

A LOW-COST VISUAL RADAR-BASED ODOMETRY FRAMEWORK WITH MMWAVE RADAR AND MONOCULAR CAMERA

Yang-En Lu¹, Syun Tsai^{1*}, Meng-Lun Tsai², Kai-Wei Chiang¹

¹ Dept. of Geomatics, National Cheng Kung University, Tainan, Taiwan –
f64074065@gs.ncku.edu.tw, dino920135@gmail.com, kwchiang@geomatics.ncku.edu.tw

² High Definition Map Research Center, Dept. of Geomatics, National Cheng Kung University, Tainan, Taiwan –
taurusbryant@geomatics.ncku.edu.tw

Commission I, WG I/7

KEYWORDS: Visual Odometry, Radar Odometry, Millimeter Wave Radar, Simultaneous Localization and Mapping, Positioning

ABSTRACT:

One of the most popular research areas is low-cost navigation and positioning systems for autonomous vehicles. Determining a vehicle's position within a lane is critical for achieving high automation. Vehicle navigation and positioning relied heavily on the Global Navigation Satellite System (GNSS) service in open-sky scenarios. Nonetheless, GNSS signals were easily degraded due to various environmental situations such as urban canyons caused by multi-path effects and Non-Line-of-Sight (NLOS) issues. To perform robustly in complex scenarios, sensor fusion is the most common solution. The following paper presents a radar visual odometry framework to improve the lack of scale factors for monocular cameras and poor angular resolution for radar. The framework is based on the characteristics of camera and radar sensors which have complementary advantages in each other. The results show that the proposed framework can be used to estimate general 2D motion in an indoor environment and correct the unknown scale factor of Monocular Visual Odometry in a real-world setting.

1. INTRODUCTION

According to the prediction of Boston Consulting Group (BCG), the global self-driving vehicle market will reach US \$42 billion in 2025, and autonomous vehicles will account for 12.4% of the overall vehicle market. Therefore, it can be seen that mapping and navigation technology have a certain market. By 2035, the market scale will double and grow. At present, GNSS has been highly relied on to obtain positioning and navigation services outdoors, but the positioning system based on GNSS has caused positioning errors due to the multipath effect caused by urban canyon, and signal cycle slips. To improve the robustness of positioning, the rapid development of multi-sensor integrated positioning systems not only improves the positioning accuracy but also gradually reduces the cost of mapping.

To overcome the limitations of complex scenarios, multi-sensor platforms such as Inertial Navigation System (INS), Global Navigation Satellite System (GNSS), cameras, Light Detection and Ranging (LiDAR), and radar have become possible solutions. Among them, camera sensors had the characteristics of low-cost and widely studied in robotic and navigation fields. Camera sensors can provide vehicles relative pose change with Visual Odometry (VO) and Simultaneous Localization and Mapping (SLAM) techniques. In the simplest V-SLAM systems, a monocular camera is used to determine the ego-motion and build a map without true scale. For example, ORB-SLAM (Raúl Mur-Artal et al., 2015) is a feature-based monocular SLAM system that operates in small and large, indoor and outdoor environments. The system allows wide baseline loop closing and relocalization and includes full automatic initialization which performs a real-

time and robust trajectory estimation. However, several drawbacks were noticed in different VO or VSLAM algorithms. Such as lack of scale factors for the monocular camera, low ranging accuracy for stereo camera due to short baseline, and sensitivity to environment interference (light, rain, fog). Radar sensors can provide stable ranging and velocity measurement with less susceptibility to environmental interference. RadarSLAM (Ziyang Hong, Yvan Petillot and Sen Wang, 2020) is a large-scale SLAM based on mechanical millimeter-wave (mmWave) radar. The algorithm demonstrates the reliability and localization accuracy in various adverse weather conditions, such as dark night, dense fog and heavy snowfall. However, the scanning radar is quite expensive and heavy. Therefore, several ego-motion estimation methods using doppler radar have been proposed (Dominik Kellner et al., 2013) which can use single-chip mmWave radar to perform robust pose estimation.

The aforementioned characteristics of camera and radar sensors show these two sensors have complementary advantages. It leads the potential to form a low-cost visual and radar base odometry. The aim of this work is to implement visual radar based odometry using low-cost mmWave radar and monocular camera.

2. RELATED WORKS

In the literature, several in-depth analyses have been presented that highlight advantages and limitations of single-chip mmWave sensor and visual sensor.

Several research groups have proposed mmWave radars as a solution for various mobile robot tasks such as navigation,

* Corresponding author: Syun Tsai. Yang-En Lu, Syun Tsai, Kai-Wei Chiang are with the Department of Geomatics, National

Cheng Kung University (NCKU), Tainan City 701, Taiwan (ROC)

localization and mapping. Several studies are proposed to investigate the imaging capabilities of the radars for environment representation (G. Brooker, 2015) and 2D/3D SLAM (M. Jaud, 2014). However, these solutions involve bulky radar systems that provide dense measurements at the cost of increased physical size and price of the system. The mmWave radar has some limitations due to the beam being wider than the LIDAR sensor, which results in lower bearing resolution and cluttered measurements. The longer wavelength of radar causes the echo to be reflected off multiple objects on its return trip to the antenna, known as the multipath effect. This effect causes false range measurement, which produces lots of “ghost points” in one scan, which is even more challenging in indoor environments due to walls, ceiling and floor reflection (M. Adams and E. Jose, 2012).

A thorough experimental evaluation of ego-motion estimation with low-cost mmWave sensor is described in (Yasin Almalioglu et al., 2021), where mmWave radar system attached on top of a moving platform is used for the indoor ego-motion estimation. The study points out the recent advances in the integrated circuit and packaging technologies, it is even possible to integrate a frequency-modulated continuous-wave (FMCW) radar system operating at a higher frequency band. The study proposed Milli-RIO, an ego-motion estimation method based on single-chip low-cost mmWave radar, complemented by an Inertial Measurement Unit (IMU) sensor. The study uses a new point association technique to match the sparse measurements of low-cost mmWave radar and a model-free motion dynamics estimation technique for Unscented Kalman filter (UKF) using Recurrent Neural Network (RNN). The experiment takes place in a typical lab environment where it is tracked with a VICON tracking system that provides ground truth with sub-millimeter accuracy. As result, the successful implementation of mmWave radar odometry fused with IMU improves the reliability and versatility of mobile systems.

Instantaneous ego-motion using Doppler radar (Dominik Kellner et al., 2013) uses doppler radar unlike scanning radars that only acquire range measurements. Doppler radars can measure the velocity of a target object. The study proposed instantly determining the velocity and yaw rate of a ground vehicle with a single Doppler radar. It used RANdom SAmple Consensus (RANSAC) to detect stationary targets and least square adjustment to estimate the linear velocity of the vehicle. 3D ego-motion Estimation using low-Cost mmWave Radars via Radar Velocity Factor for Pose-Graph SLAM (Yeong Sang Park et al., 2021) designs a unique hardware configuration by combining two low-cost Doppler radars and estimating 3D instantaneous velocity. Furthermore, by applying RANSAC to this dual configuration followed by tangential motion refinement. They design a radar velocity factor for pose-graph SLAM and complete a 3D ego-motion in the integration with IMU. The radar provides instantaneous linear velocity and the relative positional difference between nodes. Leveraging the rotation from IMU, they complete a full 3D ego-motion estimation. The result shows that in a disaster environment with thick fog, where the camera and LiDAR revealed limited visibility. The system yielded reliable ego-motion inference in testing over both 2D and 3D motion.

Visual SLAM is a camera-based SLAM algorithm. Compared with traditional lidar data, visual data has the characteristics of low cost and a large amount of information. Visual SLAM is mainly composed of sensor data, visual odometry (VO), back-end optimization, and loop closure detection. There are three main types of visual sensors: monocular camera, stereo camera and RGB-D camera. In addition to image pre-processing, these

three sensors may be equipped with sensors such as IMU, so time synchronization processing is also required.

Visual odometry is used to estimate the camera motion and the position of local feature points between adjacent photos. The positioning methods are divided into direct method and feature-based method. Based on the front end of feature points, it has long been considered as the mainstream method of visual odometry. Feature points are used to provide the basis to identify the environment. Feature point is composed of key point and descriptor. Key point refers to the position of the feature point in the image, and descriptor is the information describing the pixels around the key point. There are many feature point extraction methods, such as ORB, GFTT etc.

3. VISUAL RADAR ODOMETRY FRAMEWORK

In this section, we will introduce the visual radar odometry framework. This framework integrates the instantaneous linear velocity estimated by mmWave radar and the rotation estimated by monocular camera. The development of visual radar system can be divided into the following steps:

1. Radar-based linear velocity estimation.
2. Visual Odometry implementation.
3. Visual Radar odometry integration.

3.1 Radar-based linear velocity estimation

The principle concept of the algorithm is to estimate the ego-motion based on the doppler velocity and azimuth angle of the measured reflections (targets) in the field of view. In the literature discussion, two algorithms (Kellner et al., 2013 and Stahoviak and Carl C, 2019) for solving the linear velocity of the platform using radar are mentioned respectively. The principles can be sorted out using Figure 1:

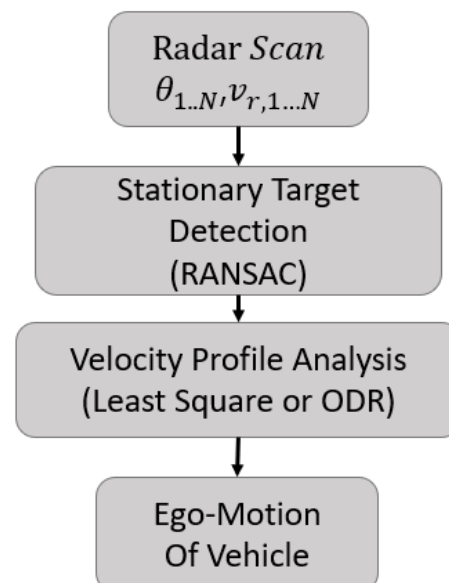


Figure 1. Steps of the ego-motion estimation using doppler radar.

For each measurement cycle, three steps are performed. First, the largest group of targets with the same movement is extracted. It is assumed that there is no group of targets with exactly the same

linear movement and a larger number of targets than all stationary targets in common. This allows classifying all targets of the largest group found as stationary targets. The movement of the radar sensor can be reconstructed by analyzing the returned radial velocity of all stationary targets with regard to their position in the azimuth angle. In the last step, the ego-motion of the vehicle is calculated from the sensor movement using the single-track model with the Ackerman condition.

If a platform is moving, from the radar sensor point of view all stationary targets move in the opposite direction. Their relative velocity is exactly equal to the sensor's velocity and heading direction. Nonetheless, it is hard to directly extract the velocity vectors of the vehicle as a Doppler radar can only measure the radial velocity component.

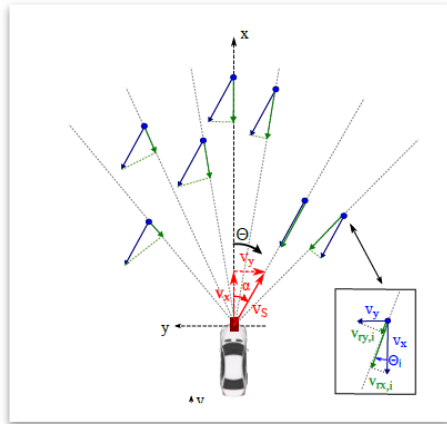


Figure 2. Stationary targets (blue) the radial velocity (green arrow) of the relative velocity (blue arrow) is measured by the Doppler Radar.

The key is that the radial velocity with respect to the angle of arrival (AoA) reveals the sinusoidal curve. The radial velocity and AoA have the relation illustrated in Figure 2. The relation can be shown as a mathematical problem as:

$$\begin{bmatrix} v_{r,1} \\ \vdots \\ v_{r,n} \end{bmatrix} = \begin{bmatrix} \cos(\theta_1) & \sin(\theta_1) \\ \vdots & \vdots \\ \cos(\theta_n) & \sin(\theta_n) \end{bmatrix} \begin{bmatrix} v_{x'} \\ v_{y'} \end{bmatrix} \quad (1)$$

Where $v_{r,i}$ = radial velocity of target i
 θ_i = angle of arrival of target i
 $v_{x'}$ = relative velocity in x-direction
 $v_{y'}$ = relative velocity in y-direction

Stationary target detection uses RANSAC. It is mainly used to remove the outliers in the data, which has a good performance in the smoothing of gross errors in the data. For this study, when there is a moving object or gross error, it is seen as an outlier. After the non-stationary objects are filtered out, substitute the remaining stationary objects into the equation mentioned above, and then use the least square adjustment (LS) or orthogonal distance regression (ODR) to obtain the estimate of the linear velocity. Least squares is arguably the standard method for fitting data to a model when there are errors in the observations. This can be interpreted as minimizing the sum of the squares of the residual from the observations. ODR is a total least squares regression method for finding the maximum likelihood estimators of parameters in measurement error models in the case of normally distributed errors which considers the error of angle estimation of the radar.

3.2 Visual Odometry implementation

The front-end visual odometry can give a trajectory and map in a short time, but due to the inevitable error propagation, the long-time and large-scale results are inaccurate. Therefore, the back-end optimization is mainly to deal with the error in the SLAM process. Considering the balance between accuracy and performance, there are many different approaches, such as EKF.

As the platform continues to move forward, the front-end VO is basically completed by the relationship between two adjacent frames. This algorithm relying on local constraints will inevitably lead to the accumulation of errors. Hence, it is necessary to select the key frame from the global perspective and determine whether the platform reaches the previous position. If a closed-loop is detected, it will send the message to the back-end, Pull the pose with cumulative error to the correct position.

In this study, the VO implements the ORB-SLAM algorithm. ORB-SLAM is a feature-based visual SLAM algorithm. ORB-SLAM adopts ORB feature, which is a combination and improvement of FAST feature points and BRIEF descriptor. ORB-SLAM includes map initialization, tracking, local mapping, and loop closure. Figure 3 shows the ORB-SLAM framework overview.

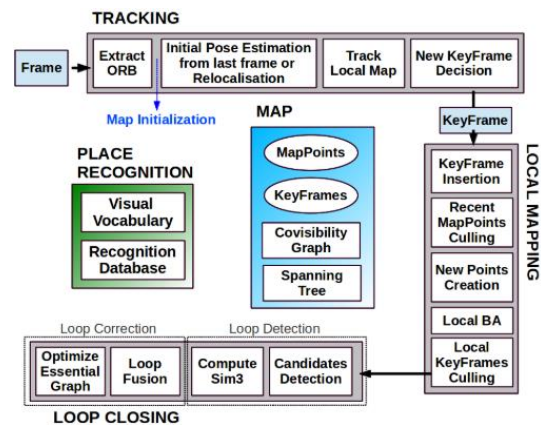


Figure 3. ORB-SLAM system overview, showing all the steps performed by the tracking, local mapping, and loop closing threads. (Raúl Mur-Artal et al., 2015)

3.3 Visual Radar odometry integration

This study integrates the instantaneous linear velocity estimated by mmWave radar and leverages the rotation from V-SLAM by monocular camera. Figure 4. shows the system integration diagram of Visual Radar Odometry.

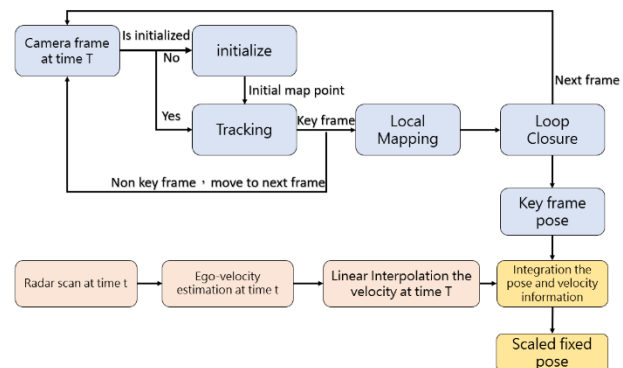


Figure 4. The Visual Radar odometry integration framework

The framework can be divided into two parts: estimation of rotation from the visual sensor and velocity estimation of radar. The rotation estimation of the visual sensor refers to the algorithm of ORB-SLAM which is shown as blue in Figure 4. The linear velocity estimation algorithm has been discussed in section 3.1. First, the non-keyframe is interpolated into its current pose according to time. The reason for interpolation is that the non-keyframe mainly has no significant pose change, so it can be regarded as a uniform constant velocity movement. Second, due to the sampling frequency of the radar being 10Hz, which is larger than the pose estimation frequency of the ORB-SLAM about 5Hz, the linear velocity estimated by radar is interpolated into the frame at that time to obtain the instantaneous linear velocity of the frame. Third, using the optimization solution after loop closing, calculate the pose change between the previous frame and the current frame. Then we could simply integrate each frame using the formula below to calculate the position of the platform in local coordinate defined by the first frame of the image. Figure 5. illustrate the integrating process between two frame t-1 and t using the angle velocity estimate from ORB-SLAM and linear velocity estimated by the mmWave radar.

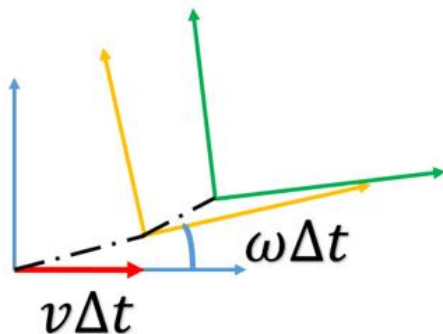


Figure 5. The displacement between t and t-1 can be calculated using the formula (2) and (3).

$$Sx = v\Delta t \quad (2)$$

$$Sy = v\omega(\Delta t)^2 \quad (3)$$

Where Sx = x-direction displacement
 Sy = y-direction displacement

4. EXPERIMENT

4.1 Experiment setup

The experimental setup of visual radar system in this study is shown in Figure 6. Single-chip mmWave radar (AWR1843 BOOST, Texas Instrument) with Raspberry-pi camera module v2 on a TurtleBot3 robot is used to develop the current Visual-Radar system. The TurtleBot3 provides the trajectory estimate by IMU and wheel encoder. And the mmWave radar provides 3-dimensional points with range, azimuth, elevation, and velocity information.

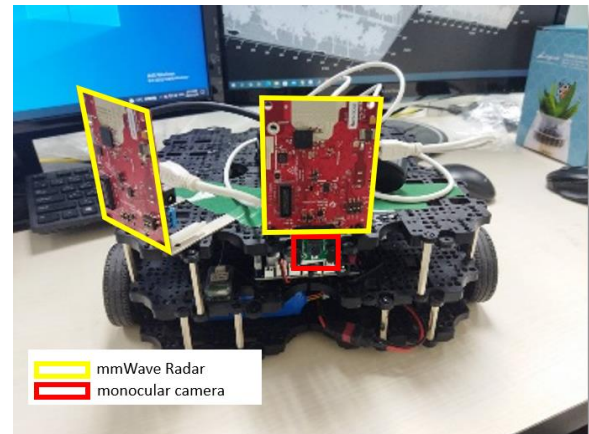


Figure 6. The experimental setup of visual radar system

4.2 Experiment environment

The experimental validation was performed on two types of environments. First, we mounted the radar on a mobile robot and moved in a straight line in the hallway. Second, we moved in a circle trajectory at the conference room. The experimental environment is shown in Figure 7.



Figure 7. (a)Hallway (b)Conference room

4.3 mmWave radar configuration

Firstly, in order to find the radar parameter suitable for this study. Therefore, we refer to many empirical formulas of radar and the limitations of hardware and try to adjust the best settings of radial velocity resolution, range resolution and angle resolution. There are three transmitting antennas and four receiving antennas. The mmWave parameters used in the of this study are shown in Table 1.

Parameter	Value
Frequency	77GHz
BandWidth	4GHz
Max range	8.42(m)
Range resolution	0.047(m)
Range accuracy	0.032(m)
Max AoA	+90 (deg)
Angle resolution	14.32 (deg)
Angle accuracy	2.81 (deg)
Max radial velocity	0.64(m/s)
Velocity resolution	0.04(m/s)
Velocity accuracy	0.02(m/s)

Table 1. The radar parameters in this study

4.4 Linear velocity estimation

Figure 8. shows the comparison between velocity estimation by radar and IMU. The time series of velocity in x and y directions analyzed by using Least Square (LS) and Orthogonal Distance Regression (ODR) are in good agreement with IMU in Figure 8. Statistic comparison between Instantaneous velocity estimated by LS and ODR in scenario 1(hallway) is depicted in Table 2. It

shows LS method has a smaller mean velocity difference from IMU in x-direction which is the forward direction. Therefore, LS method is applied in current study.

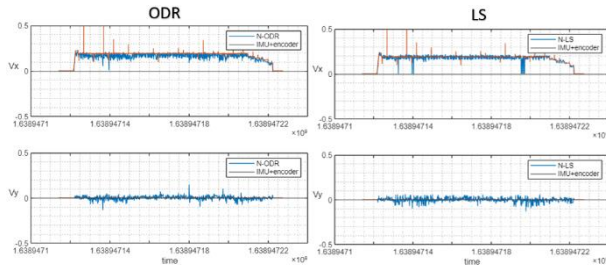


Figure 8. Comparison between Instantaneous velocity estimation using mmWave radar and IMU+encoder in scenario 1 (Hallway).

Scenario 1: Hallway		
Velocity diff.	LS	ODR
Unit: m/s		
x mean	-0.0065	-0.0187
y mean	-0.0017	0.0008
x Std	0.0265	0.0269
y Std	0.0232	0.0204

Table 2. Statistic comparison between Instantaneous velocity estimation using mmWave radar and IMU+encoder in scenario 1 (hallway)

4.5 Visual Radar odometry estimated trajectory

At last, the performance of visual radar odometry was verified in two different scenarios (hallway and conference room). The comparison of Radar aided VO and pure VO is shown in Figure 9. and Figure 10. And the comparison of trajectory between Visual Radar Odometry and IMU+encoder is shown in Figure 11 and Figure 12.

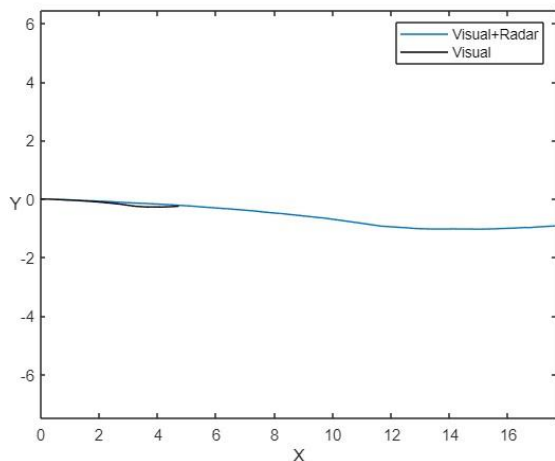


Figure 9. The comparison of trajectory between Radar Aided VO and pure VO in scenario 1 (Hallway)

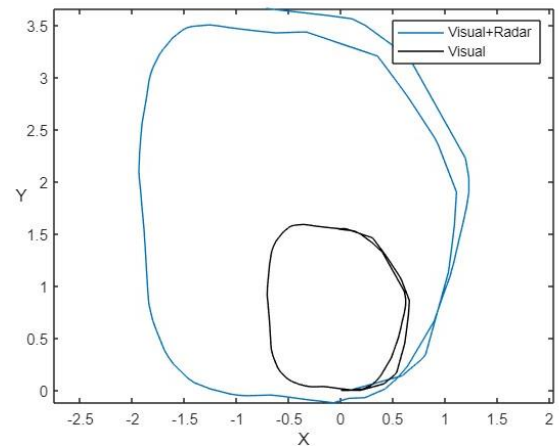


Figure 10. The comparison of trajectory between Radar Aided VO and pure VO in scenario 2 (Conference room)

In Figure 9, pure VO trajectory travels only 5 m. But with the linear velocity estimated by the radar, the travel distance is about 18 m, which is closer to the length of the hallway. In Figure 10, pure VO travel through the area for about 1×1.5 m, and the area expand to 3×3.5 m with the aid of Radar. These result clearly shown the Radar aided VO is capable of recovering the scale of true world for Monocular VO.

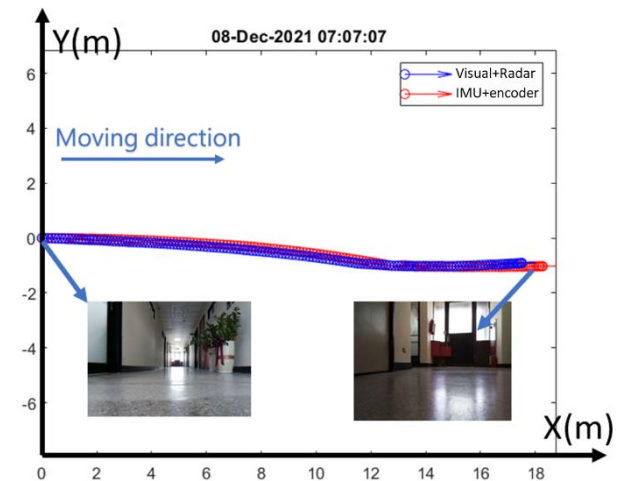


Figure 11. The comparison of trajectory between Visual Radar Odometry and IMU in scenario 1 (Hallway)

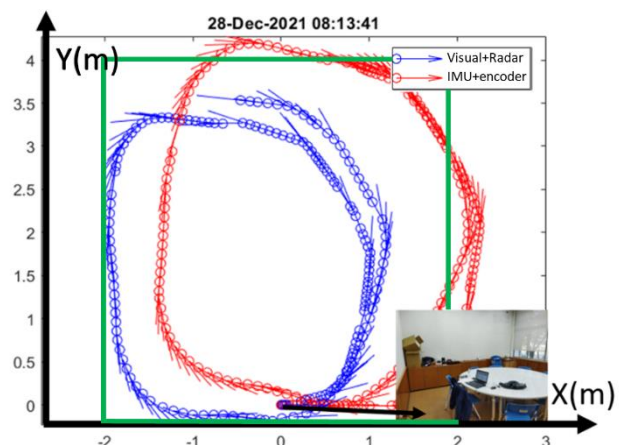


Figure 12. The comparison of trajectory between Visual Radar Odometry and IMU in scenario 2 (Conference room). The Green box indicates the boundaries of the room.

In scenario 1 (Figure 11.), Visual Radar Odometry and IMU+encoder trajectory shows good agreement. However, in scenario 2 (Figure 12.), the trajectory between Visual Radar Odometry and IMU+encoder shows substantial discrepancy due to the drift of IMU and wheel encoder after integral, causing error propagation on the pose estimation. The size of the experiment scenarios was measured using a laser rangefinder. The left boundary of scenario 2 is at $x=2.07$ m, and the top boundary is at 4.07 m, which can be confirmed Visual Radar Odometry shows a closer trajectory to the boundary.

5. CONCLUSION

This study mainly uses the characteristics of low-cost single-chip mmWave radar to apply to mapping and odometry. From the mapping results, there is more noise on the point cloud when using mmWave radar alone. Although this study uses mmWave radar velocity information, reflection intensity information and incident angle to filter the noise, it cannot effectively eliminate the noise, so using mmWave radar alone in high-precision mapping will require further effort. At present, the reason why radar point cloud cannot use lidar related SLAM algorithm is that its accuracy is low and its quantity is scarce, so it cannot describe the texture features in the environment, and use Normal Distribution Transformation (NDT), Iterative Closest Point (ICP) and other algorithms for positioning. At present, the main functions of low-cost single-chip radar are object detection and assisted driving. It is very difficult to achieve positioning and mapping. Limited by the angular resolution of single-chip radar, there are already imaging radars on the market. The main method is to increase the number of receiving antennas to achieve better angular resolution. However, increasing the number of antennas also increases its volume and power. In order to achieve 360-degree scanning, adding a mechanical rotating system is also the design method of high-resolution mmWave radar on the market, such as Navtech's terran360. However, this study mainly focuses on the characteristics of radar with velocity and angle information. Therefore, we choose to integrate the attitude estimation of a monocular camera into the Visual Radar Odometry system architecture, and it is proved that the scale factor of ORB-SLAM can be effectively corrected without great deviation.

REFERENCES

- Mur-Artal, R., Montiel, J.M.M., Tardos, J.D., 2015. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot.* 31, 1147–1163. <https://doi.org/10.1109/TRO.2015.2463671>
- Hong, Z., Petillot, Y., Wang, S., 2020. RadarSLAM: Radar-based large-scale SLAM in all weathers. *IEEE Int. Conf. Intell. Robot. Syst.* 5164–5170. <https://doi.org/10.1109/IRROS45743.2020.9341287>
- Kellner, D., Barjenbruch, M., Klapstein, J., Dickmann, J., Dietmayer, K., 2013. Instantaneous ego-motion estimation using Doppler radar. *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC* 869–874. <https://doi.org/10.1109/ITSC.2013.6728341>
- G. Brooker, D. Johnson, J. Underwood, J. Martinez, and L. Xuan, “Using the polarization of millimeter-wave radar as a navigation aid,” *Journal of Field Robotics*, vol. 32, no. 1, pp. 3–19, 2015.
- Adams, M., Adams, M. D., Mullane, J., & Jose, E., 2012. *Robotic navigation and mapping with radar*. Artech House.
- Almalioglu, Y., Turan, M., Lu, C.X., Trigoni, N., Markham, A., 2021. Milli-RIO: Ego-Motion Estimation with Low-Cost Millimetre-Wave Radar. *IEEE Sens. J.* 21, 3314–3323. <https://doi.org/10.1109/JSEN.2020.3023243>
- M. Jaud, R. Rouveure, L. M. Arvis, P. Faure, and M. Monod, “Boat borne radar mapping versus aerial photogrammetry and mobile laser scanning applied to river gorge monitoring,” *Open Journal of Remote Sensing and Positioning*, vol. 1, no. 1, pp. 48–63, 2014.
- Kellner, D., Barjenbruch, M., Klapstein, J., Dickmann, J., Dietmayer, K., 2014. Instantaneous ego-motion estimation using multiple Doppler radars. *Proc. - IEEE Int. Conf. Robot. Autom.* 1592–1597. <https://doi.org/10.1109/ICRA.2014.6907064>
- Park, Y.S., Shin, Y.S., Kim, J., Kim, A., 2021. 3D ego-motion Estimation Using low-Cost mmWave Radars via Radar Velocity Factor for Pose-Graph SLAM. *IEEE Robot. Autom. Lett.* 6, 7691–7698. <https://doi.org/10.1109/LRA.2021.3099365>
- Stahoviak, C.C., 2019. An Instantaneous 3D Ego-Velocity Measurement Algorithm for Frequency Modulated Continuous Wave (FMCW) Doppler Radar Data.
- Lu, C.X., Saputra, M.R.U., Zhao, P., Almalioglu, Y., DeGusmao, P.P.B., Chen, C., Sun, K., Trigoni, N., Markham, A., 2020. MilliEgo: Single-chip mmWave radar aided egomotion estimation via deep sensor fusion. *SenSys 2020 - Proc. 2020 18th ACM Conf. Embed. Networked Sens. Syst.* 109–122. <https://doi.org/10.1145/3384419.3430776>
- Lu, C.X., Rosa, S., Zhao, P., Wang, B., Chen, C., Stankovic, J.A., Trigoni, N., Markham, A., 2020. See through smoke: Robust indoor mapping with low-cost mmWave radar. *MobiSys 2020 - Proc. 18th Int. Conf. Mob. Syst. Appl. Serv.* 14–27. <https://doi.org/10.1145/3386901.3388945>
- ElNatour, G., Ait-Aider, O., Rouveure, R., Berry, F., Faure, P., 2015. Toward 3D reconstruction of outdoor scenes using an MMW radar and a monocular vision sensor. *Sensors (Switzerland)* 15, 25937–25967. <https://doi.org/10.3390/s151025937>
- Marck, J.W., Mohamoud, A., VdHouwen, E., VanHeijster, R., 2013. Indoor radar SLAM A radar application for vision and GPS denied environments. *Eur. Microw. Week 2013, EuMW 2013 - Conf. Proceedings; EuMC 2013 43rd Eur. Microw. Conf.* 1783–1786. <https://doi.org/10.23919/EuMC.2013.6687024>