# Object Pose Estimation using Least Non Coplanar Feature Points

**Rohit Prasad, Tejaswini Kar**

*Abstract— This paper describes the pose estimation of an object using a calibrated camera. The idea is to first calibrate the camera and then implement the algorithm to find the estimated matrices that describes the three dimensional pose of the object. The camera calibration process includes capturing the images and then processing them to find the intrinsic and extrinsic parameters, which are used to estimate the object pose. And object pose estimation is carried out by first finding the corners with the help of Harris feature extraction and then comparing the image and object matrices in the POSIT algorithm and finally eliminating the errors with the help of iterations. The algorithm estimates the pose with a minimum of four non-coplanar points from the acquired image. Both camera calibration and pose estimation processes were implemented using MATLAB® Ver.7.12.0.635 (R2011a).*

*Index Terms— POSIT, pose estimation, camera calibration, intrinsic parameters, non-coplanar feature points.*

## I. INTRODUCTION

THIS Mathematical calculation of object pose i.e. the position and orientation of an object using image's feature points, when a model or the object's geometric configuration is known. Importance of the above involves applications, such as object recognition, cartography, tracking and calibration.

In 1970s, Ivan. E. Sutherland [1] addressed a data input system which easily and conveniently accepts 3-D data from multiple 2-D views of the object that was to be digitized. In 1980s, Joseph S.-C. Yuan [2] presented for the first time a completely general solution to the exterior orientation calibration problem, which was independent of the number or distribution of the feature points selected from the target. In 1990s, a new algorithm was suggested by M. A. Abidi and T. Chandra [3] for pose estimation which featured the volume measurement of tetrahedral which included the target points and the lens center of camera or vision system, using a quadrangular target and a pinhole camera. The main contribution of this work is the discovery of efficient and accurate ways to recover interior orientation parameters using a direct and robust solution.

Noteworthy techniques can be found in the works of Tomasi [4], who introduced an inventive approach called factorization method, by application of which one can robustly recover shape and motion from a sequence of images without assuming a moving model, i.e. constant rotation and translation, and Ullman and Basri [5], in which a new approach is suggested that each model is represented by the linear combination of 2-D images of the object, which are the base of the technique used in this paper. In the recent years, some major improvements are done. In 2003, A. Ansar and K. Daniilidis [16], presented number of simulations for pose estimation involving closed-form or linear solutions free of initialization and compared the reults with other linear algorithms and iterative approaches. In 2006, O. Tahri, C. Leroux and J. M. Alexandre [17], proposed a new efficient method in which partial pose was determined from only five points. In 2011, C. Chen, H. Ming and R. Luo [18], proposed a new framework that requires still images to estimate the side view pose of human and using that walking pose was estimated. In 2014, G. Bae, S. Kwak, H. Byun and D. Park [19], efficiently introduced a method to detect humans in surveillance video, we have a keen interest in this method because it does not require any knowledge of camera parameters and it improved the efficiency and accuracy of real- world surveillance videos.

Before we move to pose estimation part let us discuss the need for camera calibration [7], [9], [11], [12], [13]. Camera calibration is required because capturing process reduces the dimensions of data from three to two and camera calibration determines which light source is associated with each pixel on resulting image. Camera calibration is also required due to the reason that there may be misaligned lenses or deformation in their structure during manufacturing, resulting complex distortions in the acquired image. A camera consists of a lens and an image plane, camera calibration is done to improve the transformation between image space and object space. Due to the presence of distortions between the object points and their respective points on the image, a perfect transformation cannot be achieved by a perspective transformation. An approximate theoretical description of the real relationship between these distortions can be made in order to understand or explain how distortions occur. And camera calibration fulfills this void. Also, the proposed method requires camera calibration's only one intrinsic parameter [10]. The method used for pose estimation in this paper depends on techniques of linear algebra and like methods of Lowe and Yuan [14] it uses iteration and unlike classic Newton's method [14] an initial pose estimate is not required, and also matrix inversion in its iteration loop is not included. The proposed method requires fewer computations than previous works mentioned. Also, stability of this algorithm appears to be valid and it smoothly degrades as increment in the noise is made. This method is known to us as POSIT (Pose from Orthography and Scaling with Iteration) [6], [14].

Previously, POSIT was associated with the uncelebrated camera but in this paper we put a light on using POSIT with a calibrated camera (intrinsic parameters only). This paper is organized as follows: Methodology of camera calibration and pose estimation using POSIT is described in Section II. Results and discussions is given in Section III. Finally, Section IV discusses the entire work in general and gives conclusion and future extensions.

## II. METHODOLOGY

The work in this paper is essentially the estimation of pose of the object using POSIT. But before we begin, we do a procedure for camera calibration using 'Camera Calibration Toolbox for Matlab®' by Jean-Yves Bouguet [8]. The images are being processed by a general purpose PC and since camera is analog, for connection between the two a frame grabber is required (shown on Figure 1). After connection, a set of 10 monochrome test images of a planar checkerboard grid differently oriented in each image (shown on Figure 2) was taken. The toolbox features an algorithm to compute a projective transformation between the image points of the n different images with the extracted corner points of the checkerboard pattern. Afterwards, using closed-form solution the camera intrinsic parameters [10] are recovered, also with a linear least- square solution, the recovery of the third and fifth order radial distortion terms are done. And finally, all the parameters are recovered by a nonlinear minimization of the reprojection error i.e. by using a Lavenberg-Marquadt method [8], [10], [13].
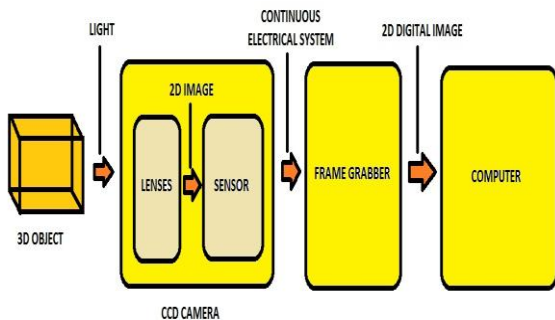


**Figure 1. Connection between camera and computer.**

From the above recovered parameters, our interest is only with the 'Focal Length', which is used in the POSIT for pose estimation. The 'Focal Length' thus obtained is a [2x1] vector, we then use the magnitude of the vector, for a particular range of distance (as in CCD camera, focal length changes with the zooming).
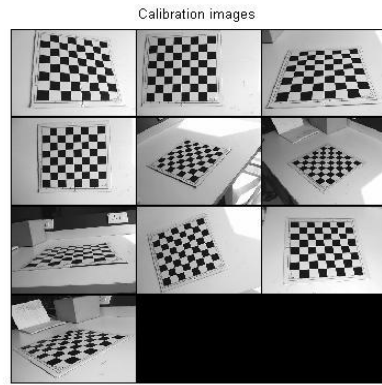


**Figure 2. Images used in calibration process.**

Now, the pose estimation phase begins. Two snapshot are taken within Matlab® and are converted into grayscale (shown in Figure 3). The first iteration step includes multiplying of two vectors i.e. a precomputed object matrix (obtained from object's feature extraction) and two vectors (which only depends on coordinates of extracted feature points). This result in two vectors which are then normalized, the obtained result is the rotation matrix's first two rows and by the normalization of these vectors results in the projection's scaling factor and the same is later used to find the translation matrix. With the help scaled orthographic projection (SOP) [6], [14], the involved image points are obtained. This part of algorithm is called as POS (Pose from Orthography and Scaling) [6], [14]. The next iteration step corrects the obtained pose to the actual pose, and for these shifted points a SOP is obtained. This part is called as POSIT (POS with ITeration) [14]. Usually four or five iteration steps can converge to an accurate pose.
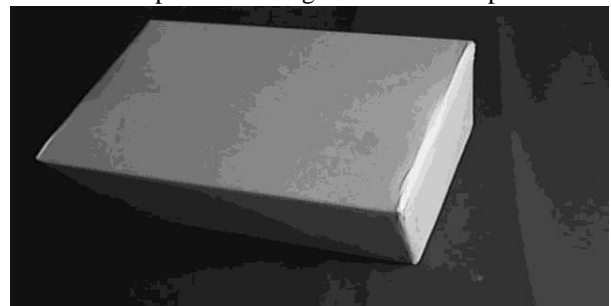


**(a)**
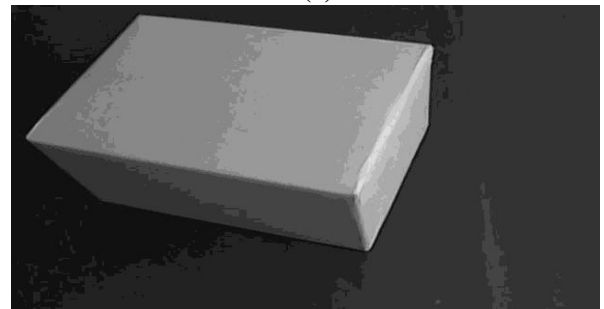


**(b)**

**Figure 3: Images converted into grayscale.**
**(a) First image taken from the camera.**
**(b) Second image taken at a different position.**

A more perceived explication of the POSIT algorithm is discussed. By perspective projection, the distribution of object's feature points and respective image's feature points is known. From a perspective image, SOP images of the object feature points is build and an exact pose is obtained by applying POS algorithm to these obtained images. It is beneficial to know the exact pose of the object while computing the SOP images. Almost exact depth for each feature point is obtained by applying POS to actual image's feature points. The feature points at these depths are then positioned on the line of sight. Doing this, SOP image is computed and again POS is applied to this obtained image to find an enhanced SOP image. These steps are repeated until the result is converged towards a precise SOP image and an exact pose. As stated, typically four or five iteration steps are required to converge to an accurate pose.In the end, an approximated pose of the object is obtained using the calculated rotation matrix and translation vector.

## III. RESULTS AND DISCUSSION

The camera used is a simple CCD camera with the resolution of [640 x 480] pixels. The software used is MATLAB® Ver.7.12.0.635 (R2011a) and the machine's processor specifications are Intel® Core™ i7-3610QM CPU @2.30GHz. After the camera calibration was performed, the following parameters were calculated: focal length, principal point, skew, distortion and pixel error, and they are the intrinsic parameters of the camera. The focal length in pixel is stored in a [2x1] vector. The principal point is also stored in a [2x1] vector. A scalar value of the angle between x and y pixel axes i.e. skew coefficient, is stored. And a [5x1] vector stores the radial and tangential distortions i.e. image distortion coefficients.

Calibration parameters after calculations:

Focal Length: [730.60139  726.85790] + [9.37023 8.65351]

Principal Point: [352.57795  269.74615] + [11.08022 12076128]

Skew: [0.00000] + [0.00000] => Angle of pixel axes = 90.00000 + 0.00000 degrees

Distortion: [0.09973 -0.55053 0.00551 0.00686 0.00000] + [0.05133 0.25032 0.00467 0.00472 0.00000]

Pixel error: [0.91590 0.91024]

From the above parameters our interest is only with focal length, i.e. [730.60139 726.85790] and after finding the magnitude we get: 1030.582747 (in pixels). We use this value in POSIT algorithm to find the pose of the object.

First, we took two snapshots of the object with black background and converted them into grayscale (shown in Figure 3). Then, the strongest eight corner points are extracted using Harris feature extraction (shown in Figure 4). These feature points are the image points (in 2D) and we provide the object points (in 3D) as:

Object Points: [0,0,0;  17.8,0,0;  17.8,23.2,0;  0,23.2,0; 0,0,13.5;  17.8,0,13.5;  17.8,23.2,13.5;  0,23.2,13.5]
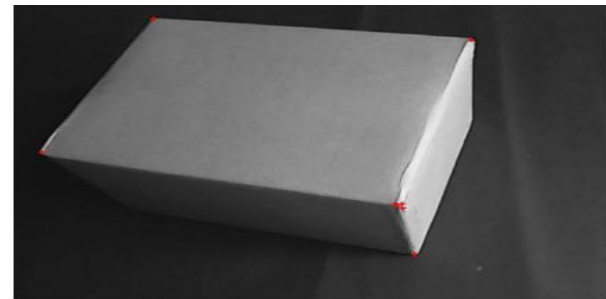


**(a)**
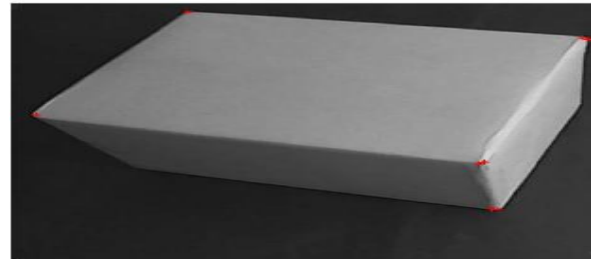


**(b)**

**Figure 4:  Feature Points marked on the objects.**
**(a) Point features extracted and marked on the object in the first image**
**(b) Point features extracted and marked on the object in the second image.**

Now the code is fed with the required inputs. The code runs and gives the approximated pose of the object in 6 iterations (each). Finally, the rotation matrix and translation vector is obtained and plotted with reference point as [0, 0] (shown in Figure 5):
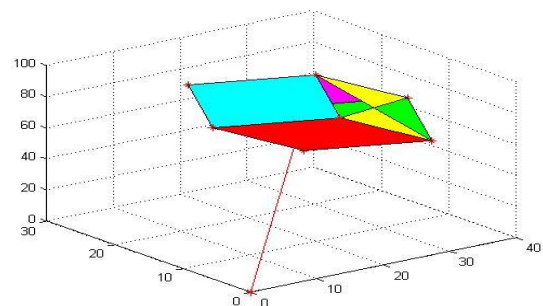


**(a)**
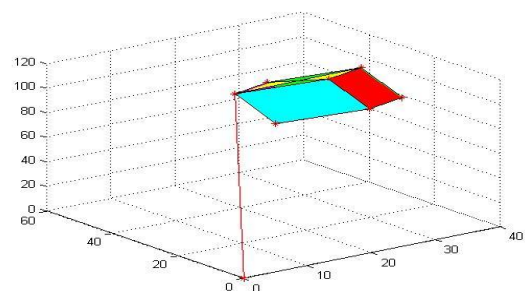


**(b)**

**Figure 5:  Estimated object pose [The X, Y, Z axes are shown in centimeters].**
**(a) Estimated pose of the object in first image.**
**(b) Estimated pose of the object in the second image.**

[Note: The obtained pose is completely depending upon the first feature point extracted.]

## IV. CONCLUSION AND FUTURE EXTENSIONS

This paper explores the process of pose estimation using the method of POSIT. Since the use of cameras are increasing in scientific applications as well as in the applications requiring precise visual information, POSIT has advantages over popular pose algorithms as an initial pose estimate is not required, the implementation of the method is simple and can perform much faster than existing algorithms. This was also noted that with different images total run time of the POSIT algorithm was different. In the first image, the iteration count was 6 and total run time of the POSIT algorithm was 1.915643 seconds but this was not same with the second image, as in second image there is reduction in noise and slight change in angle of projection, with 6 iterations, the run time was 10 times lesser than the previous image i.e. 0.186915 seconds.

Here, we have not used all of the calibration parameters and left for future extensions. Here, we have moved both camera and object while performing the test. Also, this method can be implemented for the development of drones, automated motor vehicles like Mars Exploration Rover and satellites. This method can also be implemented in different fields like medical, geology, archeology, etc. We can take a case like in archaeology, where a 3D model can be reconstructed using the broken parts of the sculptures whose some parts already been destroyed or lost. In geology, one can be motivated by the 'Virtual Los Angeles Project' [15] and implement this method for robust results.

Proposed method can only be used with 3D objects and cannot be used with planar objects. Although, it does not require initial guess but a set of 3D object points is required at the beginning. These are some requirements for this method to work properly and give useful results.

## ACKNOWLEDGMENT

## REFERENCES

1. Ivan E. Sutherland: "Three-Dimensional Data Input by Tablet", Proceedings of the IEEE, Vol. 62, No. 4, April 1974.
2. Joseph S. –C. Yuan: "A General Photogrammetric Method for Determining Object Position and Orientation", IEEE Transactions on Robotics and Automation, Vol. 5, No. 2, April 1989.
3. M.A. Abidi and T. Chandra: "A New Efficient and Direct Solution for Pose Estimation Using Quadrangular Targets: Algorithm and Evaluation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 5, May 1995.
4. Takeo Kanade and Carlo Tomasi: "Shape and Motion from Image Streams: A Factorization Method Full Report on the Orthographic Case", Cornell TR 92-1270 and Carnegie Mellon CMU-CS-92-104, March 1992.
5. Ronen Basri and Shimon Ullman: "Recognition by linear combinations of models", IEEE Transactions on pattern analysis and machine intelligence, Vol. 13, No.10, October 1991.
6. Daniel DeMenthon and Larry S. Davis: "New exact and approximate solutions of the three-point perspective problem", IEEE, 1990.
7. T. A. Clarke and J. G. Fryer: "The development of camera calibration methods and models", Photogrammetric Record, 16(91): 51–66 (April 1998).
8. Jean- Yves Bouguet: "Complete Camera Calibration Toolbox for Matlab®", Computer Vision Research Group, Dept. of Electrical Engineering, California Institute of Technology.
9. Roger Y. Tsai: "A Versatile Camera Calibration Techniaue for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, August 1987.
10. David A. Forsyth and Jean Ponce: "Computer Vision: A Modern Approach, Second Edition", Prentice Hall, ISBN-13: 978-0-13-608592-8.
11. Roger Y. Tsai: "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision", IEEE, 1986.
12. Guo- Qing Wei and Song De Ma: "A Complete Two-plane Camera Calibration Method and Experimental Comparisons*", IEEE, 1993.
13. Azra Fetić, Dinko Osmanković and Davor Jurić: "The procedure of a camera calibration using Camera Calibration Toolbox for MATLAB", MIPRO 2012, May 21-25, 2012, Opatija, Croatia.
14. Daniel F. DeMenthon and Larry S. Davis: "Model- Based Object Pose in 25 Lines of Code", Computer Vision Laboratory, Center for Automation Research, University of Maryland.
15. Urban Simulation Team: "Virtual Los Angeles Project", University of California, Los Angeles.
16. Adnan Ansar and Kostas Daniilidis: "Linear Pose Estimation from Points or Lines", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 5, May 2003.
17. Omar Tahri, CJean Marc Alexandre and Cristophe Leroux: "Pose Estimation From less than Sic Non Coplanar Points", Proceedings of the 2006 IEEE International Conference on Robotics and Automation Orlando, Florida - May 2006.
18. Cong Chen, Ronghua Luo and Huaqing Ming: "Side View Pose Estimation Of Human From Images Using Prior Knowledge", 2011 4th International Congress on Image and Signal Processing.
19. Guntae Bae, Sooyeong Kwak, Hyeran Byun and Daeyong Park: " Method to improve efficiency of human detection using scalemap", ELECTRONICS LETTERS 13th February 2014 Vol. 50 No. 4 pp. 265–267

## AUTHORS PROFILE

**Rohit Prasad,** has received his Bachelor of Technology in Electronics and Electrical Engineering in 2014, from Kalinga Institute of Industrial Technology, Bhubaneswar, India. His research interests include Digital Image Processing, Digital Signal Processing and Embedded Systems. He is currently working on project, titled: 'Wireless network based PLC automation system for moving Stacker cum Reclaimers (SCRs) for Raw material handling Plant (RMHP) of Durgapur Steel Plant (DSP), SAIL', which is directly supervised by Government of India, at Research & Development Centre for Iron & Steel of SAIL, Ranchi.

**Tejaswini Kar,** received her Bachelor's degree in Electronics and Telecommunication from BPUT in 2003 and M.Tech. in Communication System Engineering from KIIT University, Bhubaneswar in 2008. Currently she is a PhD scholar in KIIT University. She has been working as a faculty in School of Electronics Engineering, KIIT University from 2008.She currently teaches Signals and Networks & Digital Image processing courses at School of Electronics Engineering at KIIT University. Her main research interests include Image and Video Processing.