3D Environment Reconstruction Using Mobile Robot Platform and Monocular Vision

Keshaw Dewangan, Arindam Saha, Karthikeyan Vaiapury and Ranjan Dasgupta

Abstract Constructing a 3D map/perception model of an unknown indoor or outdoor environment using robotics is of compelling research nowadays because of the importance of the automatic monitoring system. Available IMU sensors and mobile robot kinematics allow 3D reconstruction to be finished in near real-time using a very low cost robotic platform. In this paper, we describe a framework for dense 3D reconstruction on an inexpensive robotic platform using a webcam and robot wheel odometry. Our experimental results show that our technique is efficient and robust to a variety of indoor and outdoor environment scenarios with different scale and size.

Keywords 3D environment reconstruction · Mobile robot · Robot operating system · Odometry · Camera calibration · Optical flow · Epipolar geometry · Robot locomotion

1 Introduction

The 3D technology is well established and accepted in manufacturing, chemical, automobile, construction industries and showing keen interest in investigating how this can be applied in practice. Robot based solution is in high demand in the market, especially manufacturing and chemical industries to inspect hazardous area where human cannot easily go. A variety of affordable mobile robots are available

K. Dewangan (✉) · A. Saha · K. Vaiapury · R. Dasgupta
TCS Innovation Labs Kolkata, Kolkata, India
e-mail: keshaw.dewangan@tcs.com

A. Saha
e-mail: ari.saha@tcs.com

K. Vaiapury
e-mail: karthikeyan.vaiapury@tcs.com

R. Dasgupta
e-mail: ranjan.dasgupta@tcs.com

R.K. Choudhary et al. (eds.), *Advanced Computing and Communication Technologies*,
Advances in Intelligent Systems and Computing 452,
DOI 10.1007/978-981-10-1023-1_22

in the market due to the emerging advancement in robotics field. These mobile robots are equipped with different low cost sensors (like camera, IMU sensors etc.) including a light weight computing unit. So the possibility of environment monitoring and verification in 3D space using such low cost mobile robot is manifold. In fact, there is a quite powerful and stable structure-from-motion pipeline readily available for reconstructing 3D model from multiple 2D images as shown in [1–3].

In a recent work [4], Pradeep et al. has described a methodology for markerless tracking and 3D reconstruction in scenes of smaller size using RGB camera sensor. It tracks and re-localizes the camera pose and allows for high quality 3D model reconstruction using a webcam. Pizzoli et al. [5] proposed a solution by adapting a probabilistic approach in which depth map is computed by combining bayesian estimation and convex optimization techniques. All these implementations are limited to a small scene reconstruction and not suitable for an entire 3D environment creation.

The 3D reconstruction of an environment from multiple images or video captured by a single moving camera has been studied for several years and is well known as Structure-from-Motion (SfM). Recently, smart phones are used for image acquisition due to its low cost and easy availability. So researchers used smart phones sensors like accelerometer, magnetometer for data collection and 3D reconstruction, it reduces computation [6, 7] and few works such as [8, 9] have accomplished this, but the output is noisy due to a fast and course reconstruction.

A system capable of dense 3D reconstruction of an unknown environment in real-time through a mobile robot requires simultaneous localization and mapping (SLAM) [10]. In our system, localization of the robot is done from odometer and the robot movement is controlled by user. The estimation of accumulated error is done from expected next position and actual odometer value of left and right wheel, so the complexity is lesser in this case.

In this context to fulfill these requirements, we present an end to end framework capable of generating 3D reconstruction of an environment based on the image/video captured through a remote platform mounted on a two wheel based robot. This work is a core part of our system presented in [11]. Firebird VI robot [12] is used in our experiments which allows navigation across a given environment. Robot captures images, odometry and IMU data and sends to backend server where, 3D view of environment is constructed using some selected key frames from the captured images and poses information. The novelty of the proposed system is reconstructing an entire environment in near real-time using a very low cost user guided mobile robot platform.

We demonstrate the framework along with the performance of the approach with computation time details. We present different type of results to show the capability and robustness of the system to work in a wide range of scenes and environments. We also evaluate the accuracy of the reconstruction and compared with ground truth.

2 Robotic Platform Description

In our work, Firebird VI robot (refer Fig. 1) is used which controls all operations through ROS [13]. The framework is capable to work on any robotic platform that supports ROS and Firebird VI is chosen due to its low cost and readily availability. The block diagram of the entire system is provided in Fig. 2.

Fig. 1 *Left to right* figure shows the Firebird VI robot used in our experiments, sample data sets and corresponding reconstructions

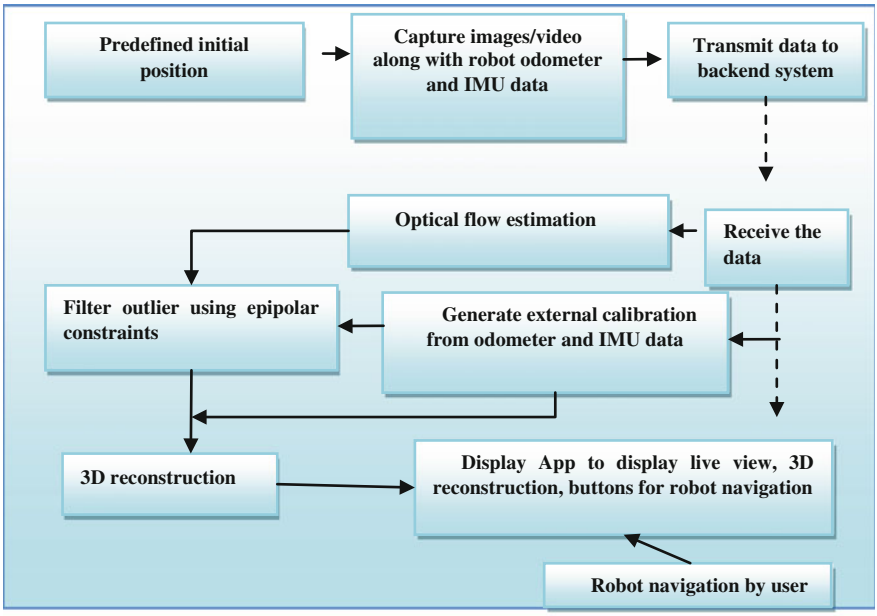

Fig. 2 Proposed system architecture: robotic vision

2.1 Data Acquisition

The robot starts navigation from its initial position and this is taken as the origin of world coordinate system by default in order to reduce the complexity. So the starting position is considered as predefined position. One very light weight and less computing power system is mounted on the top of robot, where master ROS handles tasks of robot navigation, pose estimation (using odometry and IMU sensor data) and Image capturing and transmitting to the backend server.

Image capturing is a major task in data acquisition, which is handled by the ROS package [14]. Another ROS program is running at the back-end system which is connected with the master ROS via, a wireless network and subscribes topics published by master ROS like images, odometry and IMU data. Back-end system uses odometry and IMU data to get pose parameters for some selected key frames and initiate the 3D reconstruction of the entire environment. The detail of 3D reconstruction process is explained in the next sections.

2.2 Camera Calibration

The camera mounted on the servo of the robot is an off-the-shelf webcam and it is fixed throughout our experiments. Pin-hole camera model [1] convention is used and zoom factor of the camera and resolution of captured images are kept constant for a fixed internal calibration matrix. The internal calibration process is performed offline using well known checker board methods as described by Zhang in [15, 16].

Orientation matrix of camera is calculated from servo angles and heading information of robot. Robot pose information can be estimated accurately by fusing odometry and IMU using Extended Kalman filter (EKF) data. EKF cleans up the noises in odometer data [17, 18]. The transformation between servo and robot body frame of reference would give the orientation angles of the camera with respect to the world coordinate system. The orientation matrix (R) of the camera is computed using (1).

$$\mathrm{R} = \mathrm{Rz} * \mathrm{Ry} * \mathrm{Rx}. \tag{1}$$

where $\mathrm{R_i}$ denotes the rotation matrix along axis *i*.

Since robot is moving in x–z plane and there is no camera rotation along X-axis (tilt angle), hence camera orientation depends only on robot rotation angle θ and camera pan angle α along Y axis. In this case Rz and Rx will be identity matrix and Ry will be given as shown in (2).

$$\begin{bmatrix} \cos(\alpha+\theta) & 0 & \sin(\alpha+\theta) \\ 0 & 1 & 0 \\ -\sin(\alpha+\theta) & 0 & \cos(\alpha+\theta) \end{bmatrix} \tag{2}$$

2.3 Dense Stereo Matching

The dense stereo matching is vast and we refer to [19] for a comparison of all existing methods. In fact, there are few relevant works available on real-time, dense reconstruction using a monocular moving camera.

Motion estimation by means of optical flow is standard technique for providing dense sampling in time. The predominant way of estimating dense optical flow in today's computer vision literature is by an approach of integrating rich descriptors into the variational optical flow setting as described in [20]. The main advantage of the selected approach is the ability to produce better results in a wide range of cases and also for large displacement.

Large displacement optical flow is a variational optimization technique which integrates discrete point matches with continuous energy formulation. The final goal is to find the global minima of the energy and for that the initial guess of the solution has to very close to the global minima. The entire energy is globally minimized and the details of minimization procedure are studied in [20].

The given approach is not directly applicable for any near real-time system because of its high computation. The running time between a pair of frames of 640 × 480 resolutions on a 2.13 GHz is about 55 s. The performance in terms of time is drastically improved by the use of general purpose graphics processing unit (GPU) as described in [21]. The parallel implementation on a GPU yields about 78 times faster performance compare to a serial C++ version. This implementation is further used in a dense 3D reconstruction with a hand held camera [22] where the system is not in real-time due to the tracking of every frame in captured video sequence.

The n-view point correspondence generation is carried out using the GPU implementation as described in [21]. The point trajectories are generated between some selected key frames from the captured video. Optical flow has an effect of accumulating errors in the flow vector. So, a long trajectory suffers from this error and leading to a significant drift. Short trajectories are almost free from the drift error, but the triangulation process suffers due to small base line measurements. Hence, we have not chosen consecutive frames rather selected frames that are having base line about 10 cm and the trajectory length chosen as less than 15 in our experimental setup.

2.4 Outlier Detection

Detecting outlier is a very primitive task before doing any further processing with the available information. Outlier detection process is very straight forward and it follows the epipolar constraints [1] as shown in (3). Accurate camera calibration estimation produces a better estimation of pair wise fundamental matrix (F) [1] which is used for noise cleaning. The corresponding points (x, x') ideally should follow the epipolar constraints as given in (3) [1]. In reality, the value never

becomes zero rather it goes very close to zero. We used a dynamic threshold based on percentage of rejection because static threshold does not hold good for different type of scene. The threshold is always consider as below of 3 pixels.

$$x'^{T}Fx = 0 \tag{3}$$

2.5 3D Model Generation

The 3D point cloud is created using well known triangulation process as described in [23]. Each point is back projected onto the image plane to calculate the back projection error. Any 3D point with back projection error more than 3 pixels is considered as outlier.

The whole scene reconstruction is done in an incremental way. Images are divided into small sets such that the trajectory length is not more than 15 images. Each subset is merged after triangulation to get the final reconstruction. The scale rectification is done using IMU sensors [24].

3 Results

The implementation environment consists of Firebird VI robot as shown in Fig. 1 and a back-end system having Intel(R) Xeon(R) E5606 processor running at 2.13 GHz along with a NVIDIA Tesla C2050 Graphic Card. One ZOTAC ZBOXHD-ID11 is mounted on top of the robot. ROS hydro is installed inside Ubuntu 12.04 LTS in all the systems.

The entire capture task is running on the ZOTAC box mounted on the Firebird VI. Image is captured with 640 × 480 resolution using a Logitech C920 webcam. The 3D model reconstruction is carried out on the backend system due to less processing power of ZOTAC box.

3.1 Outputs

We presented two sample reconstruction sequences in different environment to demonstrate the robustness and usability of our solution. The presented samples contain several images which are affected by occlusion, motion blur.

Figure 3 shows a sample reconstruction of a wall in an indoor office environment of size 13 × 9.5 feet. The reconstruction is carried out with only 7 images in a single iteration. The reconstruction shows noisy output at the right bottom part and

Fig. 3 *Topleft* sample image for reconstruction, *Rest* 3D point cloud after reconstruction

it is due to the fact that images that are used for reconstruction are taken only from the frontier side of the wall.

In Fig. 1 we presented another result where the data is captured in a living room. The three sides of the room are captured where different objects are placed. The dimension of the room is about 13 × 11 feet. Another data presented in Fig. 1 is captured of outdoor environment from more than hundred feet. The user can guide the robot to go closer and capture the frames to produce a more accurate and dense points in any required portion of the environment.

The presented samples justifies that our proposed solution is capable of reconstructing both indoor and outdoor environments without any size limit which is the basic requirement for any 3D reconstruction system. The robot locomotion is guided through user and this is advantageous to focus onto a specific object in the environment for better observation through high quality reconstruction.

3.2 Execution Time

The timing details of the samples presented above is given in Table 1.

Table 1 Timing details

Data sets	Frames	Tracking	Inlier detection and triangulation	Total
Wall	7	13	3 s	16 s
Living room	62	127	25 s	152 s

Fig. 4 Reconstructed structure verification with ground truth

3.3 Ground Truth Verification

The accuracy of our results verified against the ground truth by measuring the distance between two points. One such measurement comparison is shown in Fig. 4 where outer distance between two legs of the reconstructed chair is 21.2651 in. and the actual measured value is 21.25 in. It shows our work accuracy is 98–99 %.

4 Conclusion

We have presented an approach for dense 3D reconstruction of any uncontrolled indoor and outdoor environment through monocular robotic vision, IMU and odometry. Our results shows that proposed work is useful for constructing 3D of indoor and outdoor environment in near real time. Further work is planned to integrate an IR thermal sensor and fuse thermal information onto the 3D structure to create an opto-thermal 3D and corrosion, erosion measurement.

References

1. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press. ISBN: 0-521-54051-8 (2003)
2. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: Dense tracking and mapping in real-time, ICCV, pp. 2320–2327 (2011)

3. Pollefeys, M., Nister, D., Frahm, D.J.M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G., Towles, H.: Detailed real-time urban 3D reconstruction from video. IJCV **78**(2–3), 143–167 (2008)
4. Pradeep, V., Rhemann, C., Izadi, S., Zach, C., Bleyer, M., Bathiche, S.: MonoFusion: Real-time 3D reconstruction of small scenes with a single web camera. In: The 13th IEEE International Symposium on Mixed and Augmented Reality, pp. 83–88 (2013)
5. Pizzoli, M., Forster, C., Scaramuzza, D.: REMODE: probabilistic, monocular dense reconstruction in real time. In: IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, pp. 2609–2616 (2014)
6. Saha, A., Bhowmick, B., Sinha, A.: A system for near real-time 3D reconstruction from multi-view using 4G enabled mobile. In: IEEE International Conference on Mobile Services (MS), pp. 1–7, (2014)
7. Tanskanen, P., Kolev, K., Meier, L., Paulsen, F.C., Saurer, O., Pollefeys, M.: Live metric 3D reconstruction on mobile phones. In: ICCV, pp. 65–72 (2013)
8. Bhowmick, B., Mallik, A., Saha, A.: Mobiscan3D: A low cost framework for real time dense 3D reconstruction on mobile devices. In: IEEE 11th International Conference on Ubiquitous Intelligence and Computing, IEEE 11th International Conference on and Autonomic and Trusted Computing, and IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops (UTC-ATC-ScalCom), pp. 783–788 (2014)
9. Mallik, A., Bhowmick, B., Alam, S.: A multi-sensor information fusion approach for efficient 3D reconstruction in smart phone. In: International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), pp 291–298 (2015)
10. Davison, A.: Real-time simultaneous localisation and mapping with a single camera. In: IEEE International Conference on Computer Vision, pp. 1403–1410 (2003)
11. Deshpande, P., Reddy, V.R., Saha, A., Vaiapury, K., Dewangan, K., Dasgupta, R.: A next generation mobile robot with multi-mode sense of 3D perception. In: International Conference on Advanced Robotics (ICAR) Istanbul, pp 382–387, (2015)
12. Firebird VI. http://www.nex-robotics.com/fire-bird-vi-robot-platform.html (2015). Access 20 Oct 2015
13. Martinez, A., Fernández, E.: Learning ROS for robotics programming. PACKT Publishing Ltd. (2013). ISBN: 978-1-78216-144-8
14. ROS usb Camera Package. http://wiki.ros.org/usb_cam (2015). Accessed 20 Oct 2015
15. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: International Conference on Computer Vision (ICCV'99), pp 666–673, (1999)
16. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)
17. ROS Robot Pose EKF Package. http://wiki.ros.org/robot_pose_ekf (2015). Accessed 20 Oct 2015
18. ROS Robot Localization Package. http://wiki.ros.org/robot_localization. Accessed 20 Oct 2015
19. Hirschmuller, H., Scharstein, D.: Evaluation of stereo matching costs on images with radiometric differences. IEEE Trans. Pattern Anal. Machine Intell. **31**(9), 1582–1599 (2009)
20. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. IEEE Trans. Pattern Anal. Mach. Intell. **33**(3), 500–513 (2011)
21. Sundaram, N., Brox, T., Keutzer, K.: Dense point trajectories by GPU-accelerated large displacement optical flow. In: European Conference on Computer Vision (ECCV), pp. 438–451, Crete, Greece, Springer, LNCS (2010)
22. Ummenhofer, B., Brox, T.: Dense 3D reconstruction with a hand-held camera. Springer, Berlin Heidelberg (2012)
23. Hartley, R.I., Sturm, P.: Triangulation. Comput. Vis. Image Underst. **68**(2), 146–157 (1997)
24. Nützi, G., Weiss, S., Scaramuzza, D., Siegwart, R.: Fusion of IMU and vision for absolute scale estimation in monocular SLAM. J. Intell. Robot Syst **61**(1–4), 287–299 (2011)