OL], 2010. http://szeliski.orgiBooki.

[18]. Liu Wei, Wu Yihong, Hu Zhanyi. A Survey of 2D to 3D Conversion Technology for Film [J]. Journal of Computer-Aided Design & Computer Graphics, 2012, 24(1): 14-28.

[19]. Computer Vision, Imaging and Computer Graphics. Theory and Applications: International Joint Conference, VISIGRAPP 2011, Vilamoura, Portugal, March 5-7, 2011. Revised Selected Papers (pp.86-101)