



# AGRI-SLAM: a real-time stereo visual SLAM for agricultural environment

Rafiqul Islam<sup>1</sup> · Habibullah Habibullah<sup>1</sup> · Tagor Hossain<sup>1</sup>

Received: 12 March 2022 / Accepted: 10 May 2023 / Published online: 4 July 2023  
© The Author(s) 2023

## Abstract

In this research, we proposed a stereo visual simultaneous localisation and mapping (SLAM) system that efficiently works in agricultural scenarios without compromising the performance and accuracy in contrast to the other state-of-the-art methods. The proposed system is equipped with an image enhancement technique for the ORB point and LSD line features recovery, which enables it to work in broader scenarios and gives extensive spatial information from the low-light and hazy agricultural environment. Firstly, the method has been tested on the standard dataset, i.e., KITTI and EuRoC, to validate the localisation accuracy by comparing it with the other state-of-the-art methods, namely VINS-SLAM, PL-SLAM, and ORB-SLAM2. The experimental results evidence that the proposed method obtains superior localisation and mapping accuracy than the other visual SLAM methods. Secondly, the proposed method is tested on the ROSARIO dataset, our low-light agricultural dataset, and O-HAZE dataset to validate the performance in agricultural environments. In such cases, while other methods fail to operate in such complex agricultural environments, our method successfully operates with high localisation and mapping accuracy.

**Keywords** Stereo visual SLAM · Simultaneous localisation and mapping (SLAM) · Place recognition · Bundle adjustment

## 1 Introduction

In the last two decades, vision-based localisation and mapping has drawn much attention for its broader applications in robotics and computer vision (Zhou et al., 2015; Mur-Artal & Tardós, 2017; Cvisic, 2017; Qin et al., 2018; Bayle et al., 2020; Jiao et al., 2021; Cao & Beltrame, 2021; Liang & Wang, 2021; De Croce et al., 2019; Pire et al., 2020). Visual simultaneous localisation and mapping (VSLAM) is a key concept that emerges for mapping an unknown environment while estimating camera pose in that environment. SLAM technique is specially used in an occluded environment where GPS and any other wireless signals are mostly unavailable. An autonomous agricultural robot navigating in

a shaded orchard can be an example of such a scenario. Many SLAM methods have been proposed over time, and they are mainly dissimilar by the usage of different sensors, i.e., camera, lidar, etc. Mapping an unknown environment using a camera sensor provides rich information about the environment compared to other available sensors, i.e., colour, texture and depth information. Meanwhile, it is a cheap, compact, noise-free, lifelong and sustainable device. Subsequently, a SLAM system deployed with a camera has been a significant interest in robotics and computer vision nowadays.

Visual SLAM methods are mainly subdivided into two categories: Feature-based method (Mur-Artal & Tardós, 2017; Quan et al., 2021) and direct method (Engel et al., 2014). Feature-based methods solely rely on feature extraction both for map point creation and accurate camera pose estimation by minimising reprojection error. On the other hand, the direct method utilises image pixel density for map point creation and camera pose estimation by minimising photometric error (Ma et al., 2019). Among these two methods (Mur-Artal & Tardós, 2017; Engel et al., 2014), the feature-based method (Mur-Artal & Tardós, 2017) is proven to be more robust in illumination changes and geometric errors. The feature-based SLAM methods can produce very compet-

✉ Rafiqul Islam  
md\_rafiqul.islam@mymail.unisa.edu.au

Habibullah Habibullah  
habibullah.habibullah@unisa.edu.au

Tagor Hossain  
md\_tagor.hossain@mymail.unisa.edu.au

<sup>1</sup> UniSA STEM, University of South Australia, Mawson Lakes Blvd, Adelaide, SA 5095, Australia



**Fig. 1** Sample image and detected features of a shaded macadamia orchard using the proposed method

itive accuracy similar to its rival laser range scanner precision when abundant features are extracted. However, those methods very often fail due to inadequate feature extraction from images of a scene. Inadequate feature extraction sometimes causes a loss of tracking of a feature-based SLAM system (Ma et al., 2019). The feature extraction and matching can be reduced by many factors such as underexpose and hazy images (Huang, 2001; Alismail et al., 2017). Nevertheless, some applications, which require semi-dense mapping, can still be benefited from the direct method.

In this paper, several improvements have been proposed for a visual SLAM system—enhancing meaningful correspondences from a pair of images in an extremely changing environment is one of the main contributions. Figure 1 shows meaningful feature extraction using the proposed method. The proposed method is deployed with image enhancement; therefore, it potentially improves the visibility and colour shift caused by the scattered light in the air. Moreover, the image enhancement technique for a visual SLAM system opens up a great possibility that enhances the machine’s capability to recover sufficient information from an uncertain environment. Therefore, localisation and mapping have been potentially improved. On the flip side, most of the visual SLAM algorithms assume that the input images are fine-tuned scene radiance (Mur-Artal & Tardós, 2017; Engel et al., 2014; Quan et al., 2021; Ma et al., 2019), and are mostly evaluated using preprocessed datasets. However, the assumption might not always be practical in real applications as the raw images have to be processed onboard. Thus, the proposed method can be applied to a variety of scenarios that suffers from extremely changing environments.

The primary contributions of this research are as follows:

1. To the best of the author’s knowledge, the proposed method is the first stereo-visual SLAM for highly shaded and GPS-denied agricultural orchards.
2. A local point and line feature recovery technique has been developed using a low-light image enhancement and dehazing method for an agricultural stereo visual SLAM.
3. A low-light agricultural dataset has been recorded for the first time from a Macadamia orchard for the evaluation of visual SLAM methods in difficult and GPS-denied agricultural orchards. The dataset is publicly available at <https://rafiqrana.github.io/low-light-dataset>.

In the rest of the paper, the sections are discussed in the following orders: related work in Sect. 2, discussion of the proposed method in Sect. 3, presenting the experimental result and discussion in Sect. 4 and 5 respectively, and finally ends with a conclusion in Sect. 6.

## 2 Related Work

In recent years, remarkable progress has been achieved in the development of stereo visual SLAM systems (Mur-Artal & Tardós, 2017; Gomez et al., 2017; Cvisic, 2017; Sumikura et al., 2019; Lemaire et al., 2007; Marks et al., 2008; Nalpantidis et al., 2011; Wen et al., 2021). This particular area is becoming promising for real-time operation, and researchers are now striving to make the visual SLAM system adaptable for different uncertainties and atmospheric conditions (Huang & Liu, 2019; Lazaros et al., 2008). For instance, adverse weather conditions, including low atmospheric light, motion blur, and hazy atmosphere, are some relatively new shortcomings in visual SLAM systems (Zhu et al., 2015). As a consequence, visual SLAM system, especially for low-light environment, has recently gained significant research interest. The robustness of the SLAM system is reduced by rapid illumination change in the environment. Automatic camera control is one of the common reasons for the illumination change of an image (Alismail et al., 2017). Unfortunately, a

vision-based system has little hope of tracking recovery if the feature-based pipeline fails due to the illumination change.

While extended Kalman filter-based SLAM was the earlier convention to work in a large working environment, visual-inertial SLAM and pure feature-based visual SLAM systems are becoming more robust and popular nowadays (Mur-Artal & Tardós, 2017; Gomez et al., 2017; Kerl et al., 2013; Zhong & Chirarattananon, 2021; Chen et al., 2019). Visual-inertial SLAM is also a popular research stream nowadays for its cheap and simple hardware requirements. VI-SLAM has now been used in many applications in the field of localisation and mapping, such as autonomous navigation systems, augmented reality and virtual reality (Chen et al., 2018; Liu et al., 2018). The pure feature-based algorithms entirely rely on correspondences to estimate the ego motion of the camera sensor and map the surroundings. Therefore, sufficient features need to be extracted from the environment. However, some algorithms that solely rely on key-points are failed to extract meaningful features from a low-textured environment (Huang & Liu, 2019; Yang & Zhai, 2019; Gomez et al., 2017; Alismail et al., 2017). The authors of PL-SLAM tried to solve the problem by appending line-segment features with point features, which comparatively gives better results compared to ORB-SLAM2 in such a scenario (Mur-Artal & Tardós, 2017; Gomez et al., 2017). Unfortunately, non of these algorithms produce a good result in the low-light and hazy environments. Moreover, insufficient feature extraction leads to a high reprojection error which results in a high ego-motion error and sometimes causes a loss of tracking of visual SLAM systems. In addition to that, insufficient correspondences reduce the place recognition performance of the visual SLAM system while it plays a vital role in the loop detection and loop closure of the system (Milford & Wyeth, 2012; Islam & Habibullah, 2022).

Visual SLAM system with the fine-vocabulary technique was introduced to work in an invariant image lighting condition (Ranganathan et al., 2013; Mikulík et al., 2010; Schubert et al., 2021). The authors focus on light-invariant feature descriptors; in other words, the descriptors of features are expected to be invariant even the illumination of the image of a scene changes. This approach improves the performance of the visual SLAM system to some extent in low-light imaging conditions. However, this method assumes that sufficient features are extracted from the low-light images. But detecting features in a low-light environment is very challenging as most feature detection algorithms are highly sensitive to pixel density value. Therefore, such methods can not compensate for the overall performance of the visual SLAM system. Similarly, A binary feature descriptor has been proposed to tackle such a low-textured environment due to illumination changes in a direct method (Alismail et al., 2017). But the main challenge of a direct method is to cope with a large

number of mismatched feature points efficiently in real-time (Tykkälä & Comport, 2011). Therefore, the current state-of-the-art feature-based method, which is very forthcoming for real-time operation, does not have any effective solutions for changing the imaging environment.

In agricultural robotics and precision agriculture, the vision-based navigation system has been paved with advances in Technology (Rovira-Más et al., 2008; Ball et al., 2016; Paturkar et al., 2017; Chen et al., 2018; Shu et al., 2020; Islam & Habibullah, 2021). In (Paturkar et al., 2017; Chen et al., 2018), the authors investigated potential works of the vision-based systems in the agricultural context and reported challenges and limitations in current progress. It is found that major limiting factors are changing illumination and low-light condition in agricultural scenarios. In another work (Ball et al., 2016), a global positioning system (GPS), an inertial measurement unit (IMU) and a stereo vision system have been used for obstacle detection and navigation of a mobile agricultural robot. The system has been tested for numerous weeks of field trials during the day and night, and it was found from the experiment that the robot was able to operate for only 5 min without any GPS data. Therefore, such systems, that rely on additional sensor information, i.e., GPS, are not operational in the shaded orchard. For solving SLAM in an agricultural olive groves environment, Extended Information Filter (EIF) was proposed (Cheeín et al., 2011; Shu et al., 2021), and the map was constructed by detecting olive stems from a larger range sensor and a monocular camera. The Support Vector Machine (SVM) is used to detect olive stems from the captured image of the environment (Cheeín et al., 2011). In other works (Cheeín & Guivant, 2014; Matsuzaki et al., 2018), A SLAM method with GPS localization is utilised for determining the treetops volume information, and semantic mapping is also introduced to deal with the contextual navigation in greenhouses for agricultural robot navigation (Matsuzaki et al., 2018). SLAM for agricultural applications is proposed by generating a map in the simulated agricultural environment (Habibie et al., 2017), and the map was constructed from grid-based/volumetric using fine-tune SLAM-GMapping algorithm. However, the simulated environment is well constructed and does not change over time; on the flip side, a real agricultural environment is very uncertain where the camera frames and other sensory data fed to the system change instantaneously due to uncertain atmospheric conditions. (Aguiar et al., 2020; Huang & Liu, 2019).

Finally, we found a few analogous works to our proposed method where the authors attempted to solve the flaws of a visual SLAM system for the low-light environment (Huang & Liu, 2019; Yang & Zhai, 2019; Kim et al., 2021). In Huang and Liu (2019), the authors employed a simple histogram equalisation technique that is intended to work as a frontend of the ORB-SLAM2 system for preprocessing low-light input

**Fig. 2** Proposed system architecture

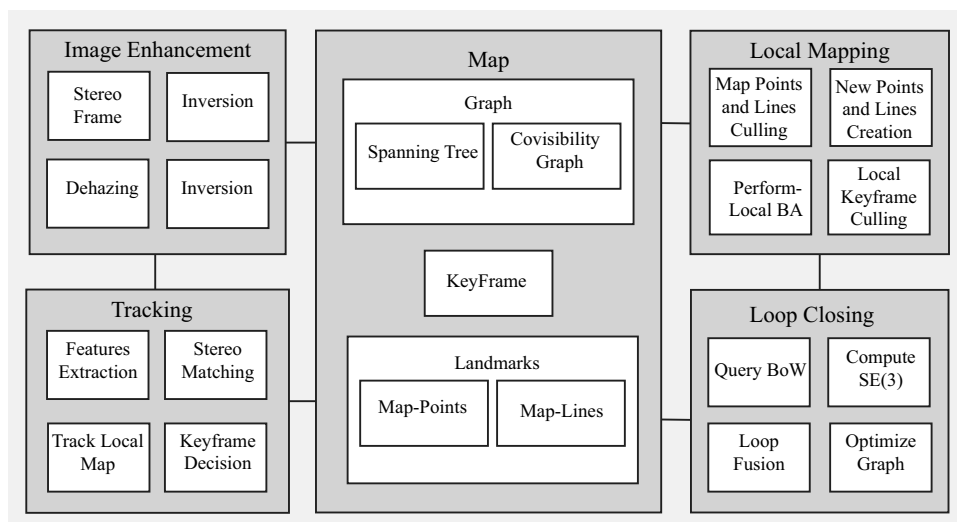


image stream. Although their method reasonably improves image visibility and subsequent feature extraction to some extent, they did not address haze in the image even when it is a common issue found in sudden illumination changes. Moreover, their implementation was only tested on a point feature-based SLAM system, which suffers in a low-textured environment (Gomez et al., 2017). In another method (Kim et al., 2021), the variation of the illumination has been improved by selecting the seed low dynamic range (LDR) image for high dynamic range (HDR) fusion to secure inter-frame consistency. Then they exploited the camera response function (CRF) to synthesise HDR images. However, point and line features based on visual SLAM systems for the low-textured, low-light and hazy environment are still left undeveloped.

### 3 System overview

A structural overview of the proposed SLAM system is illustrated in Fig. 2. This system comprises mainly four subsystems - image enhancement, tracking, local mapping and finally loop closure. The goal of these four subsystems is to construct a 3D map of a stereo camera recorded environment and track a camera location in that environment. The operation of the SLAM system is described here considering a single pair of stereo images. Therefore, the whole process needs to be placed in a loop in order to simultaneously map and localise the sensor in the 3D-constructed map.

#### 3.1 Image enhancement model

Colour inversion, which is also known as the negative effect, has been utilised here to determine haze in low-light images (Dong et al., 2011). Assuming the input is an 8-bit RGB colour image, the colour components can take values rang-

ing from 0 to 255. Therefore, the inversion mapping can be defined as:

$$Q^c(m, n) = 255 - P^c(m, n), \quad (1)$$

where  $c$  is the index of the colour channel,  $(m, n)$  is the index of the pixel in  $m$ -th row and  $n$ -th column,  $P^c(m, n)$  is the pixel intensity of input image  $P$ , and  $Q^c(m, n)$  is the pixel intensity of the inverted image  $Q$ .

In image processing and computer vision, the atmospheric scattering model (Lee et al., 2016; Dong et al., 2011) can be expressed as follows:

$$Q^c(m, n) = J^c(m, n)t(m, n) + A^c(1 - t(m, n)), \quad (2)$$

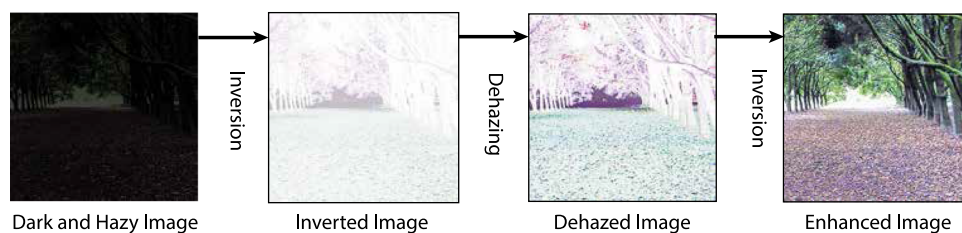
where  $(m, n)$  is the index of the pixel,  $Q(m, n)$  is the observed inverse intensity of the low-light hazy pixel, which is determined in Eq. 1. In this case,  $J(m, n)$  is the radiance of the scene that represents the de-hazed image,  $t(m, n)$  is the medium of transmission that describes the amount of light enters to camera sensor without being scattered, and  $A$  is the light in the atmosphere during image capture and a global constant.  $A$ ,  $P$  and  $J$  are three-dimensional arrays as the input image is in RGB space. The medium of transmission  $t(m, n)$  (Long et al., 2014; Xu et al., 2012; Shuai et al., 2012) for min operation of normalised haze image is determined as:

$$t(m, n) = 1 - \omega \min_{c \in \{r, g, b\}} \left( \min_{y \in \Omega} \frac{Q^c(y)}{A^c} \right), \quad (3)$$

where  $\Omega$  is a local block centred at  $(m, n)$  having a size of 9 in this paper.  $\omega$  is the parameter that tunes the amount of haze to preserve the optimal depth of the scene. The value of  $\omega$  is set to 0.8 in this paper, which produces a good result. The global atmospheric light  $A$  has been calculated by selecting 200 pixels with the lowest intensities in all colour channels.



**Fig. 3** Process of low-light and hazy image enhancement



Among them, a single pixel has been chosen to have the highest value for the sum of all channels. As  $P$  is given,  $J$  can be restored by substituting  $t$  and  $A$  into Equ. 2 as:

$$J^c(m, n) = \frac{Q^c(m, n) - A^c}{t(m, n)} + A^c \quad (4)$$

Once the dehazed image  $J$  is obtained, the enhanced image can be produced by performing an inverse operation as follows:

$$I_{enhanced}^c(m, n) = 255 - J^c(m, n). \quad (5)$$

The process is conceptually shown in Fig. 3, which presents a surprising image enhancement.

## 3.2 Feature tracking

Feature tracking subsystem leverages the enhanced images, where feature extraction, stereo matching and local map tracking operation are actioned. This subsystem mainly deals with visual odometry estimation between consecutive frames and the new keyframe insertion policy.

### 3.2.1 Point and line features

In the proposed method, the well-known Oriented FAST and Rotated BRIEF (ORB) method for point feature and Line-segment Detector (LSD) for line feature have been employed to detect sufficient features from wide imaging scenario (Rublee et al., 2011; Von Gioi et al., 2008). ORB feature has been chosen for its fast feature detection performance while LSD has also been chosen for its superiority to detect features from low-texture environments. Moreover, the line segment gives more meaningful geometrical and spatial information about the environment that might be later used for trajectory planning from the 3D reconstructed map. Both ORB and LSD methods provide the binary nature of the descriptor that allows fast and efficient feature matching so that the binary bag of words (DBoW3) method can utilise the binary descriptor for fast and efficient place recognition. Finally, employing both feature detection methods ensures a high magnitude of correspondences from the image of a scene which results in a precise camera pose estimation in the proposed method.

### 3.2.2 Stereo matching and finding correspondences

For stereo matching and correspondence problems, image rectification is a useful technique that reduced the 2-D stereo correspondence problem into a 1D problem. Image rectification works in such a way that the corresponding points always lie on the same row coordinates, i.e., parallel with the  $x$ -axis and match up between views (Hartley & Zisserman, 2004). Consequently, disparities between the images are in the  $x$ -direction only. We extract ORB and LSD features in both images and for every feature in the left image, we search for a match in the right image. This way, the point and line correspondences are determined. For the stereo matching and finding correspondences, we used open-source implementations from the OpenCV library.

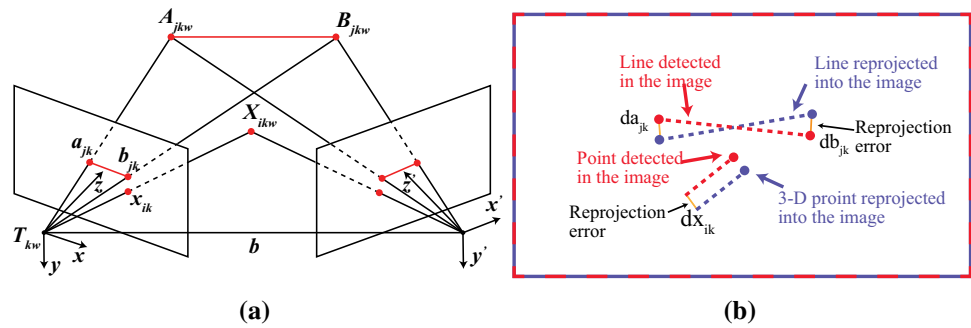
### 3.2.3 Motion estimation

The camera motion estimation is a fundamental problem in the visual SLAM system. The goal of the camera pose estimation is to find the camera pose  $T$  from point and line features. The camera pose consists of 3D rotation (roll, pitch, and yaw) and 3D translation ( $x$ ,  $y$  and  $z$ ) of a total 6D pose in the world coordinate system. To solve the problem, camera intrinsic parameters are determined by calibrating the camera, and an essential matrix has been determined using the corresponding points and lines of stereo image pairs. Once the essential matrix and camera intrinsic from the camera calibration are known, the camera motion has been solved by the perspective-n-point (PnP) method (Lepetit et al., 2008). The resultant camera pose is, therefore, the composition of rotation matrix ( $r$ ) and translation vector ( $t$ ) in homogeneous coordinates.

### 3.2.4 Keyframe selection

Keyframe selection has significant importance for local mapping because it reduces unnecessary and repetitive keyframe insertion into a map. Therefore, a new keyframe is inserted depending on how much the current image is matched the last image. There are several strategies available for keyframe selection. One of the most common strategies is: for every  $n$  frame, at a particular camera pose displacement when the number of matched features in both images is less than a cer-

**Fig. 4** A camera projection model and reprojection error: **a** A stereo camera projection model, and **b** projection error of point and line-segment features



tain threshold, a new keyframe is required to be inserted. In our approach, we employ the strategy for keyframe selection as proposed in Kerl et al. (2013), where user define parameters are reduced in the following ways: The differential entropy of gaussian  $x \sim \mathcal{N}(\mu, \Sigma)$  with  $m$  dimensions is define as:

$$h(x) = 3(1 + \ln(2\pi) + 0.5\ln(|\Sigma|)), \quad (6)$$

where  $\Sigma$  is the scale of the error distribution. Since the entropy changes as the scene of the trajectory changes, i.e.,  $H(x) \propto \ln(|\Sigma|)$ , the uncertainty of the covariance matrix is shifted into a scalar form.

Then, the entropy ratio between motion estimation is as follows:

$$\alpha = \frac{H(T_{k:k+1})}{H(T_{k:k+j})}. \quad (7)$$

where the entropy ratio  $\alpha$  has been calculated from the estimated motion of the last keyframe  $k$  to the current keyframe  $k + j$ . The estimated movement from the last keyframe  $k$  to the immediate next frame  $k + 1$ . The entropy ratio decreases upon the increase in distance between the last keyframe and the current frame. In the proposed approach, we empirically get the best result for  $\alpha = 0.8$ . Finally, a new keyframe is selected when the current frame  $k + j$  lies below the predefined threshold.

### 3.3 Local mapping

The local mapping thread is responsible for performing bundle adjustment when a new keyframe is inserted. Moreover, local mapping is also responsible for optimising the position of the triangulated landmarks from different camera viewpoints. Local bundle adjustment is used to refine the 3D coordinates of inconsistently constructed 3D landmarks in the local map. Therefore, the actions of the subsystem are described as follows:

#### 3.3.1 Keyframe insertion

Every time when a new keyframe is chosen from the tracking thread as described in Sect. 3.2.4, it has to be inserted into the local map. For the keyframe insertion, a similar approach has been used in other V-SLAM methods (Gomez et al., 2017; Mur-Artal & Tardós, 2017; Qin et al., 2018). Estimating the camera poses inevitably results in errors, which accumulate over time, and this effect is called the drift error. The estimation of the relative camera pose between the current and previous keyframes is therefore needed to be refined. This is essentially comprised of performing the bundle adjustment on the local map. Since all the keyframes are connected with the current one by the visibility graph and the landmarks observed by those local keyframes. Therefore, we employ local bundle adjustment to optimise the local map which has been discussed broadly in Sect. 3.3.2.

#### 3.3.2 Local bundle adjustment

Camera pose estimation from one viewpoint to the next always contains some errors due to imprecise key-points and line-segments in the images, incorrect matching in correspondences, and imprecise camera calibration. These errors accumulate as the number of camera views incremented. The bundle adjustment is used to refine the camera poses and 3D landmarks' position in the 3D constructed map. When a new keyframe is inserted in the map, bundle adjustment of the local map is performed. A stereo camera model has been illustrated in Fig. 4a, which has been used to derive the mathematical model of local bundle adjustment as follows:

Let  $T_{kw} \in SE(3)$  be the camera pose, where  $k$  stands for keyframe and  $w$  stands for world coordinate,  $X_{ikw} \in \mathbb{R}^3$  be the position of world 3D point in the  $i$ -th key-points.  $\hat{x}_{ik} \in \mathbb{R}^2$  be the projected 2D key-points into the image plane.  $A_{jkw}, B_{jkw} \in \mathbb{R}^3$  be the world 3D endpoints of the  $j$ -th line. and  $\hat{a}_{jk}, \hat{b}_{jk} \in \mathbb{R}^2$  be their projected 2D line endpoints in the image plane. The detection of  $X_{ikw}$  in the image plane is  $x_{ik} \in \mathbb{R}^2$ ; Similarly, the detections of  $A_{jkw}, B_{jkw}$  in the image plane are  $h_{a_{jk}}, h_{b_{jk}} \in \mathbb{R}^2$ , where  $h$  stands for homogeneous form. Thus, the re-projection error of the line

feature is illustrated in Fig. 4b, where  $d_{a_{jk}}, d_{b_{jk}}$  are the pixel distance from  $a_{jk}, b_{jk}$  to the projected line  $\widehat{a}_{jk}, \widehat{b}_{jk}$ . Let  $\pi$  be the projection function of a stereo camera, and  ${}^h\pi$  is its homogeneous form.

$${}^h\pi = \begin{bmatrix} f_x \frac{X_{kw}}{Z_{kw}} + c_x \\ f_y \frac{Y_{kw}}{Z_{kw}} + c_y \\ f_x \frac{X_{kw}-b}{Z_{kw}} + c_x \end{bmatrix}, \tag{8}$$

where,  $(f_x, f_y)$  is the focal length,  $(c_x, c_y)$  is the principal point of the camera, and  $b$  is the baseline, and all those parameters are known from camera calibration. The reprojection error of key-points and line-segment are respectively  $e_{ik}^p$  and  $e_{jk}^l$ :

$$e_{ik}^p = [x_{ik} - \pi(T_{kw}, X_{ikw})], \tag{9}$$

$$e_{jk}^l = \begin{bmatrix} d_{p_{i,k}} \\ d_{q_{i,k}} \end{bmatrix} = \begin{bmatrix} l_{jk}\pi(T_{kw}, A_{jkw}) \\ l_{jk}\pi(T_{kw}, B_{jkw}) \end{bmatrix}, \tag{10}$$

where  $l_{jk}$  is the infinite line corresponding to the  $j$ -th line in the  $k$ -th keyframe. Combining the points and line segment, the loss function is:

$$L = \sum_{i \in \mathcal{K}} \left[ \sum_{i \in \mathcal{P}} (e_{ik}^p)^T \Sigma_{ik}^{-1} e_{ik}^p + \sum_{j \in \mathcal{L}} (e_{jk}^l)^T \Sigma_{jk}^{-1} e_{jk}^l \right]. \tag{11}$$

where the left and right parts of the loss function are residual functions for the points and lines features, respectively.  $\Sigma_{ik}^{-1}, \Sigma_{jk}^{-1}$  are the inverse covariance matrices of points and lines,  $\mathcal{K}, \mathcal{P}, \mathcal{L}$  refers to the groups of a local keyframe, points and lines, respectively.

Minimising the loss function  $L$ , the camera pose, as well as the position of the point and line landmarks, gets their optimal values. However, the optimisation problem requires the derived Jacobian matrices. The Jacobians for the reprojection error of points  $e_{ik}^p$  and line  $e_{jk}^l$  are already deduced in Zhou et al. (2015); Zuo et al. (2017); Mur-Artal and Tardós (2017). Once all the Jacobians about the loss function are known, the optimisation of camera pose, point and line landmarks is solved using the Levenberg-Marquardt method implemented in the open-source library  $g^2o$  (Kümmerle et al., 2011).

### 3.4 Loop closing

The loop closure subsystem is assigned to recognise a loop based on whether or not the sensor has already visited the current scene. It is responsible for triggering a pose graph optimisation (PGO) method to optimise the cumulative error being distributed along the recognised loop (Fuentes-Pacheco et al., 2015). In the case of tracking failure or relocalisation in an already mapped scene, loop closure plays

an essential role. This subsystem includes two steps, as follows:

#### 3.4.1 Loop closure detection

Loop detection simply involves searching a potential amount of features of a currently being processed scene to the previously visited scene in the database. If the search returns the desired threshold, the loop closure is considered to be detected. However, a good loop closure detection technique expects no false-positive and least false-negative loop detection, which is the fundamental requirement of a good loop closure algorithm (Fuentes-Pacheco et al., 2015). The loop is accurately detected by employing a newly developed bag of words (DBoW3) approach (Muñoz-Salinas & Medina-Carnicer, 2020), which is an improved version of the DBoW2 (Galvez-Lopez & Tardos, 2012). For the binary nature of DBoW3, it is comparatively fast for such a problem. It was initially developed for BRIEF binary descriptor only; however, it was later proposed to work with LSD as well (Gomez et al., 2017). Therefore, at each time step, both ORB and LSD features are compared and saved parallelly for accurate loop closure detection. Searching for both images is motivating as some scenes might be described more precisely with segments compare to key-points or vice versa.

#### 3.4.2 Global bundle adjustment

The global bundle adjustment is the specific case of local bundle adjustment, where all keyframes and landmarks are optimised. The process optimises the overall map by adjusting the distributed drift while fusing both sides of the loop. This problem is well known as PGO, where all the nodes are considered to be the keyframes within the detected loop and edges are produced by the essential graph and spanning tree. In the proposed method, a full bundle adjustment is incorporated after a pose graph to achieve an optimal solution. The PGO problem is solved by employing an open-source implementation of  $g^2o$  library (Kümmerle et al., 2011). The entire trajectory is refined after the PGO operation is completed.

## 4 Experimental validation

The proposed method’s performance has been evaluated using standard datasets and evaluation metrics. The evaluation metrics include relative pose error (RPE) and absolute pose error (APE) of a visual SLAM system that is accumulated over the distance travelled by the camera sensor. The RPE is a standard metric for evaluating the local consistency of a SLAM trajectory. It compares the relative poses along with the estimated and the reference trajectories (ground truth). This is based on the delta pose difference between

**Table 1** Properties of different datasets used for the experimental validation

Datasets	KITTI	EuROC	Low light agri. dataset	O-haze
Cameras	Stereo gray 2 × ~1242 × 376 (global shutter)	Stereo gray 2 × 752 × 480 @20Hz (global shutter)	Stereo RGB 2 × 1920 × 1080 @30 Hz (electronic synchronized rolling shutter)	Mono RGB 5456 × 3632
IMUs	OXTS RT3003 IMU & GPS, 6 axis, @100 Hz	ADIS16488 3-axis acc/gyro @200Hz	–	–
Environment	Outdoor/street	Indoor	Outdoor/agricultural	Outdoor

them. The detailed formulations of the evaluation metrics have been included in Appendix B. The formulations for these evaluation metrics were implemented in EVO evaluation tool (Grupp, 2017). Therefore, this tool has been used to evaluate all those metrics. The proposed method has been experimented upon four different datasets and compared with three different state-of-the-art visual SLAM methods to validate the performance in different environmental conditions. In the experiments, the feature selection threshold has been set to 3k for each of the compared systems. Firstly, the results of the proposed method are compared with VINS-Fusion (Qin et al., 2019), PL-SLAM (Gomez et al., 2017) and ORB-SLAM2 (Mur-Artal & Tardós, 2017) by experimenting with their open-source software in KITTI (Geiger et al., 2013) and EuRoC Micro Aerial Vehicle (MAV) datasets (Burri et al., 2016). Secondly, the proposed method has been experimented on ROSARIO (Pire et al., 2019) agricultural dataset. Thirdly, the proposed method has been experimented on our low-light agriculture dataset of a macadamia orchard and compared with other methods to test the tracking loss or success in such a low-light condition. Finally, the performance of the proposed method has been tested on the outdoor hazy (O-HAZE) dataset (Ancuti et al., 2018) to check the operability of the proposed method in hazy conditions. The properties of different datasets used for the experimental validation have been included in Table 1

Additionally, the application runtime has also been evaluated for the real-time operation of the system by changing different parameters and settings. All the experiments have been computed on a DELL Latitude 5490 having Intel Core i5-8250U quad-core 3.38GHz, Intel HD Graphics 620 and 16GB DDR4 RAM (2400MHz). Each sequence of the datasets was run five times because of the non-deterministic nature of the multithreading system (Mur-Artal & Tardós, 2017). Finally, we median the results for an accurate evaluation of the estimated trajectories from the proposed visual SLAM system. In the following four subsections, the experimental results have been presented for the KITTI dataset,

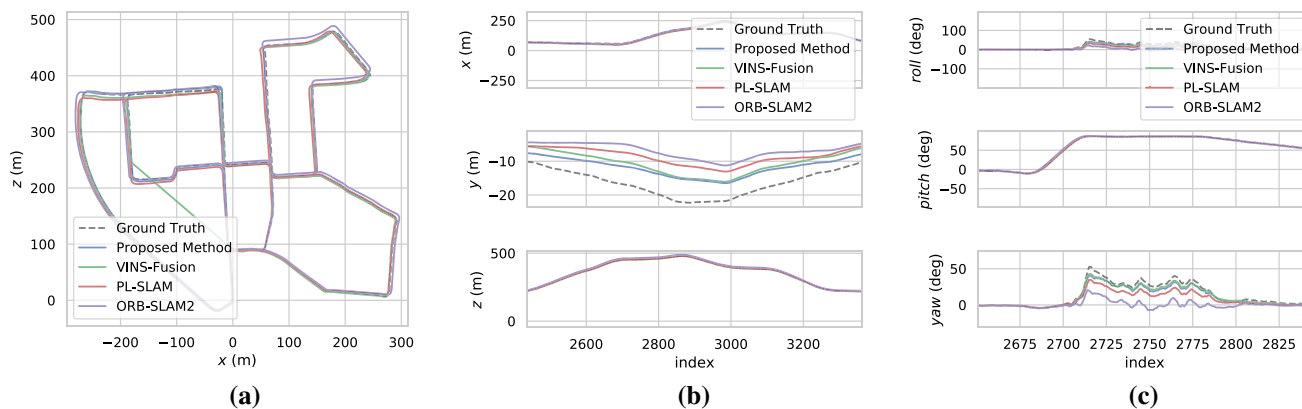
EuRoC MAV dataset, Low-light Agricultural dataset of a Macadamia Orchard, and O-Haze Dataset, respectively.

#### 4.1 KITTI dataset

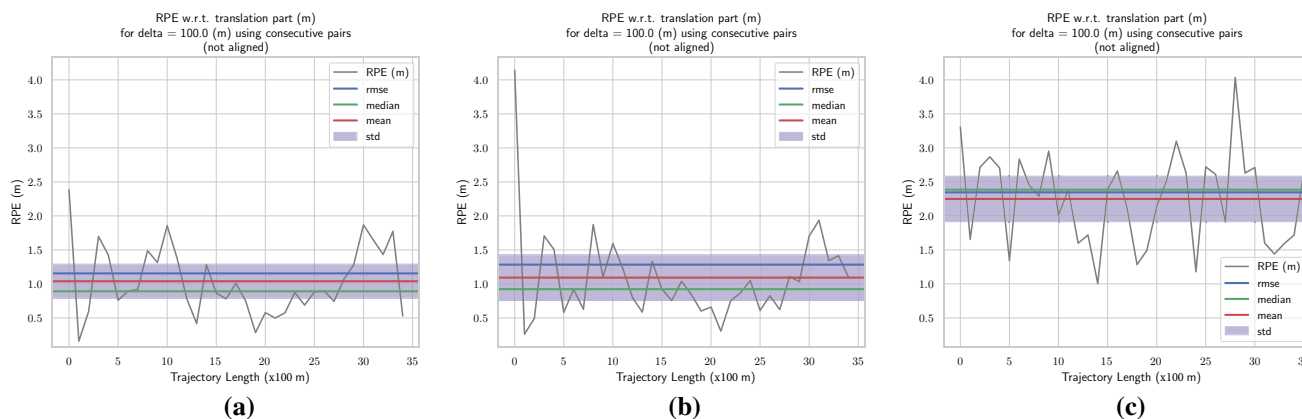
The KITTI dataset comprises high-resolution colour and grayscale stereo sequences filmed from a standard station wagon around Karlsruhe, in motorways and rural areas. A GPS and a Velodyne laser scanner were employed to ensure accurate ground truth of the sequences. Some of the sequences contain underexposed scenes in the pathway, and the proposed method applies image enhancement on those frames to recover enough features from the low-light area. Many sequences such as 00, 02, 05, 06, 07, and 09 contain loops; please refer to Table 2 to see the KITTI sequences that have been used in the experiment. Thus, the proposed method performs place recognition and bundle adjustment on these sequences to reduce the localisation and mapping error.

The estimated trajectories of the proposed method and other methods are illustrated in Fig. 5 for KITTI 00 sequence. In Fig. 5a, the trajectories are shown in the  $xz$  camera axis. The KITTI 00 sequence has a sufficient loop and close features and was recorded in a highly textured environment. In such an environment, both keypoint-based and line-segment-based approaches perform very similarly. For example, almost all tested methods produce similar results in the KITTI 00 sequence, and yet the proposed method yields a slightly superior result in trajectory estimation than the other visual SLAM methods. Additionally, each component of translation errors ( $x$ ,  $y$ ,  $z$ ) are individually shown in Fig. 5b that gives a thorough overview of the error rate in each frame index along the pathways. Only a rapid variation in the  $y$  component is found for all tested methods. Here the proposed method was too close to the ground truth trajectory. Similarly, the rotation part ( $roll$ ,  $pitch$ ,  $yaw$ ) of the RPE is shown in Fig. 5c. It can be seen that there is almost no variation in the estimation of the pitch component of the ( $roll$ ,  $pitch$ ,  $yaw$ ) because a ground vehicle usually





**Fig. 5** Estimated trajectory and ground truth comparison, where, **a** The accuracy of the estimated trajectory from the top view, and **b** The accuracy of  $XYZ$  coordinates, respectively, with respect to distance travelled. **c** The accuracy of the roll, pitch and yaw relative to distance travelled



**Fig. 6** RPE of the proposed method, PL-SLAM and ORB-SLAM2 on KITTI 00. **a–c** are illustrating the RPE, RMSE, median, and standard deviation respectively for the proposed method, PL-SLAM and ORB-

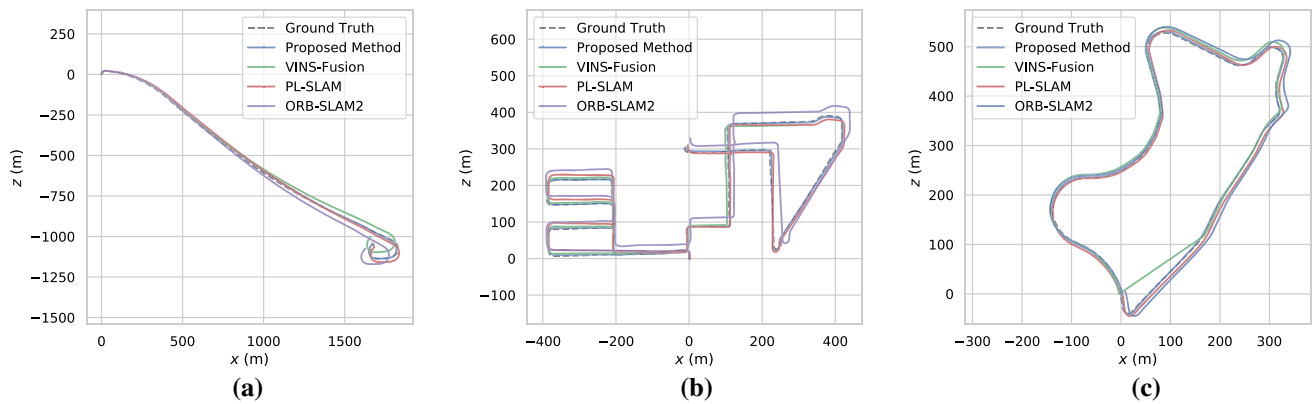
SLAM2. Note that, the RPE in the  $y$ -axis has been updated for every 100 m of camera travel up to 3500 m ( $35 \times 100$  m) of total trajectory length

faces little pitch variation on the plain surface, and the KITTI dataset was captured in similar terrain.

The relative pose error (RPE) is presented in different statistical forms in Fig. 6 that shows the root mean square error (RMSE), median, mean and standard deviation (SD) of the RPE among the proposed method and other methods in KITTI-00 sequence. It is illustrated in Fig. 6 how much the proposed methods and other methods disagree with the ground truth data. For the evaluation of pose estimation error, RMSE is considered to be a standard evaluation metric among all other forms. The lower the value of RMSE, the more it is accurate in pose estimation, see Fig. 6 for the observation. From the evaluation, it is found that the proposed method reduces around 3% RMSE of the translation part in the KITTI-00 sequence. Figure 6a, 6b and 6c show RPE of the proposed method, PL-SLAM, and ORB-SLAM2 respectively. The RPE for the VINS-Fusion method was not illustrated as there was a tracking loss during the test. The

standard deviation represents how the estimated pose error is dispersed from its mean value along with the range of frames.

The comparisons of estimated trajectories by the proposed method and the other methods have been shown in Fig. 7 for KITTI 01, 08, and 09 sequences. The keypoint-based method, i.e., ORB-SLAM2, is very error-prone when no loop closure is detected. The errors are accumulated until any loop closures are detected, which finally results in a large estimation error, see Fig. 7a around  $x,z$  (1700,-1200), and Fig. 7b around  $x,z$  (400,400). Such keypoint-based methods are very fragile in finding sufficient correspondences in a sharp turning around, and they start to deviate from the actual path, see Fig. 7a approximately between  $x,z$  (900,-500), and  $x,z$  (1700,-1200). Similarly, in and Fig. 7b approximately between  $x,z$  (100,350), and  $x,z$  (400,400). However, fusing inertial measurement unit (IMU) data with such a keypoint-based method reduces rotation error in such sharp turning around, i.e., VINS-Fusion (Qin et al., 2019) utilises the IMU data and produces a better result in such scenario, shown in



**Fig. 7** Comparison of the estimated trajectories of the proposed method, VINS-Fusion, PL-SLAM, and ORB-SLAM2 with ground truth in KITTI a 01, b 08 and c 09 sequences

**Table 2** The accuracy of the proposed method is compared with VINS-fusion, PL-SLAM and ORB-SLAM2 in the KITTI Dataset

Seq.	Proposed method		VINS-fusion		PL-SLAM		ORB-SLAM2	
	$t_{rel}$	$r_{rel}$	$t_{rel}$	$r_{rel}$	$t_{rel}$	$r_{rel}$	$t_{rel}$	$r_{rel}$
00	<b>1.154</b>	<b>0.707</b>	–	–	1.456	0.803	2.346	1.158
01	<b>1.601</b>	0.403	1.706	<b>0.375</b>	2.633	2.075	1.786	0.442
02	1.142	0.536	1.124	<b>0.491</b>	<b>1.121</b>	0.618	1.129	0.517
03	1.932	<b>0.416</b>	1.950	0.6548	2.123	1.251	<b>1.627</b>	0.708
04	<b>1.742</b>	0.201	1.793	<b>0.146</b>	1.842	0.264	1.798	0.152
05	<b>0.017</b>	<b>0.001</b>	0.025	0.003	0.020	0.001	0.027	0.003
06	1.227	0.697	<b>0.529</b>	0.396	1.082	0.475	0.595	<b>0.395</b>
07	<b>0.634</b>	0.397	1.086	<b>0.274</b>	1.193	0.537	1.678	0.440
08	<b>1.318</b>	<b>0.678</b>	–	–	1.798	1.812	2.277	0.826
09	<b>1.088</b>	0.628	–	–	1.120	<b>0.507</b>	1.224	0.521
10	<b>0.880</b>	<b>0.479</b>	1.426	0.853	1.054	0.654	1.703	1.047

The RMSE of translation and rotation parts of relative pose error (RPE) is presented, where the translation errors are expressed in (m/100 m) and the rotation errors are expressed in (deg/100 m). Dash (–) indicates that there was a tracking loss during the experiment

Bold values highlight the method with the lowest error

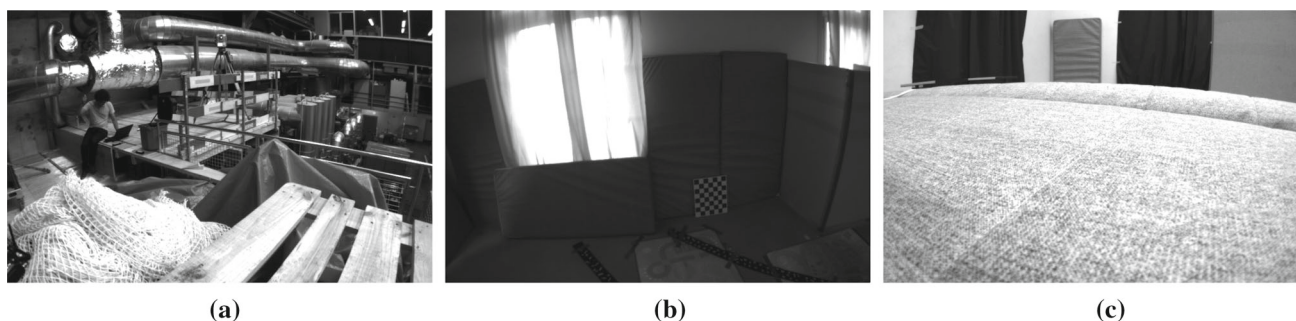
Fig. 7. But VINS-Fusion is unable to reduce overall translation error, which has been discussed in the subsequent paragraph using Table 2. On the other hand, the proposed system, which is a key-point and line segment-based method equipped with a low-light image enhancement technique, does not suffer from the number of correspondences in the relatively low-textured environment. Therefore, it produces good ego-motion estimation without any additional sensors.

Illustrating the accuracy of a system with figures' information only might not be enough for precise evaluation because of the subtle variation in results among different methods. Additionally, the scaling of a figure is limited by paperwork. Therefore, the RMSE has been presented in Table 2 to summarise the overall performance of the proposed method and other state-of-the-art methods in KITTI 00 - 10 sequences. In Table 2,  $t_{rel}$  = translation part of RMSE error in meter per 100 m camera movement, and  $r_{rel}$  = rotation part of RMSE error in degrees per 100-meter camera movement.

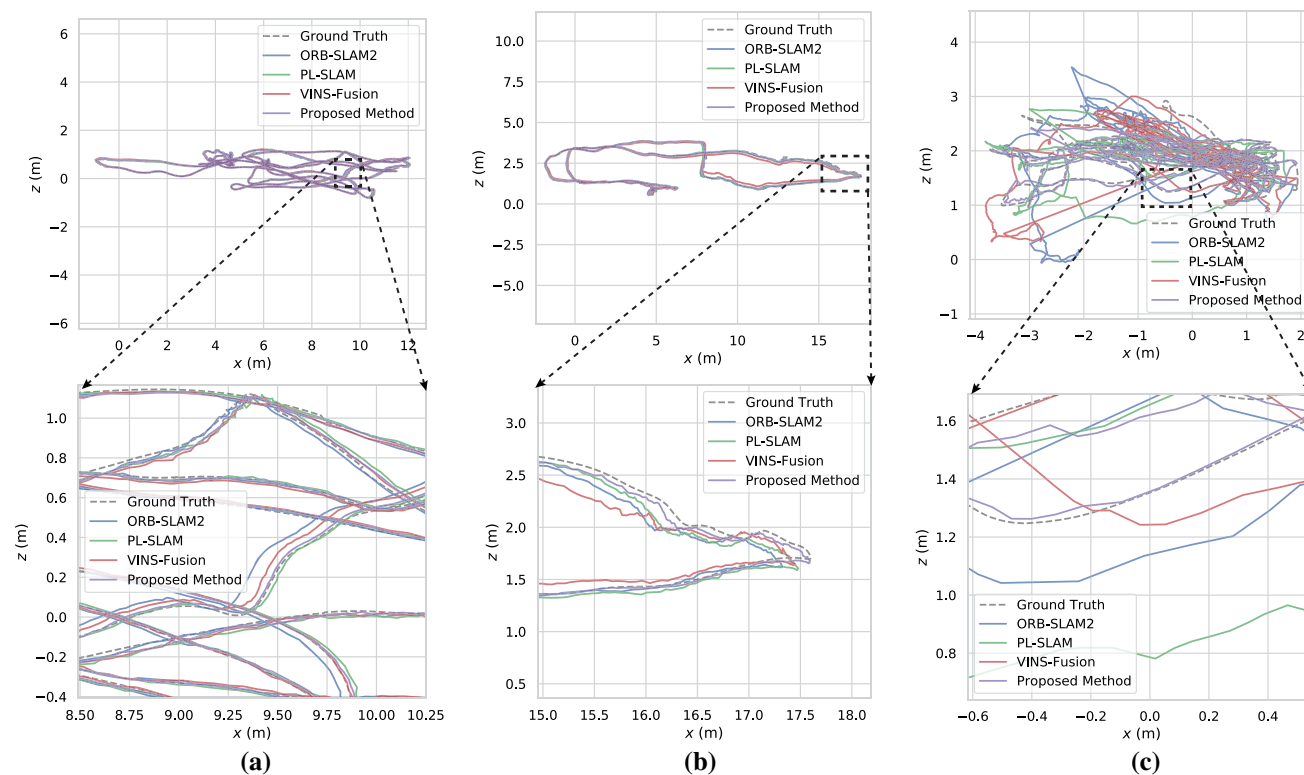
The highlighted values represent the lowest error among all the experiments done on the sequence. It is noticeable that the proposed method achieves higher accuracy (low  $t_{rel}$  and  $r_{rel}$ ) in most of the sequences. In other words, the proposed method outperforms other methods by reducing error in most of the sequences (e.g., on 00, 05, 08, 10).

## 4.2 EuRoC dataset

The EuRoC dataset (Burri et al., 2016) offered eleven stereo colour sequences filmed by a Micro Aerial Vehicle (MAV) which was flying in two large rooms and a big industrial setting. The sequences are categorised based on different challenges - easy, medium and difficult based on the MAV linear velocity, orientation, angular velocity, illumination and scene texture. The difficulty level of the scene texture and illumination has been illustrated by the sample frames from the EuRoC dataset as shown in Fig. 8, and the orientation



**Fig. 8** Sample frames of the EuRoC dataset that represent the difficulty level of the sequences **a** EuRoC MH-01-Easy, having sufficient texture, **b** EuRoC V1-03-Difficult, having low-light image, **c** EuRoC V2-03- Very Difficult, having inadequate texture information



**Fig. 9** Trajectories of **a** EuRoC MH-03-medium, **b** EuRoC MH-04 difficult, **c** EuRoC V2-03-very-difficult

and camera movement have been illustrated by the ground truth trajectories as shown in Fig. 9. Since the MAV revisits the same place several times, the proposed SLAM reuses the existing map and closes the loops when it detects. Moreover, the proposed method successfully tracks the camera pose with high accuracy in the dark frames of the sequences.

A comparison of trajectories of the proposed method and the other state-of-the-art methods is shown in Fig. 9. Each sub-figure is presented in a top to bottom order, where the bottom one represents the scaled version of the top one for showing subtle differences in the trajectory estimation. If we notice the trajectories in Fig. 9a and 9c, it can be seen that the estimated trajectories of the proposed method are relatively better than the other methods. However, The differences are

pretty close because those sequences are comparatively well-exposed and have less camera motion. The accuracy of VINS-Fusion (Qin et al., 2019) and ORB-SLAM2 (Mur-Artal & Tardós, 2017) is abruptly reduced if the texture of the images is low, and the camera motion is high, i.e., the EuRoC v2-03-difficult sequence shown in Fig. 8c. PL-SLAM (Gomez et al., 2017) on the other hand, work slightly better in low-light and low-texture environment. But in every condition, the proposed method produces the best results.

Finally, a summary of the relative RMSE of the translation part has been presented in Table 3. The highlighted values represent better accuracy and lower values. PL-SLAM achieves better accuracy in two sequences, and VINS-Fusion in one sequence. On the other hand, the proposed method

**Table 3** The comparison of the accuracy of the translation error of the proposed method, VINS-Fusion, PL-SLAM and ORB-SLAM2 on the EuRoC dataset

Seq.	Proposed SLAM $t_{rel}$	VINS-fusion $t_{rel}$	PL-SLAM $t_{rel}$	ORB-SLAM2 $t_{rel}$
MH-01-easy	<b>0.9014</b>	0.9020	0.9246	0.9036
MH-02-easy	0.8404	<b>0.8017</b>	0.8839	0.8389
MH-03-med	<b>1.1409</b>	1.1469	1.1425	1.1521
MH-04-diff	<b>1.1030</b>	1.1840	1.1853	1.2735
MH-05-diff	<b>1.0148</b>	1.1075	1.3265	1.1322
V1-01-easy	0.9232	0.9013	<b>0.8835</b>	0.9080
V1-02-med	<b>0.9873</b>	0.9899	0.9959	0.9993
V1-03-diff	<b>0.8446</b>	–	0.9146	0.8912
V2-01-easy	<b>0.7221</b>	0.7822	0.7834	0.7657
V2-02-med	0.9030	0.9070	<b>0.8984</b>	0.9091
V2-03-diff	<b>0.9838</b>	–	1.0943	1.1211

The root mean square error (RMSE) of the translation part is presented, where Delta = 1 m. Dash (–) indicates that there was a tracking loss during the experiment  
Bold values highlight the method with the lowest error

achieved higher accuracy in most of the sequences, especially in the EuRoC medium and difficult sequences. VINS-Fusion was also found to be in a state of tracking loss in a few sequences, i.e., V1-03-difficult and V2-03-difficult. Therefore, it can be concluded that the proposed method succeeded in achieving superior trajectory estimation in more robust conditions.

### 4.3 Rosario dataset

The ROSARIO (Pire et al., 2019) is a multisensory dataset recorded using a weeding robot in an agricultural environment. It offers six sequences from soybean fields, and it contains time-synchronised sensory from a wheel encoder, IMU, stereo cameras and GPS-RTK. The stereo frames cover repetitive scenes, reflection, and rough terrain; and thus put the visual SLAM in a challenge.

Figure 10 shows the comparison of the estimated trajectories for the proposed method and baseline methods in the ROSARIO dataset. The trajectories are shown in the world coordinate system similar to the ROSARIO ground truth. It can be seen in Fig. 10 that the tested visual SLAM methods produced relatively poor RMSE in the ROSARIO dataset than in the KITTI and EuRoC datasets even for the small distance dataset. It is because the ROSARIO contains high repetitive patterns in the agricultural environment than KITTI and EuRoC which are recorded in mostly structural environments. PL-SLAM incorporated with the proposed image enhancement relatively works better than the original PL-SLAM, ORB-SLAM2 and VINS-Fusion in the ROSARIO dataset, and even it achieved better accuracy in sequence-04 than the proposed method with a small difference. It is also noticeable from the subfigures (a), (b) and (e), that the baseline methods are error-prone at the sharp turning points.

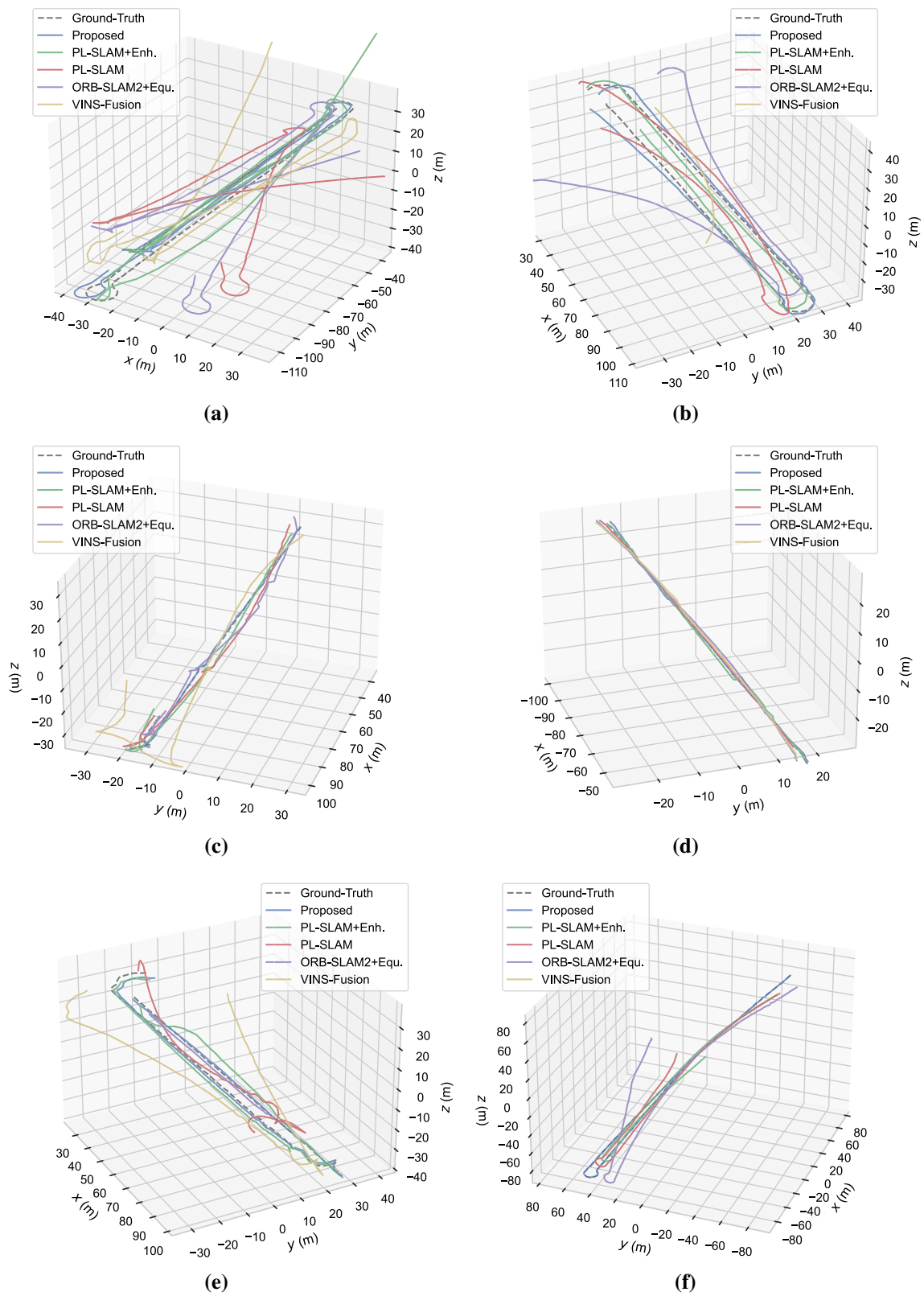
However, the proposed method produces excellent trajectory estimation with the lowest RMSE in such difficult conditions.

Table 4 shows the relative pose error (RPE) of the proposed method and the baseline V-SLAM methods in the ROSARIO dataset. The measurement in bold text is the lowest RMSE which is desirable. The RPE is presented for the translation part only as the ROSARIO dataset doesn't provide a rotational part in the ground truth data. From the presented RMSE, It can be observable that point feature-based methods such as the ORB-SLAM2 and the VINS-Fusion suffer from tracking loss in such an agricultural environment. The ORB-SLAM2 with image equalization method produces better results than the only point feature-based method like VINS-Fusion. However, in sequence-05, the ORB-SLAM2 incorporated with the image equalization method fails to track the complete trajectory. PL-SLAM and PL-SLAM+Enh. have also been tested to investigate the performance compared to the proposed method. It was found that PL-SLAM+Enh. produces slightly better results than only with PL-SLAM but still suffers in the turning points causing deviation from the actual trajectories, which can be noticed from the Fig. 10 and Table 4. On the other side, the proposed method produces superior results than the baseline methods in most of the sequences, employing image enhancement, haze removal, and point and line features.

### 4.4 Low-light agricultural dataset of a Macadamia Orchard

The low-light dataset of a Macadamia Orchard was recorded on a very sunny day in McLeans Ridges NSW 2480, Australia, and yet the sequence was underexposed due to the shade of the trees, as shown in Fig. 11b. There is no loop to the sequence which made the sequence a little bit challeng-





**Fig. 10** Trajectories of the proposed methods and baseline methods on ROSARIO dataset. **a** Sequence-01, **b** Sequence-02, **c** Sequence-03, **d** Sequence-04, **e** Sequence-05, and **f** Sequence-06

**Table 4** The comparison of the translation accuracy of the proposed method, and baseline methods on the ROSARIO dataset

Seq.	Proposed $t_{rel}$	PL+Enh $t_{rel}$	PL-SLAM $t_{rel}$	ORB+Equ $t_{rel}$	VINS $t_{rel}$
Sequence 01	<b>0.3759</b>	1.1053	1.3785	1.3641	2.5758
Sequence 02	<b>0.4154</b>	0.8530	1.1550	2.4524	–
Sequence 03	<b>0.1456</b>	0.7562	1.1332	1.1253	2.0800
Sequence 04	0.8566	<b>0.8491</b>	1.2765	1.3893	2.3645
Sequence 05	<b>1.0139</b>	1.6834	1.9167	–	3.3812
Sequence 06	<b>0.0162</b>	0.5486	0.9251	1.0179	–

The root means square error (RMSE) of the translation part is presented, where Delta =1 frame. Dash (–) indicates that there was a tracking loss during the experiment  
Bold values highlight the method with the lowest error

ing. A TORO Timecutter Zero Turn Mower has been used as the main framework to record the sequence. The mower was retrofitted with a ZED camera from the STEREO LABS, NVIDIA Jetson computational unit and other required hardware as shown in Fig. 11a.

Figure 12 illustrates the constructed map and the estimated trajectory of the proposed method on the low-light sequence of a macadamia orchard. A 3D colour map has been constructed to illustrate the spatial structure of the orchard. In the experimental setting, the threshold was set to a short range, which is why the 3D map from the next rows of trees did not appear on the map. However, a relatively more area was constructed when the camera direction was changed during the turning of the mower at the end of the sequence.

During the attempt of capturing GPS data for ground truth, the GPS signal was almost unreachable and was completely inconsistent even when it was available for a short time. The inconsistent measurement sometimes located the vehicle outside the orchard. Therefore, it was not possible to provide ground truth data for RMSE measurement. However, a simple measure called tracking loss/success has been taken into consideration for the evaluation of different tested methods. A method is considered to be more successful when it does not fail to track the complete trajectory. The (%) loss/success has been measured based on the portion of the successful tracking distance of the complete trajectory. It can be seen from Fig. 12d that PL-SLAM comparatively tracks better than the ORB-SLAM2 and VINS-Fusion in such low-light conditions. PL-SLAM failed after approximately 59.2% of the sequence length, whereas the ORB-SLAM2 and VINS-Fusion failed after 26.5% and 28.3% of the sequence length. In low-light conditions, point and line feature-based methods, i.e., the proposed method and PL-SLAM, work better because they extract more features compared to point feature-based methods, i.e., ORB-SLAM2, VINS-Fusion, in a low-light environment. Based on this measurement, only the proposed method successfully tracked the complete trajectory and the baseline methods failed in such a difficult environment.

**Table 5** The comparison of tracking loss/success rate of the proposed method and other state-of-the-art methods in a shaded macadamia orchard sequence

Methods	Tracking status	% Success (Approx.)
Proposed method	Success	100
VINS-fusion	Lost	28.3
PL-SLAM	Lost	59.2
ORB-SLAM2	Lost	26.5

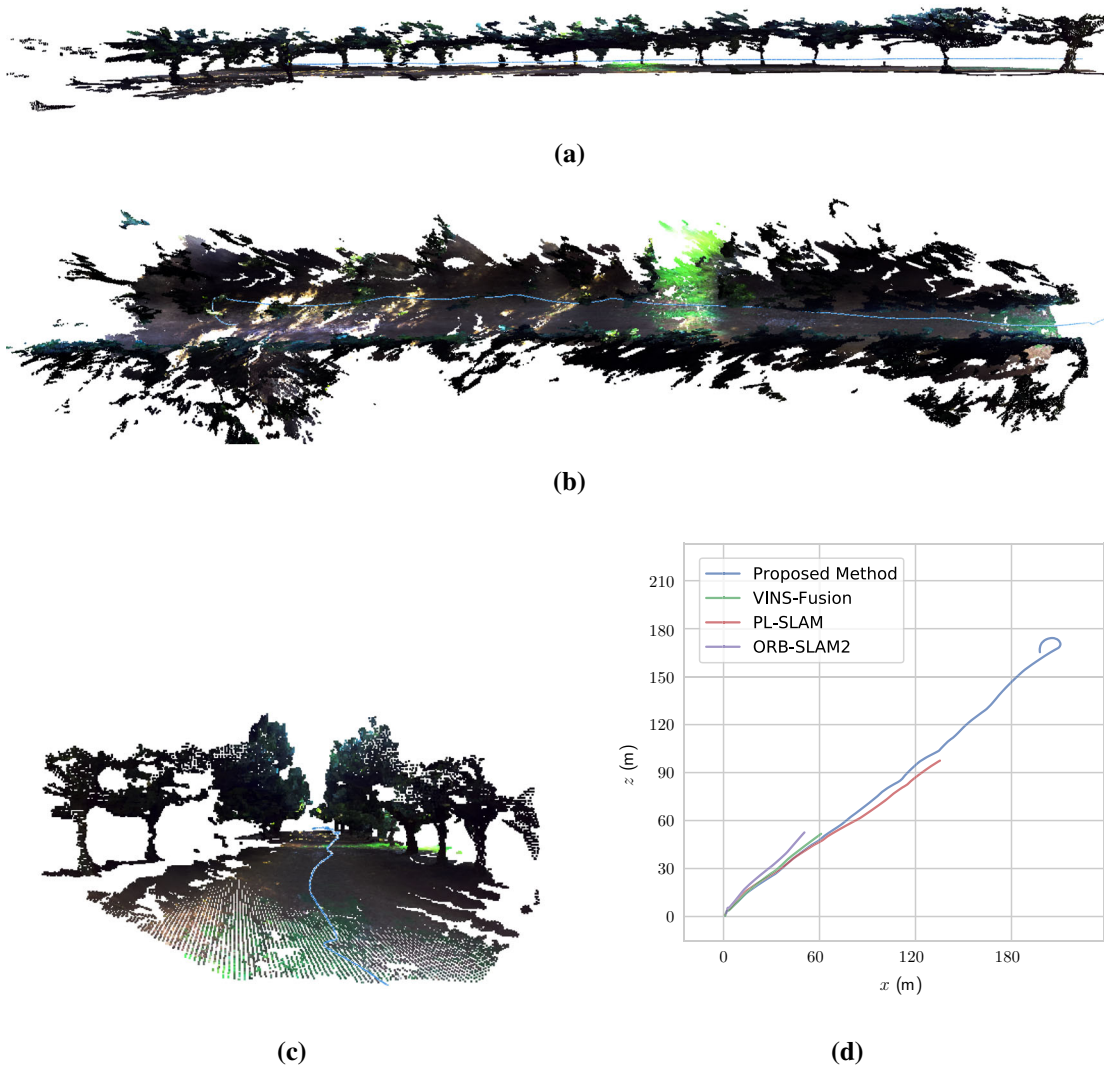
Finally, the tracking loss and success have been summarised in Table 5. It can be seen from Table 5 that the proposed method successfully tracked 100% of the trajectory while VINS-Fusion, PL-SLAM and ORB-SLAM2 partially succeeded in map tracking. Therefore, it can be summarised that point and line features-based methods might be promising in a well-lit environment, but the proposed method achieves better trajectory tracking results in such a low-light environment.

#### 4.5 O-Haze dataset

The O-HAZE dataset (Ancuti et al., 2018) was captured in an outdoor environment. It contains pairs of real hazy and corresponding haze-free ground truth images. The hazy images were captured in the presence of real haze, generated by a professional haze machine producing high-fidelity real hazy conditions. O-HAZE dataset contains 45 different outdoor scenes, and some of them are very similar to agricultural environments. The experimental results of the proposed method and a state-of-the-art method have been illustrated in Fig. 13, where the 1<sup>st</sup>-column shows the hazy images, 2<sup>nd</sup>-column shows the results of Cai et al. (2016), 3<sup>rd</sup> column represents the result of the proposed method, and finally 4<sup>th</sup>-column shows ground truth images. From Fig. 13, it can be seen that the proposed method substantially reduces the haze of the hazy images and more close to the ground truth images.



**Fig. 11** Experimental setup and sequence properties. **a** A retrofitted lawn-mower equipped with a ZED camera and Jetson computational unit, and **b** the image properties of the low light sequence



**Fig. 12** Estimated trajectory and map of the proposed method in macadamia orchard, **a** is a side view of the map and trajectory, **b** is the top view, **c** is the front view, and finally, **d** shows the comparison of the estimated trajectory of the proposed method and other state-of-the-art methods from the top view





**Fig. 13** Comparative dehazing results on O-Haze dataset (Ancuti et al., 2018). Where column **a** is hazy images, **b** is dehazed images of Cai et al. (2016), **c** is dehazed image by the proposed method, and finally **d** is the ground truths

**Table 6** Comparison of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) for dehazed and hazy images against reference images

Sets of images	PSNR (higher is better)	SSIM (higher is better)
Img 01 [Our]	19.5404	0.6800
Img 01 [Cai et al.]	15.6605	0.4830
Img 01 Hazy	14.9922	0.4113
Img 02 [Our]	17.2985	0.6061
Img 02 [Cai et al.]	12.5887	0.3903
Img 02 Hazy	11.6397	0.3012
Img 03 [Our]	27.1873	0.7823
Img 03 [Cai et al.]	13.0333	0.3927
Img 03 Hazy	11.6830	0.2888



Table 6 shows the calculated Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) values for a set of dehazed images against the corresponding reference images. The PSNR and SSIM values are common metrics used to evaluate the quality of hazy images (Wang et al., 2004; Taubman et al., 2002). The PSNR value measures the peak signal-to-noise ratio between the dehazed image and the reference image, while the SSIM value measures the structural similarity between the dehazed image and the reference image.

A higher PSNR value implies better image quality. Similarly, a higher SSIM value indicates higher perceptual quality. For example – looking at the table, the PSNR and SSIM values can be compared for the image pair “Img-01-Ours” and “Img-01-Hazy”. The PSNR value for the “Img-01-Ours” is 19.5404, whereas for the “Img-01-Hazy”, it is 14.9922. This suggests that the dehazed image has significantly better quality than the hazy image in terms of PSNR. Similarly, the SSIM value for the “Img-01-Ours” image is 0.6800, while for the “Img-01-Hazy”, it is 0.4113. Again, this indicates that the dehazed image is considerably higher quality in terms of SSIM. Overall, the table provides a quantitative analysis of the quality of dehazed images using PSNR and SSIM values, and the presented data suggest that the proposed method produces high-quality haze-free images than other methods.

## 5 Discussion

The proposed method is distinguishable in many ways from other similar state-of-the-art methods (Huang & Liu, 2019; Yang et al., 2013), where the authors employed a simple histogram equalisation technique that is intended to work as a front-end of the ORB-SLAM2 system for preprocessing low-light input image stream. Although their method reasonably improves image visibility and subsequent feature extraction to some extent, they did not address haze in the image even when it is a common issue found in sudden illumination changes. Moreover, their implementation was only tested on a point feature-based SLAM system and in well-lit standard datasets, which suffer in low textured environment (Gomez et al., 2017).

The low-light dataset of a Macadamia orchard provides the stereo sequence without any ground truth data; therefore, it was not possible to compare the estimated trajectories of the tested methods with the ground truth data. However, the dataset was very effective in evaluating the tracking loss and success rate of the visual SLAM methods in the low-light condition in the real agricultural environment. Regarding the trajectory estimation accuracy, the proposed method has been tested in the KITTI, EuRoC and ROSARIO datasets.

## 6 Conclusion

In this paper, we have presented a novel stereo-visual SLAM system, which is capable of working in a challenging agricultural environment. The approach introduces a local point and line features recovery technique, which enables the proposed method to work in a variety of shaded and hazy agricultural environments, i.e., working under tree canopies where GPS single is unreachable or not reliable. For the performance evaluation, firstly, the proposed method is tested on standard benchmarking datasets such as KITTI, and EuRoC MAV. The testing results have been compared with the state-of-the-art methods, VINS-Fusion, PL-SLAM, ORB-SLAM2; and our method presents superior results in most cases. Secondly, the proposed method has been tested on ROSARIO and our low-light agricultural datasets and compared with baseline methods. The proposed method achieved superior RMSE results on the ROSARIO dataset. The low-light agricultural dataset of a macadamia orchard has been used to validate tracking success/failure in such a difficult environment, and our proposed method successfully keeps tracking the complete trajectory while the other visual SLAM methods fail. Finally, the functionality of the proposed method has been tested in hazy conditions using the O-HAZE dataset.

**Acknowledgements** The authors would like to acknowledge the support from Southern Cross University and the University of South Australia for the stipend and Tuition Fee Offset scholarships. Moreover, the author would like to thank the team of AUTO ORCHARD PTY LTD, McLeans Ridges NSW 2480, Australia, for the experimental and fieldwork facilities.

**Author contributions** Conceptualization, methodology, formal analysis and investigation, writing—original draft preparation: RI; Writing—review and editing, supervision: HH; Review, supervision: RVP; Review: TH.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions.

**Data availability** <https://rafiqrana.github.io/low-light-dataset/>.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copy-

**Table 7** Approximate CPU and RAM utilization statistics of different VSLAM methods in different datasets (LLAD stands for, Low-light Agricultural Dataset), where CPU utilization in percent (%) @3.38GHz, and RAM utilization in megabyte (MB)

Datasets	Proposed method		PL-SLAM		VINS-fusion		ORB-SLAM2	
	CPU	RAM	CPU	RAM	CPU	RAM	CPU	RAM
KITTI	33	560	35	580	25	502	27	530
EuRoC	32	542	34	553	22	505	25	523
LLAD	34	1154	36	1325	28	1039	30	1092

right holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A CPU and RAM utilization

The approximate CPU and RAM utilization are presented in Table 7 according to different datasets and different tested methods (Mur-Artal & Tardós, 2017; Gomez et al., 2017; Qin et al., 2018). Note that it is found that there was a subtle difference in CPU and RAM utilization for different sequences of each dataset; therefore, approximate average measurements have been taken for the given datasets.

## B SLAM evaluation metric

The relative pose error (RPE) is a standard evaluation metric for measuring the local accuracy of the estimated trajectory. The RPE is determined based on the delta pose difference between the estimated trajectory  $P_i \in SE(3)$  and the ground truth trajectory  $G_i \in SE(3)$ . As per the standard (Grupp, 2017; Prokhorov et al., 2019), the RPE is defined as:

$$E_i^\Delta = (G_i^{-1}G_{i+\Delta})^{-1}(P_i^{-1}P_{i+\Delta}) \quad (12)$$

The root mean square error (RMSE) for the RPEs is calculated as:

$$RMSE^{i,\Delta} = \sqrt{\frac{1}{n} \sum_{i=1}^n \|E_i\|^2} \quad (13)$$

The parameter  $\Delta$  determines the distance between the pose pairs along the trajectories. For evaluating the proposed method,  $\Delta = 100$  (m) is considered which means the pose error is measured every 100 m of camera movement.

## References

Aguiar, A. S., dos Santos, F. N., Cunha, J. B., Sobreira, H. M. P., & Sousa, A. J. (2020). Localization and mapping for robots in agriculture and forestry: A survey. *Robotics*, 9, 97.

- Alismail, H., Kaess, M., Browning, B., & Lucey, S. (2017). Direct visual odometry in low light using binary descriptors. *IEEE Robotics and Automation Letters*, 2(2), 444–451.
- Ancuti, C. O., Ancuti, C., Timofte, R., & Vleeschouwer, C. D. (2018). O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images.
- Ball, D., Upcroft, B., Wyeth, G., Corke, P., English, A., Ross, P., & Bate, A. (2016). Vision-based obstacle detection and navigation for an agricultural robot. *Journal of Field Robotics*, 33, 1107–1130.
- Bavle, H., De La Puente, P., How, J. P., & Campoy, P. (2020). VPS-SLAM: Visual planar semantic SLAM for aerial robotic systems. *IEEE Access*, 8, 60704–60718. <https://doi.org/10.1109/ACCESS.2020.2983121>
- Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., & Siegwart, R. (2016). The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research*. <https://doi.org/10.1177/0278364915620033>
- Cai, B., Xu, X., Jia, K., Qing, C., & Tao, D. (2016). DehazeNet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11), 5187–5198. <https://doi.org/10.1109/TIP.2016.2598681>
- Cao, Y., & Beltrame, G. (2021). VIR-SLAM: Visual, inertial, and ranging slam for single and multi-robot systems. *Autonomous Robots*, 45, 905–917.
- Cheein, F. A. A., & Guivant, J. E. (2014). SLAM-based incremental convex hull processing approach for treetop volume estimation. *Computers and Electronics in Agriculture*, 10(2), 19–30.
- Cheein, F. A. A., Steiner, G., Paina, G. P., & Carelli, R. O. (2011). Optimized EIF-SLAM algorithm for precision agriculture mapping based on stems detection. *Computers and Electronics in Agriculture*, 78, 195–207.
- Chen, C., Zhu, H., Li, M., & You, S. (2018). A review of visual-inertial simultaneous localization and mapping from filtering-based and optimization-based perspectives. *Robotics*. <https://doi.org/10.3390/robotics7030045>
- Chen, C., Zhu, H., Wang, L., & Liu, Y. (2019). A stereo visual-inertial SLAM approach for indoor mobile robots in unknown environments without occlusions. *IEEE Access*, 7, 185408–185421.
- Cvisic, I. (2017). SOFT-SLAM: Computationally efficient stereo visual SLAM for autonomous UAVs
- De Croce, M., Pire, T., & Bergero, F. (2019). DS-PTAM: Distributed stereo parallel tracking and mapping SLAM system. *Journal of Intelligent & Robotic Systems*, 95(2), 365–377.
- Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., & Lu, Y. (2011). Fast efficient algorithm for enhancement of low lighting video. In *2011 IEEE international conference on multimedia and expo* (pp. 1–6).
- Engel, J., Schöps, T., & Cremers, D. (2014). LSD-SLAM: Large-scale direct monocular SLAM. In *European conference on computer vision* (pp. 834–849).
- Fuentes-Pacheco, J., Ascencio, J., & Rendon-Mancha, J. (2015). Visual simultaneous localization and mapping: A survey. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-012-9365-8>

- Galvez-Lopez, D., & Tardos, J. (2012). Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics*, 28, 1188–1197. <https://doi.org/10.1109/TRO.2012.2197158>
- Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The KITTI dataset. *International Journal of Robotics Research (IJRR)*.
- Gomez, R., Moreno, F. A., Scaramuzza, D., & González-Jiménez, J. (2017). PL-SLAM: A stereo SLAM system through the combination of points and line segments. *IEEE Transactions on Robotics*. <https://doi.org/10.1109/TRO.2019.2899783>
- Grupp, M. (2017). evo: Python package for the evaluation of odometry and SLAM. <https://github.com/MichaelGrupp/evo>
- Habibie, N., Nugraha, A. M., Anshori, A. Z., Ma'sum, M. A., & Jatmiko, W. (2017). Fruit mapping mobile robot on simulated agricultural area in Gazebo simulator using simultaneous localization and mapping (SLAM). In *2017 International symposium on micro-nanomechatronics and human science (MHS)* (pp. 1–7). <https://doi.org/10.1109/MHS.2017.8305235>
- Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511811685>
- Huang, J., & Liu, S. (2019). Robust simultaneous localization and mapping in low-light environment. *Computer Animation and Virtual Worlds*, 30(3–4), e1895. <https://doi.org/10.1002/cav.1895>
- Huang, W. H. (2001). Optimal line-sweep-based decompositions for coverage algorithms. In *Proceedings of ICRA IEEE international conference robotics and automation* (cat. no.01ch37164) (Vol. 1, pp. 27–32). <https://doi.org/10.1109/ROBOT.2001.932525>
- Islam, R. & Habibullah, H. (2021). A semantically aware place recognition system for loop closure of a visual SLAM system. In *2021 4th International conference on mechatronics, robotics and automation (ICMRA)* (pp. 117–121).
- Islam, R., & Habibullah, H. (2022). Place recognition with memorable and stable cues for loop closure of visual slam systems. *Robotics*. <https://doi.org/10.3390/robotics11060142>
- Jiao, J., Wang, C., Li, N., Deng, Z., & Xu, W. (2021). An adaptive visual dynamic-SLAM method based on fusing the semantic information. *IEEE Sensors Journal*. <https://doi.org/10.1109/JSEN.2021.3051691>
- Kerl, C., Sturm, J., & Cremers, D. (2013). Dense visual SLAM for RGB-D cameras. In *2013 IEEE/RSJ international conference on intelligent robots and systems* (pp. 2100–2106).
- Kim, J., Jeon, M. H., Cho, Y., & Kim, A. (2021). Dark synthetic vision: Lightweight active vision to navigate in the dark. *IEEE Robotics and Automation Letters*, 6, 143–150.
- Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., & Burgard, W. (2011). G2o: A general framework for graph optimization. In *2011 IEEE international conference on robotics and automation* (pp. 3607–3613).
- Lazaros, N., Sirakoulis, G. C., & Gasteratos, A. (2008). Review of stereo vision algorithms: From software to hardware. *International Journal of Optomechatronics*, 24, 435–462. <https://doi.org/10.1080/15599610802438680>
- Lee, S., Yun, S. M., Nam, J. H., Won, C. S., & Jung, S. W. (2016). A review on dark channel prior based image dehazing algorithms. *EURASIP Journal on Image and Video Processing*, 2016, 1–23.
- Lemaire, T., Berger, C., Jung, I. K., & Lacroix, S. (2007). Vision-based SLAM: Stereo and monocular approaches. *International Journal of Computer Vision*, 74, 3343–364.
- Lepetit, V., Moreno-Noguer, F., & Fua, P. (2008). EPnP: An accurate  $O(n)$  solution to the PnP problem. *International Journal of Computer Vision*, 81, 155–166.
- Liang, Z., & Wang, C. (2021). A semi-direct monocular visual slam algorithm in complex environments. *Journal of Intelligent & Robotic Systems*, 10(1), 11–19.
- Liu, H., Chen, M., Zhang, G., Bao, H., & Bao, S. Y. Z. (2018). ICE-BA: Incremental, consistent and efficient bundle adjustment for visual-inertial SLAM. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, 1974–1982.
- Long, J., Shi, Z., Tang, W., & Zhang, C. (2014). Single remote sensing image dehazing. *IEEE Geoscience and Remote Sensing Letters*, 11, 59–63.
- Ma, J., Wang, X., He, Y., Mei, X., & Zhao, J. (2019). Line-based stereo SLAM by junction matching and vanishing point alignment. *IEEE Access*, 7, 181800–181811.
- Marks, T. K., Howard, A., Bajracharya, M., Cottrell, G. W., & Matthies, L. (2008). Gamma-SLAM: Using stereo vision and variance grid maps for SLAM in unstructured environments. In *2008 IEEE international conference on robotics and automation* (pp. 3717–3724). <https://doi.org/10.1109/ROBOT.2008.4543781>
- Matsuzaki, S., Masuzawa, H., Miura, J., & Oishi, S. (2018). 3D semantic mapping in greenhouses for agricultural mobile robots with robust object recognition using robots' trajectory. In *2018 IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 357–362).
- Mikulík, A., Perdoch, M., Chum, O., & Matas, J. (2010). Learning a fine vocabulary (pp. 1–14). [https://doi.org/10.1007/978-3-642-15558-1\\_1](https://doi.org/10.1007/978-3-642-15558-1_1)
- Milford, M. J., & Wyeth, G. F. (2012). SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In *2012 IEEE international conference on robotics and automation* (1643–1649).
- Muñoz-Salinas, R., & Medina-Carnicer, R. (2020). UcoSLAM: Simultaneous localization and mapping by fusion of keypoints and squared planar markers. *Pattern Recognition*, 101, 107193.
- Mur-Artal, R., & Tardós, J. D. (2017). ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5), 1255–1262.
- Nalpantidis, L., Sirakoulis, G. C., & Gasteratos, A. (2011). Non-probabilistic cellular automata-enhanced stereo vision simultaneous localization and mapping. *Measurement Science and Technology*, 22(11), 114027.
- Paturkar, A., Gupta, G. S., & Bailey, D. (2017). Overview of image-based 3D vision systems for agricultural applications. In *2017 International conference on image and vision computing New Zealand (IVCNZ)* (pp. 1–6).
- Pire, T., Corti, J., & Grinblat, G. (2020). Online object detection and localization on stereo visual SLAM system. *Journal of Intelligent & Robotic Systems*, 98(2), 377–386.
- Pire, T., Mujica, M., Civera, J., & Kofman, E. (2019). The Rosario dataset: Multisensor data for localization and mapping in agricultural environments. *The International Journal of Robotics Research*, 38(6), 633–641. <https://doi.org/10.1177/0278364919841437>
- Prokhorov, D., Zhukov, D., Barinova, O., Konushin, A., & Vorontsova, A. (2019). Measuring robustness of visual SLAM. In *2019 16th International conference on machine vision applications (MVA)* (pp. 1–6).
- Qin, T., Li, P., & Shen, S. (2018). VINS-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4), 1004–1020. <https://doi.org/10.1109/TRO.2018.2853729>
- Qin, T., Pan, J., Cao, S., & Shen, S. (2019). A general optimization-based framework for local odometry estimation with multiple sensors.
- Quan, M., Piao, S., He, Y., Liu, X., & Qadir, M. Z. (2021). Monocular visual slam with points and lines for ground robots in particular scenes: Parameterization for lines on ground. *Journal of Intelligent & Robotic Systems*, 10, 172.
- Ranganathan, A., Matsumoto, S., & Ilstrup, D. (2013). Towards illumination invariance for visual localization. In *2013 IEEE international conference on robotics and automation* (pp. 3791–3798).



- Rovira-Más, F., Zhang, Q., & Reid, J. F. (2008). Stereo vision three-dimensional terrain maps for precision agriculture. *Computers and Electronics in Agriculture*, 60, 133–143.
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision* (pp. 2564–2571).
- Schubert, S., Neubert, P., & Protzel, P. (2021). Graph-based non-linear least squares optimization for visual place recognition in changing environments. *IEEE Robotics and Automation Letters*, 62, 811–818. <https://doi.org/10.1109/LRA.2021.3052446>
- Shu, F., Lesur, P., Xie, Y., Pagani, A., & Stricker, D. (2020). SLAM in the field: An evaluation of monocular mapping and localization on challenging dynamic agricultural environment. [arXiv:2011.01122](https://arxiv.org/abs/2011.01122)
- Shu, F., Lesur, P., Xie, Y., Pagani, A., & Stricker, D. (2021). SLAM in the field: An evaluation of monocular mapping and localization on challenging dynamic agricultural environment. In *2021 IEEE winter conference on applications of computer vision (WACV)* (pp. 1760–1770).
- Shuai, Y., Liu, R., & He, W. (2012). Image haze removal of wiener filtering based on dark channel prior. In *2012 Eighth international conference on computational intelligence and security* (pp. 318–322).
- Sumikura, S., Shibuya, M., & Sakurada, K. (2019). OpenVSLAM: A versatile visual SLAM framework. In *Proceedings of the 27th ACM international conference on multimedia* (pp. 2292–2295). USAACM. <https://doi.org/10.1145/3343031.3350539>
- Taubman, D. S., Marcellin, M. W., & Rabbani, M. (2002). JPEG2000: Image compression fundamentals, standards and practice. *Journal of Electronic Imaging*, 11(2), 286–287.
- Tykkälä, T., & Comport, A. I. (2011). A dense structure model for image based stereo SLAM. In *2011 IEEE international conference on robotics and automation* (pp. 1758–1763).
- Von Gioi, R. G., Jakubowicz, J., Morel, J. M., & Randall, G. (2008). LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4), 722–732.
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13, 600–612.
- Wen, S., Li, P., Zhao, Y., Zhang, H., Sun, F., & Wang, Z. (2021). Semantic visual SLAM in dynamic environment. *Autonomous Robots* 1–12.
- Xu, H., Guo, J., Liu, Q., & Ye, L. (2012). Fast image dehazing using improved dark channel prior. In *2012 IEEE international conference on information science and technology* (pp. 663–667).
- Yang, J., Chung, S., Hutchinson, S., Johnson, D., & Kise, M. (2013). Vision-based localization and mapping for an autonomous mower. In *Proceedings of IEEE/RSJ international conference on intelligent robots and systems* (pp. 3655–3662). <https://doi.org/10.1109/IROS.2013.6696878>
- Yang, W., & Zhai, X. (2019). Contrast limited adaptive histogram equalization for an advanced stereo visual SLAM system. In *2019 International conference on cyber-enabled distributed computing and knowledge discovery (CyberC)* (pp. 131–134).
- Zhong, S., & Chirarattananon, P. (2021). An efficient iterated EKF-based direct visual-inertial odometry for MAVs using a single plane primitive. *IEEE Robotics and Automation Letters*, 6, 486–493.
- Zhou, H., Zou, D., Pei, L., Ying, R., Liu, P., & Yu, W. (2015). StructSLAM: Visual SLAM with building structure lines. *IEEE Transactions on Vehicular Technology*, 64(4), 1364–1375.
- Zhu, Q., Mai, J., & Shao, L. (2015). A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24, 3522–3533.
- Zuo, X., Xie, X., Liu, Y., & Huang, G. (2017). Robust visual SLAM with point and line features. In *2017 IEEE/RSJ international conference*

*on intelligent robots and systems (IROS)*. <https://doi.org/10.1109/IROS.2017.8205991>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



autonomy of mobile robotic platforms.

**Rafiqul Islam** received a Bachelor of Science (BSc) degree in Electrical and Electronic Engineering from Eastern University (EU), Bangladesh in 2016, and a PhD from the University of South Australia (UniSA), Australia in 2022. He is currently a research associate at the University of New South Wales (UNSW), Australia with a strong background in robotics, AI, and intelligent systems. His research focuses on the development of intelligent systems to improve the performance and



Scientific Collaborator with the BEAMS Department, Université Libre de Bruxelles, Belgium, and a Research Associate position at the University of New South Wales, Canberra, Australia. His research interests are in the joint areas of mechatronics, robotics, machine learning, artificial intelligence, industrial automation, biomechanics, nanotechnology, and low-frequency vibration isolation.

**Tagor Hossain** received a Bachelor of Science (BSc) degree in Electrical and Electronic Engineering from Eastern University (EU), Bangladesh in 2016, and a Master of Engineering degree in Mechatronics Engineering from the University of South Australia (UniSA), Australia in 2021. He is pursuing a PhD in Mechatronics Engineering at the University of Adelaide, Australia. His research interests include autonomous vehicles' path planning, path tracking and motion control.

