Scientific Journal of Impact Factor (SJIF): 4.72
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 10, October -2017

3D Trajectory Reconstruction of a Moving Object from a Stereo Video using Particle Filter

Dr. Rachna Verma[1]

[1]*Department of Computer Science and Engineering, Faculty of Engineering, JNV University, Jodhpur, Rajasthan, India*

Abstract: *Moving object detection and tracking in videos is an actively researched area for the last two decades due to its practical applications in many areas, such as trajectory analysis of moving objects, making game playing robots, human computer interaction, etc. This paper presents an experimental study to reconstruct the 3D trajectory of a coloured moving object by combing particle filter and stereo vision. Although, the initial results obtained are encouraging, but the generated trajectory has waviness. Further research is required to reduce the waviness and generate more accurate trajectory.*

Keywords: *Object detection, object tracking, particle filter, 3D trajectory, Kalman filter, optical flow*

I. INTRODUCTION

Detecting and tracking objects in videos and reconstructing their 3D trajectories are an active research area of computer vision. These technologies have many potential applications, such as analysis of sport videos to detect faulty motions of a sportsman, trace the trajectory of a ball, etc[1], understanding human postures to improve human machine interactions, prediction trajectories of moving objects to detect and avoid collisions, catching moving objects by robots, surveillance, etc. The reconstruction of the 3D trajectory of a moving object is impossible from a monocular video alone without making prior assumptions about the motion of the object [2]. Biological and artificial systems use binocular stereoscopy to calculate the 3D co-ordinates of a scene point by utilizing the disparity in the projections of the scene point in the two views captured by a pair of two cameras positioned side by side at a distance. However, the calculation of disparities of all the points of a scene is a time consuming process and no real time solution is yet available for commonly available hardware. To efficiently reconstruct the 3D trajectory of a moving object, the combination of the stereo vision concept and moving object detection techniques, such as optical flow, background subtraction, particle filter, Kalman filter, etc, can be combined. The combination reduces the correspondence matching to only objects of interest, making the process more efficient. In this paper, the particle filter object tracking concept is combined with the stereoscopy concept to efficiently calculate the 3D trajectory of a moving object from a stereo video. The remaining part of the paper is organized as follows: section 2 reviews the related work. It introduces the basic concepts of stereoscopy and gives an overview of the particle filter object tracking concept. Section 3 describes the proposed system. Section 4 presents some experimental results and finally, section 5 concludes the paper.

II. RELATED WORK

Vision based object tracking has been actively researched for the past three decades. However, most of the research is limited to monocular videos, which alone is insufficient to create the 3D trajectories of the tracked objects in a scene without relying on assumptions about the scene that are too strong for any practical application [3]. To overcome the inherent limitations of monocular vision to recreate 3D trajectory, many researchers proposed to use stereoscopy. Harville [4] and Zhao et al [5] proposed to use dense stereo and static background model to recreate 3D trajectories of moving objects. Since they use dense stereo vision, the process is computationally inefficient to generate the trajectory in real time and their assumption of static background model also limits the domain of the object tracking.

Mittal and Davis [6] proposed to use region based stereo vision to track moving object. The method is more efficient than the methods proposed by Harville [4] and Zhao et al [5], but they also used the static background model. Hence, the method has the similar limitations.

Tsutsui et al [7] uses stereovision and optical flow to recreate the 3D trajectory of a moving person. The optical flow method is very sensitive to illumination change and require static background model. Zhongwei et al [8] used stereoscopy in combination with camshaft algorithm to reconstruct the 3D Trajectory. They used dynamic programming for the correspondence matching and disparity calculation.

Park at el. [2] used the multiple perspective projections of a scene to reconstruct 3D trajectory of moving objects. They used co-ordinate independent basis vectors derived from the stationary areas of the scene, which reduces the computational complexity of the trajectory reconstruction.

The method proposed in this paper is an extension of the method presented by Heath and Guibas [4] which uses sparse stereo vision and particle filter to reconstruct the 3D trajectory. In contrast to multiple stereo cameras used by

Heath and Guibas [4], the proposed method uses one stereo camera to reconstruct the trajectory. Further, the proposed method uses raw stereo videos to locate the moving object and the stereoscopy is only used for the centroid of the detected object to reconstruct the 3D trajectory. The use of the stereoscopy for the centroids makes the method very efficient so that it can be used in real time applications.

2.1. Stereo vision

The binocular stereo vision, also called stereopsis or stereoscopy, uses two images of a scene captured from two slightly different viewpoints for the purpose of gaining the depth of a scene point by exploiting the difference in projections of the scene point in the two images [9]. The research in stereopsis is inspired by the capabilities of human vision system which is capable of recreating three dimensional model of the real world from the two images captured by eyes.

From a computational standpoint, stereo vision is a two step process. In the first step, correspondences among the pixels of the two images are determined. The correspondence problem refers to finding which item in the first image (reference image) corresponds to which item in the second image (target image). Normally, the left image is taken as the reference image and the right image is taken as the target image. In the second step, the displacement of corresponding pixel pairs and stereo camera parameters are utilized to reconstruct the 3-D model of the captured scene.

The geometry of binocular stereo vision is shown in figure 1. A simple stereo vision system consists of two identical cameras placed side by side with non-zero baseline distance *b*. The image planes of both the cameras are coplanar. A feature point *P* in the scene is projected by the two cameras at different positions on the image plane. The displacement between the locations of the two features in the image plane is called disparity. The plane passing through the camera centers and the feature point in the scene is called the epipolar plane. The intersection of the epipolar plane with the image plane defines the epipolar line. In a calibrated stereo system, each point in one image (left image) lies on the same horizontal scan line in the other image (right image). This is called the epipolar constraint and is an essential requirement of all stereo algorithms developed till now. Given this geometrical formation, the corresponding point P_r (in the right image) of any point P_l (of the left image) or vice versa may be found along its epipolar lines. From figure 1, the point *P* is observed at points P_l and P_r in the left and right image planes, respectively. Let the origin of the coordinate system coincides with the left lens center. Comparing the two triangles PMC_l and P_lLC_l

$$\frac{x}{z} = \frac{x_l}{f} \tag{1}$$

Similarly, from the similar triangles PNC_r and P_rRC_r

$$\frac{x-b}{z} = \frac{x_r}{f} \tag{2}$$

Combining equations (1) and (2) and using principle of triangulation, the depth *z* of a point in the scene is calculated as

$$z = \frac{bf}{d} \tag{3}$$

Where,
b = baseline distance between the two cameras,
f = focal length of the cameras,
z = depth of the scene point,
x_l = distance of scene point projection from left camera axis in the left image,
x_r = distance of scene point projection from right camera axis in the right image,
d = disparity, which is equal to $x_l - x_r$.

Thus, it is clear from equation (3) that the depth of a scene point is inversely proportional to its disparity. If the disparity of a scene point is known, its depth can be easily estimated. Disparities can be computed for only those features which are visible in both the images. Features visible in one image but not the other are said to be occluded. Recovering accurate depth of occluded regions is a difficult task and can only be estimated by information based on neighboring features.

2.2 Particle filter object tracking system

Particle filtering is a sequential Bayesian inference technique widely used for 2D object tracking. Particle filtering inherently maintains multiple hypotheses and does not require parametric representation of posterior distribution [4]. Computationally, particle filter used for object tracking is a four step process. In the first step, N particles with random locations are assigned to each tracker, with an assumption that the object can be found at any location. This is the initialization step and is executed only once. In the second step, for the next time step, the new potential location in the next video frame is predicted based on a motion model. The motion model need not be linear. This is the prediction step. In the third step, the measurement data, also called observation model, are used to assign importance weights to various particles. The higher weights are assigned to particles that agree with the measurements. In the last step, N particles of high likelihood probability for the potential future locations of the object are selected using resampling with replacement.

Arulampalam et *al.* [11] describes various resampling algorithms in detail. In this paper, a color based particle filter is used to track the object in both stereo videos, which is described below.

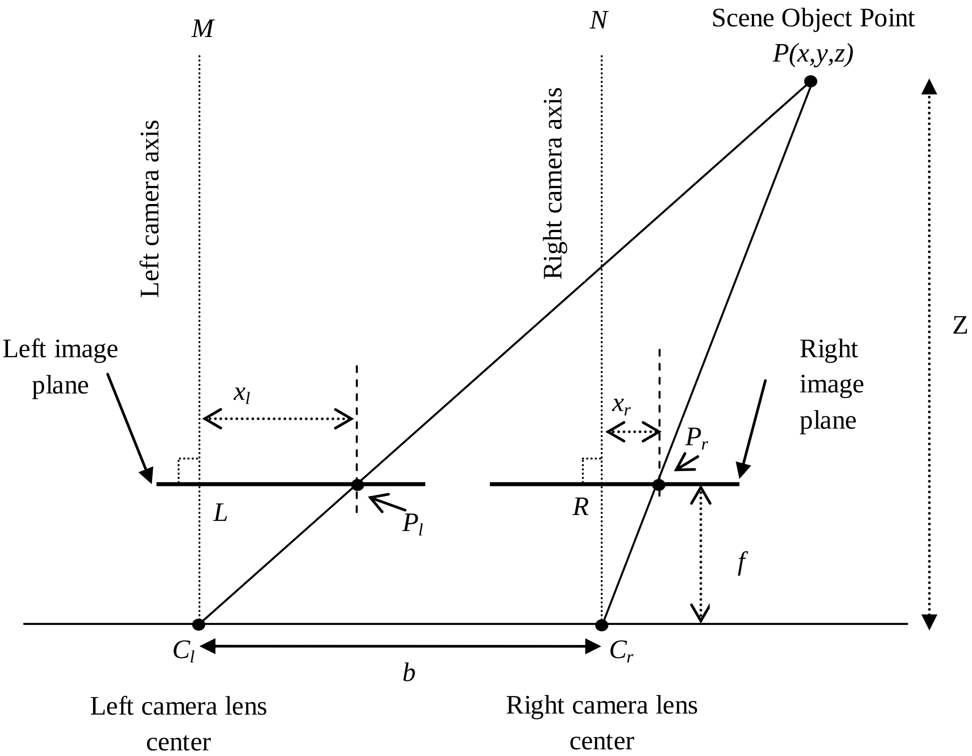


Figure 1: Depth estimation from a stereo system [10]

5.2.1 Color Based Particle Filter

The color of an object is one of the most widely used features in a tracking system. The major advantage of using color distribution is that it achieves tracking robustness against non-rigidity, rotation and partial occlusion.

In a particle filter based tracker, the state vector of each particle is given by equation 4.

$$s = [x, y, \dot{x}, \dot{y}] \qquad (4)$$

where x and y represent the location of a particle and $\dot{x}$ *and* $\dot{y}$ represent its velocity at a given time step. The motion model and observation model used are discussed next.

After initialing the state vector of each particle with random values, a motion model, given in equation 5, is used to predict the state vector for the next time step.

$$s_t = A.s_{t-1} + \varepsilon_{t-1} \qquad (5)$$

where s_t is the predicted state at time t, s_{t-1} is the previous state, A is the state transition model and ε_{t-1} is the noise term, usually sampled from a Gaussian distribution.

In a tracking situation where the object is assumed to be moving with a constant velocity, the motion model can be given by equation 6.

$$\begin{pmatrix} x_t \\ y_t \\ \dot{x}_t \\ \dot{y}_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x-1} \\ \varepsilon_{y-1} \\ \varepsilon_{\dot{x}-1} \\ \varepsilon_{\dot{y}-1} \end{pmatrix} \qquad (6)$$

In a color based tracking system, the particles whose color distributions are similar to the target model are assigned a likelihood probability for their propagation to the next step. The Bhattacharya distance *(w_t)*, which is calculated using equation 7 for the specified variance σ of the color distribution and color distance d, is widely used to calculate likelihood probability.

$$w_t = \frac{1}{\sqrt{2\pi}\sigma} . \exp(-\frac{d^2}{2\sigma^2}) \qquad (7)$$

For tracking a red object, the color distance d is calculated using equation 8, which assigns a larger weight to a particle located at a pixel whose color is closer to the red color.

$$d = \sqrt{(r-255)^2 + g^2 + b^2} \qquad (8)$$

where r, g and b are the red, green and blue components of a pixel's color in the RGB color model.

Finally, the particles for the next step are selected using resampling with replacement to avoid degeneracy problem.

III. THE PROPOSED METHOD

In this paper, the stereo vision concept is combined with a color based particle filter object tracker to reconstruct the 3D trajectory of a moving object. The basic flow diagram of the proposed method is given in figure 2. Before using the stereo camera for 3D trajectory reconstruction, it is calibrated using checker board images at different locations and orientations captured by the stereo camera. The stereo camera calibration toolbox of Matlab is used for this purpose. The calibration parameters of the stereo camera used in this study are shown in table 1. For the detailed description of these parameters, readers can refer Matlab online help of the stereo camera calibration tool box.

Table 1: Calibration parameters of the stereo camera

S. No.	**Parameter**	**Value**
	Intrinsic parameters of the left camera	
1	Focal length	[555.80785; 557.34962];
2	Principle point	[330.17268 ; 244.70437];
3	Skew	[-0.21713; 0.12312; 0.00487; 0.00163; 0.00000];
4	Distortion	[0.00000];
	Intrinsic parameters of the right camera	
5	Focal length	[558.11555; 559.49020];
6	Principle point	[314.72552 ; 232.64247] ;
7	Skew	[-0.22750; 0.14968 ; 0.00352 ; 0.00058; 0.00000]
8	Distortion	[0.00000];
	Extrinsic parameters(position of the right camera with respect to the left camera center	
9	Rotation Vector	[0.00189 ; -0.00040 ; 0.00653];
10	Translation vector	[-152.51204 ; -0.21901; -5.28905] ;

In the beginning of object tracking, the camera parameters are loaded into the memory, a stereo video is captures and two independent color-based particle filters are initialized. One particle filter tracks the object in the left video and the other in the right video. Table 2 shows the particle filter tracker parameters used in the present study to track a red colored object. For more detail about the relevance of these parameters, readers can refer [13].

Table 2: The Values of the particle filter parameters used in this study

. No.	**Parameter**	**Value**
1	Number of particles	2000
2	Standard deviation of the color distribution model	5
3	Translation model	25
4	Velocity model	5

In the next step, one frame of the left video and one frame of the right video are read. Once the frames are available, the particle filters are used to track the locations of the object in both the image frames and the centroid locations of the object in both the images are calculated as the mean of the current locations of all the particles.

In the final step, the triangulation principle of randomly oriented cameras is used to calculate the current location of the tracked object in three dimensions with reference to the left camera co-ordinate system. Finally, the 3D trajectory is plotted as a 3D plot.

IV. PROTOTYPE SYSTEM AND EXPERIMENTAL RESULTS

A prototype system is developed to test the above concepts. The prototype system consists of a stereo camera and a 2.27 GHz Intel(R) Core(TM) i3 CPU laptop running Windows 7. The stereo videos are captured in a normal lighting condition. For generating the ground truth data, the tracked object is moved along a handled chair whose measurements and the distance from the camera center are measured with the help of a measuring tape. Matlab is used as the programming environment.

For the evaluation of the prototype system, a stereo video of 150 frames is captured, during which a red colored bat is moved along the chair handles and the back support edges. Figure 3 shows some frames along with particles as blue dots and centroids as yellow circles.

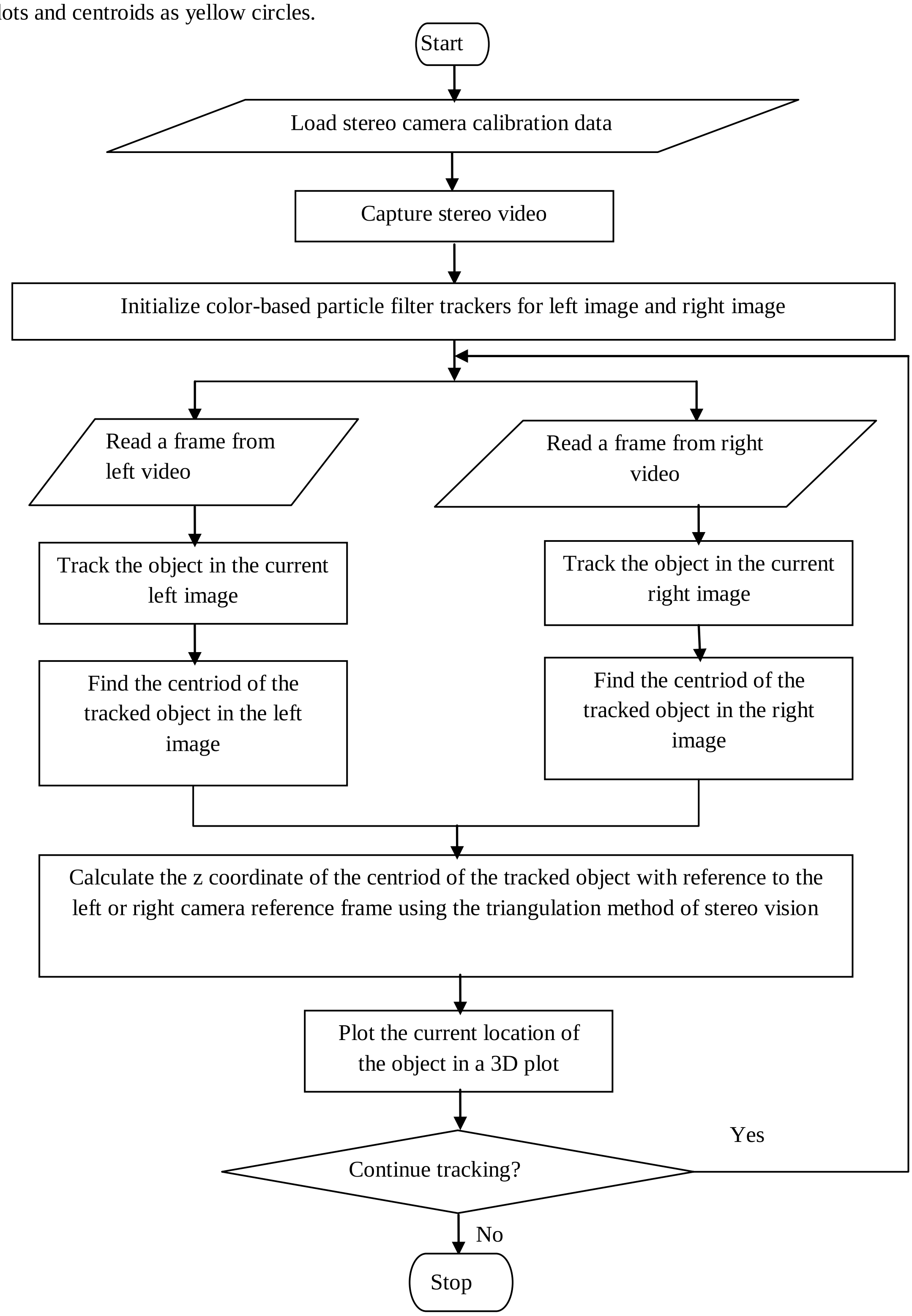


Figure 2: Flow diagram of the proposed 3D trajectory reconstruction

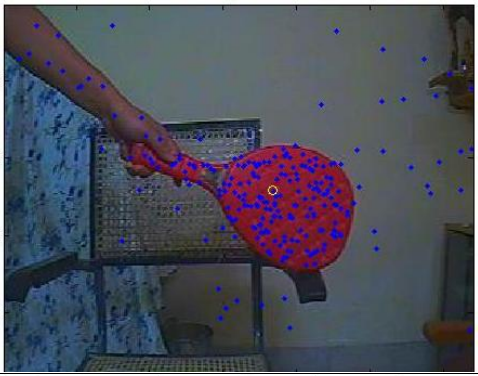
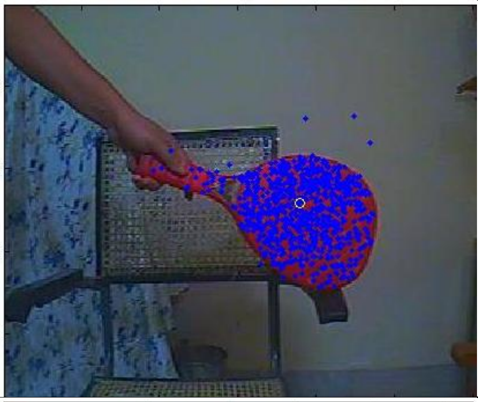
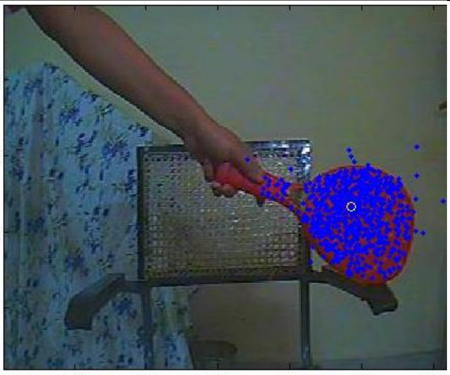
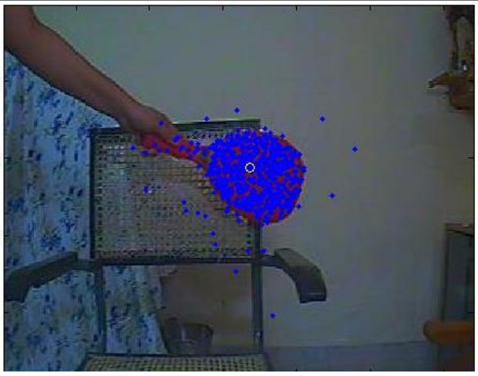
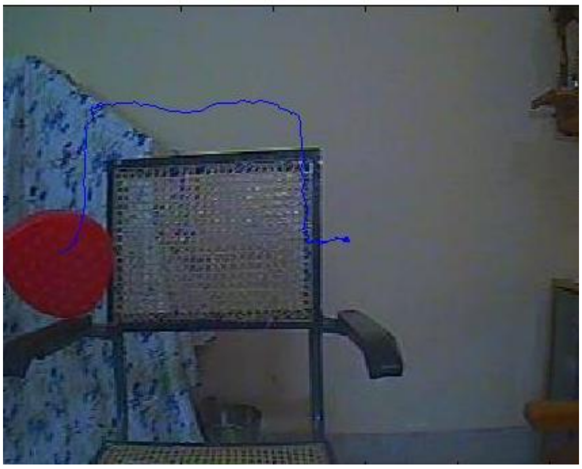
Frame No.	Left video	Right Video
1		
6		
50		

Figure 3: Some sample frames of the test stereo video with particles

The first five frames are used for warming up the particle filter tracker. Figure 4(a) and 4(b) show the 2D trajectories of the object for left and right videos, respectively. Figure 4(c) shows the computed 3D trajectory using the triangulation principle of stereo vision. The waviness in the trajectories is primarily due to jerky motion of the hand and change in the orientation of the object as it moves along the chair. In future research attempt will be made to reduce the waviness of the reconstructed 3D trajectory.

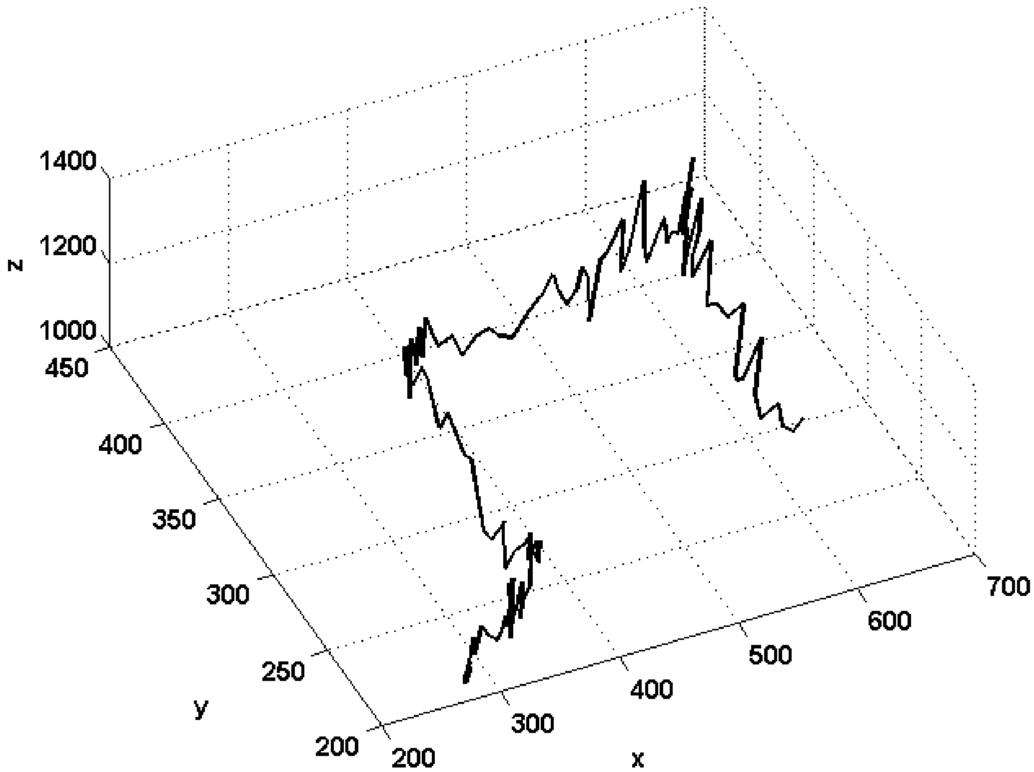
V. CONCLUSIONS

Moving object detection and tracking in videos is an actively researched area for the last two decades, but most of the research is limited to two dimensional trajectory and tracking. However, many practical applications demands 3D trajectory of moving objects, such as collision detection, game playing robots, human machine interaction, etc. This paper presents an experimental study to reconstruct the 3D trajectory of a coloured moving object by combing particle filter and stereo vision. These two techniques have been combined as it is not possible to generation 3D trajectories without making some assumptions about scenes from monocular videos. Although, the initial results obtained are quite encouraging, but the generated trajectory has error and waviness. Further research is required to reduce the waviness and generate more accurate 3D trajectories.



(a) 2D trajectory in the left video

(b) 2D trajectory in the right video



(c) 3D trajectory of the object

Figure 4: The generated trajectories of the object

VI. REFERENCES

1. J. Sköld, "Estimating 3D-trajectories from Monocular Video Sequences", Degree Project in Computer Science and Communication, School of Computer Science and Communication, KTH Royal Institute of Technology, Stockholm, Sweden, 2015.
2. H. S. Park, T. Shiratori, I. Matthews and Y. Sheikh, "3D Trajectory Reconstruction under Perspective Projection", International Journal of Computer Vision, Vol 115, Issue 2, pp 115–135, 2015.
3. K. Heath and L. Guibas, "Multi-person Tracking from Sparse 3D Trajectories in a Camera Sensor Network". Second ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC 2008.
4 M. Harville, "Stereo person tracking with short and long term plan-view appearance models of shape and color," IEEE International Conference on Advanced Video and Signal based Surveillance, 2005.
5 T. Zhao, M. Aggarwal, R. Kumar, and H. Sawhney, "Real-time wide area multi-camera stereo tracking", Computer Vision and Pattern Recognition, pp. 976–983, 2005.

6 A. Mittal and L. S. Davis, "M2tracker: A multiview approach to segmenting and tracking people in a cluttered scene using region-based stereo," European Conference on Computer Vision, pp. 18–36, 2002.
7 H. Tsutsui, J. Miura, and Y. Shirai, "Optical flow-based person tracking by multiple cameras," International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp. 91–96, 2001.
8 Z. Zhou, M. Xu, W. Fu and J. Zhao, Object Tracking and Positioning Based on Stereo Vision, Applied Mechanics and Materials, Vol: 303-306, pp 313-317, 2013.
9 Rachna Verma, 3-D Object Reconstruction From Stereo Image Pairs, PhD Thesis, Department of Physics, J N V University, Jodhpur, India, 2013.
10 Jain R., Kasturi R., Schunck B. G., "Machine Vision", *McGraw Hill International Edition,* 1995.
11 M.S. Arulampalam , S. Maskell , N. Gordon , T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking", IEEE Transactions on Signal Processing Vol. 50, Issue: 2, 2002.
12 Rachna Verma, Moving Object Detection and Tracking in Videos, ME Seminar (Submitted), Department of Computer Science, JNV University, Jodhpur, 2017.
13 Chong Chen and Dan Schonfeld," A Particle Filtering Framework for Joint Video Tracking and Pose Estimation" IEEE Transactions On Image Processing, Vol. 19, No. 6, June 2010.

