Automotive Radar Image Segmentation with Frame Fusion

Yang Xiao, Liam Daniel, Scott Cassidy, Sukhjit Pooni, Edward Hoare, Anum Ahmed Pirkani, Mikhail Cherniakov, Marina Gashinova*

Microwave Integrated Systems Laboratory (MISL), The University of Birmingham, UK
**yxx752@student.bham.ac.uk*

Keywords: Automotive radar, image segmentation, autonomous driving, frame fusion, IMU, frame registration, Kalman filter, affine transformation.

Abstract

Image segmentation on automotive radar imagery is the key technique for identifying the passable and impassable regions for path planning in autonomous or assistive driving. The availability of consecutive frames which measure the driving scene shifted along with the timeline enables improved segmentation on radar imagery. The frame fusion on automotive radar map is implemented as a two-step procedure: 1) The pixel-to-pixel mapping between consecutive frames is achieved based on an inertial measurement unit (IMU); 2) The information fusion of consecutive frames is achieved based on the Kalman filter. The frame fusion operation leads to correct classification of the initially 'unknown" regions and overall improves the confidence of classification compared to single frame segmentation. The segmentation results with frame fusion are presented and compared with the results of single frame segmentation to demonstrate the segmentation improvement.

1 Introduction

Microwave radar has been widely used as the primary automotive sensor system to sense the surrounding environment and serves for advanced driver assistance systems (ADAS) and autonomous driving (AD) in modern vehicles [1]. The automotive sensors generally include automotive radar, LiDAR, camera, and ultrasonic sensors. Automotive radar has the following advantages compared with the other sensors: 1) it is the most reliable one which can mitigate the effects of the environment such as bad weather, light conditions, etc; 2) it can directly measure the range and velocity of objects. Therefore, information extraction of the driving environment based on data from the automotive radar can give solid support to achieve the functions like path planning and further benefit the AD system. The identification of passable and impassable regions is the key information for path planning, and image segmentation is the general approach for achieving this.

Image segmentation technique has been widely utilized on optical imagery obtained from camera [2], as well as the LiDAR point cloud data [3]. Recent advances in the development of high-resolution mm-wave and low-THz radar allow to deliver high-resolution imagery where the contrast between "clutter" regions can be used for image segmentation. Thus, the full scene reconstruction purely based on automotive radar data was considered in our previous research in [4, 5] where a customized hybrid segmentation method has been proposed that fits the format of automotive radar data. Although methodology in [4, 5] shows good potential for image segmentation of automotive radar imagery in the single radar frame (scanned or beamformed single scene map), its performance can be further improved by fusing information from several consecutive frames. This paper presents two techniques to facilitate improved image segmentation with frame fusion: 1) pixel mapping between consecutive frames, which is the key step for further frame information fusion, and 2) use of the Kalman filter to combine the information of several frames.

The approach for achieving the geometric pixel mapping between consecutive frames is tracking the region movement based on the GPS and real-time driving information measured by the inertial measurement unit (IMU). This is the straightforward way of achieving frame registration. The information fusion of multiple frames can be further developed by the Kalman filter, which is the general approach that utilizes a series of measurements observed over time to improve the performance of a single measurement [6]. The frame fusion is therefore achieved by combining frame registration based on IMU measurement, and Kalman filter approaches in this paper.

2 Single Frame Image Segmentation Overview

The previous implementation of single frame segmentation [5] is composed of two main steps: first, pre-segmentation using image processing methods, edge detection in particular, where different clutter regions in the radar map are resolved and provide a contrast in terms of distribution parameters; the second step is the region classification based on multi-variate Gaussian distribution (MGD) classifier, which uses distribution parameters as the input. When the whole scene within the image is resolved (distributed clutter returns from regions of different classes are above the noise floor), the segmentation leads to a full scene reconstruction of radar map, so that each pixel is labelled as belonging to one of the chosen classes, such as asphalt, grass, objects, shadows, as well as the 'unknown area', which is a class used to describe low confidence of classification in case of high confusion or data

corruption. In such case probability of correct classification can be increased if the same region is tracked and re-classified within a number of consecutive frames.

In this paper, the proposed method for achieving frame fusion is demonstrated for improving the classification of initially 'unknown' class region, however, the method is equally applicable to any class to increase the confidence of correct classification.

3 Frame Registration of Automotive Radar Imagery using IMU

When considering the automotive radar application in the world coordinate system, the consecutive frames of the radar map detect the scene shifts along with the movement of the vehicle platform. To achieve the frame registration, the geometric transformation between radar maps needs to be understood to achieve pixel mapping, which can be represented mathematically as:

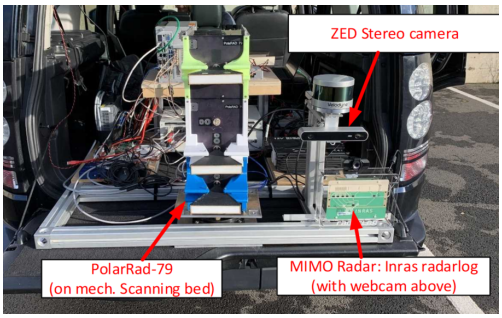
$$\begin{cases} x' = s_x \cos(\theta)x - sh_x \sin(\theta)y + t_x \\ y' = sh_x \sin(\theta)x + s_y \cos(\theta)y + t_y \end{cases} \quad (1)$$

where (x, y) and (x', y') are the coordinates in previous and current radar frame, s_x and s_y are the scale factors along x and y axis, sh_x and sh_y are the shear factors along x and y axis, θ specifies the rotation angle and t_x and t_y are the position shifts. When considering the registration between PPI (plan position indicator) radar maps, the scale and shear transformations are not expected since the radar maps are shifted and rotated under the same world coordinate. Therefore, the parameters of s_x, s_y, sh_x, and sh_y can be assumed as "1". The remaining position shifts and the rotation angle are obtained from the synchronous measurement of the IMU setup as described below.

The IMU system utilized for our automotive radar measurement platform is the Spatial FOG unit from Advanced Navigation [7], which is a GPS-aided inertial navigation system. The installation of IMU and radar systems used for data collection is presented in Fig. 1, which is placed on the roof of the vehicle since it requires to be mounted close to the centre of gravity of the vehicle. The sensor axes are well aligned with the axes of the vehicle. The radar systems are rear installed in the vehicle as shown in Fig. 1 (b), which includes 79 GHz FMCW mechanical scanning radar as well as MIMO radar. Here, we only use the data collected by FMCW radar for this part of research and the details of the characteristic of the radar system and data collection have been given in [5].

The output of the IMU setup includes: 1) precise GPS information which includes latitude, longitude, and height; 2) the real-time driving information which includes velocity, angular velocity, acceleration, and orientation. All the information is synchronized with our consecutive radar frame by timestamp. First of all, the required rotation angle θ can be calculated by subtracting the output orientations of two considered frames by $\theta = o' - o$, where o' and o are the orientations of current and previous frames.

(a)



(b)

Fig. 1 (a) The installation of IMU setup on roof of the car; (b) The installation of radar systems utilized for data collection.

Due to the fact that the output velocity and acceleration information are calculated based on GPS information, the GPS positions are considered to be the more precise way of estimating the position shifts between frames. This means we need to transfer the information of latitude, longitude into cartesian coordinates using the following equations:

$$\begin{cases} x_c = R \cdot \cos(lat) \cdot \cos(lon) \\ y_c = R \cdot \cos(lat) \cdot \sin(lon) \end{cases} \quad (2)$$

where (x_c, y_c) is the cartesian coordinate of the vehicle which corresponds to the original point of the radar map (at the range of 0). $R = 6371$ m is the approximate radius of the earth, lat and lon are the latitude and longitude values output by the IMU system.

Assuming that the original point of another map is calculated using the same method as (x_c', y_c'), then the position shift can be represented as:

$$\begin{cases} t_x = x_c - x_c' \\ t_y = y_c - y_c' \end{cases} \quad (3)$$

The pixel-to-pixel mapping between every two frames can therefore be obtained using Eq. (1) based on the obtained position shifts t_x, t_y, and the rotation angle θ.

The instance of the implementation of pixel mapping between two consecutive radar maps is given in Fig. 2, where the stationary car visible in the radar map as a tetragon region of

high return, highlighted by the red bounding box in the frame of Fig. 2 (b), is to be projected into the next frame. The projected area coincides well with the position of the car in the consecutive frame. Similarly, all regions in the first (seed) frame will be projected to the next consecutive frames, so that once classified the information can be used to improve the probability of correct classification.

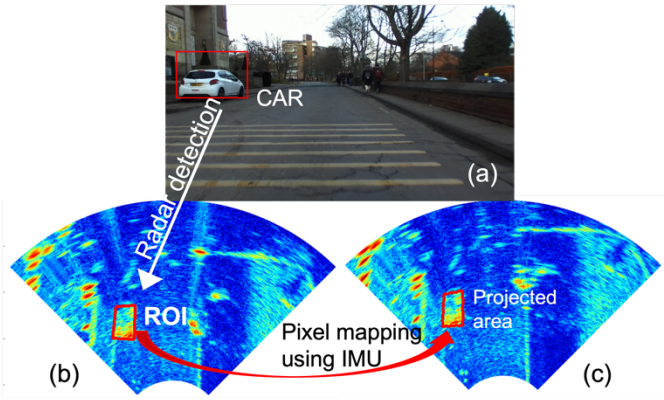


Fig. 2 The instance of projected region obtained based on IMU information. (a) the optical image; (b) the current frame with ROI; (c) the previous frame with projected area which is produced by pixel mapping of the ROI in (b).

4 Frame Fusion based on Kalman Filter

Kalman filter is an algorithm that provides an estimation of the current system state based on previous measurement and current measurement. The general mathematical definition of Kalman filter is:

$$\hat{\mathbf{x}}_{n,n} = \mathbf{F}\hat{\mathbf{x}}_{n-1,n} + \mathbf{G}\mathbf{u}_n + \mathbf{w}_n \tag{3}$$

where $\hat{\boldsymbol{x}}_{n,n}$ is a predicted system state at time step n; $\hat{\boldsymbol{x}}_{n-1,n}$ is an estimated system state at time step $n-1$; F is the transition matrix between $\hat{\boldsymbol{x}}_{n,n}$ and $\hat{\boldsymbol{x}}_{n-1,n}$; $\boldsymbol{u}_n$ is the input from the system at time step n; G is the control input model which is applied to the control vector $\boldsymbol{u}_n$; and $\boldsymbol{w}_n$ is the process noise of disturbance.

In our application, Kalman filter is utilized to mitigate the error caused by the feature variation of radar frames by considering the classification information of multiple frames. The segmentation performance of CF can be improved according to the estimation of PFs that the system state we considered here is the region class weights contributed from multiple frames.

The definition of the number of required frames can be fine-tuned based on the overlap ratio between frames which depends on the velocity of the host vehicle, the frame rate of radar measurement, and the maximum detection range of radar map. The more detailed mathematical estimation of the number of overlap frames in different scenarios will be given in the next publication [8]. As an example with the driving speed of 40 km/h and the frame rate of 10 frame/s, region tracking on four previous frames is still possible within the 25 m detection range. Therefore in this paper, we would use four frames just to illustrate the approach, though the measured data were recorded with the velocity of the host vehicle of 5 m/s, the detection range of 25 m and the limited frame rate of the demonstrator of 1 frame/s (1 Hz scan rate).

The frame fusion is done based on the voting procedure of the region class weights, and the mathematical representation of the fusion over four frames (as shown in Fig. 3) can be given as:

$$\hat{\boldsymbol{x}}_{n,n} = \hat{\boldsymbol{x}}_{n-1,n} + \hat{\boldsymbol{x}}_{n-2,n} + \hat{\boldsymbol{x}}_{n-3,n} + \boldsymbol{u}_n \tag{4}$$

Here, $\hat{\boldsymbol{x}}_{n,n}$ is the final predicted weights for the classification of the ROI in current frame (CF) which is at the time step of n. $\boldsymbol{u}_n = [p_a^{CF}, p_g^{CF}, p_s^{CF}, p_o^{CF}]$ are the weights obtained from the output of MGD classifier in the single frame segmentation of CF, where 'a, g, s, o' stand for asphalt, grass, shadows, and objects [5]. $\hat{\boldsymbol{x}}_{n-1,n}$, $\hat{\boldsymbol{x}}_{n-2,n}$, and $\hat{\boldsymbol{x}}_{n-3,n}$ are the region class weights obtained based on the classification results of the projected regions in the previous frames (PFs) which are at the time steps of $[n-1, n-2, n-3]$. The calculation details for obtaining $\hat{\boldsymbol{x}}_{n-1,n}$, $\hat{\boldsymbol{x}}_{n-2,n}$, and $\hat{\boldsymbol{x}}_{n-3,n}$ are going to be discussed in the frame fusion implementation based on Fig. 3 below.

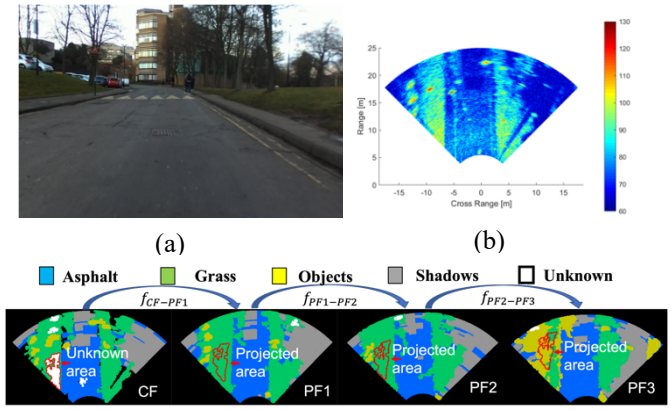


Fig. 3 Example of the identification of unknown area using frame fusion. (a) optical image of the scene; (b) radar map; (c) tracking of the unknown areas in the segmented maps of PF1-PF3.

The value of **G** in Eq. (3) determines the amount of contribution from the current single frame which is assumed as "1" here for simply the calculation. The $\boldsymbol{w}_n$ combine the noises due to the various impact factors such as the error produced in frame registration and single frame segmentation. The discussion on these errors is out of the scope of this paper and $\boldsymbol{w}_n$ is assumed as '0' here.

The example of frame fusion implementation to identify the 'unknown area' in CF is shown in Fig. 3. The calculation of the variables of $\hat{\boldsymbol{x}}_{n-1,n}$, $\hat{\boldsymbol{x}}_{n-2,n}$, and $\hat{\boldsymbol{x}}_{n-3,n}$ is the key stage for achieving frame fusion which includes the geometric

transformation using IMU information and the class weight calculation based on the segmentation results of PFs. We will use the approach to re-classify the 'unknown area' in CF in Fig 3(c) on the left, whose pixel coordinate set can be represented as C_{CF}. When simplifying the geometric transformation given in Eq. (1) as function of $f(C)$, the coordinates of projected regions in PF1-PF3 (red contoured projected regions in Fig. 3 (c)) can be calculated as:

$$\begin{cases} C_{PF1} = f_{CF-PF1}(C_{CF}) \\ C_{PF2} = f_{PF1-PF2}(C_{PF1}) \\ C_{PF3} = f_{PF2-PF3}(C_{PF2}) \end{cases} \quad (5)$$

The projected region might include multiple area classes, and the region class weights obtained from PFx can be represented as:

$$\hat{x}_{n-x,n} = [p_a^{PFx}, p_g^{PFx}, p_s^{PFx}, p_o^{PFx}] \quad (6)$$

where each item of p_i^{PFx} is calculated as:

$$p_i^{PFx} = \frac{N_i}{N_s} \quad (7)$$

Here N_i is the number of pixels which belong to class i in the projected region, and N_s is the number of pixels of the projected region.

The final predicted weights for the classification of the ROI can therefore be calculated by adding all the weights contributed from multiple frames of CF, PF1-PF3 as:

$$\hat{\boldsymbol{x}}_{\boldsymbol{n,n}} = [p_a, p_g, p_s, p_o] \quad (8)$$

The area class corresponding to the highest-class weight is selected as the final classification result of the 'unknown area'.

5 Results of Frame Fusion on Automotive Radar Map Segmentation

In this section, we estimate the performance of segmentation results obtained with and without frame fusion based on both segmentation examples and Jaccard similarity coefficients (JSCs) [5] estimation.

Fig. 4 and 5 show the segmentation results of two frames, in which insets of (c) and (d) are the segmentation results obtained from the algorithm with and without frame fusion. In Fig. 4(c), area A is the rougher tarmac on the car park area according to the ground truth optical image and is identified as an 'unknown area' since its backscattering return pdf shows low bias to the finer tarmac on the driving path. Area B is a tree on the lawn area along the road. Both of them are correctly classified as 'asphalt' and 'objects' respectively in Fig. 4 (d) after frame fusion. For region C, although most of it is correctly identified as 'grass' in the frame fusion, the extremely small tree areas are merged into grass due to the under-segmentation caused by single frame segmentation [5]. This also leads to the limitation of our frame fusion method that it highly relies on the segmentation result of single frame segmentation.

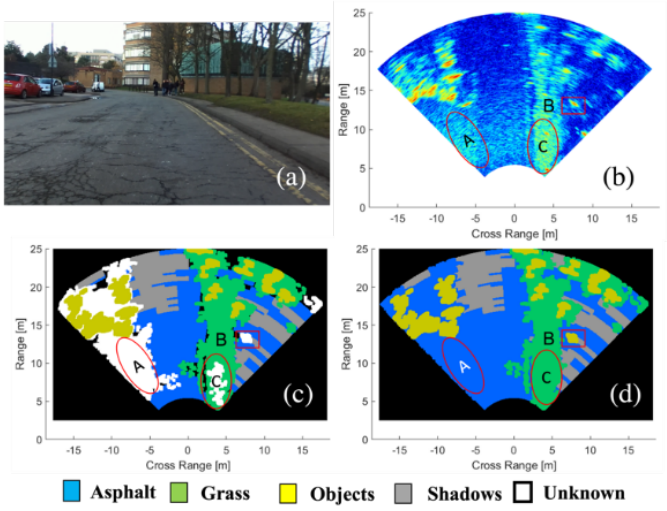


Fig. 4 First instance of segmentation before and after frame fusion. (a) is the optical imagery shows the driving scene; (b) is the radar map; (c) is the segmented map obtained from single frame segmentation; (d) is the segmented map obtained after frame fusion operation.

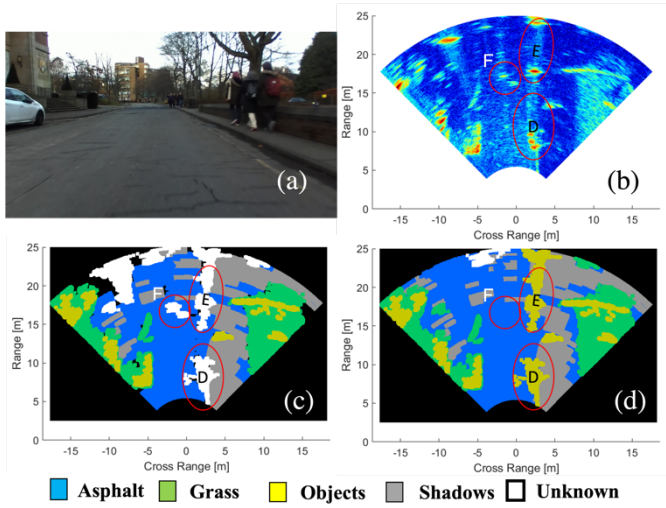


Fig. 5 Second instance of segmentation results before and after frame fusion. The caption definitions are the same with Fig. 4.

Fig. 5(c) shows that the regions of D and E are identified as 'unknown areas' since they are extended objects - low brick fences along the radar signal illumination direction, which means a lower power return than from the other objects. Finally, they are correctly identified as 'objects' based on the frame fusion and can be used to plan the passable path, though further processing steps, such as Hough transform should be used to highlight linearly extended objects (curbs) for further path planning strategy. The region F in Fig. 5 (b) is a metal manhole cover which has higher backscattering than tarmac according to the observation through consecutive frames. It has been correctly identified as the passable region in Fig. 5(d) after frame fusion.

The JSC estimation on both segmentation results based on the mathematical definition [5]:

$$JSC = \frac{A_{fs} \cap A_{label}}{A_{label}} \quad (8)$$

Here, A_{fs} is the number of correctly classified pixels overlapping with the labelled data regions of that class and A_{label} is the total number of pixels of the corresponding class in the labelled data. As shown in Table 1, the JSCs of areas of asphalt, grass, and objects are significantly improved after using frame fusion. However, the segmentation of shadows has not been improved due to the fact that they are segmented using the thresholding strategy instead of MGD classifier due to their particular nature of being a radar noise, rather than physical return.

Table 1. The comparison of JSCs before and after frame fusion implementation.

Area classes	Asphalt	Grass	Objects	Shadows
JSCs of single frame	0.69	0.7	0.68	0.83
JSCs after frame fusion	0.77	0.86	0.72	0.83

6 Conclusion

This paper proposes the frame fusion technique on consecutive radar maps, which is the extension work of single frame segmentation algorithm developed in [5]. The proposed method includes two stages: 1) the pixel mapping between consecutive frames based on the IMU measurement; 2) the information fusion of multiple frames using Kalman filter. The proposed method is implemented on our dataset to improve the identification of the 'unknown areas' which cannot be classified purely based on the feature extracted from single frame. The segmentation performance is estimated visually and mathematically that frame fusion results show significant improvement compared with the results obtained by single frame segmentation. The proposed segmentation with frame fusion is applicable to automotive radar imagery, where MIMO and SAR can be used to deliver high-resolution maps of the scene around the vehicle.

7 Acknowledgements

This work is supported by Innovate UK grant 104268 and is part of the project "CORTEX-Cognitive Real-Time System for Autonomous Vehicles".

8 References

[1] W. J. Fleming, "Overview of automotive sensors," *IEEE sensors journal,* vol. 1, no. 4, pp. 296-308, 2001.

[2] D. Feng *et al.*, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems,* vol. 22, no. 3, pp. 1341-1360, 2020.

[3] A. Paigwar, O. Erkent, C. Wolf, and C. Laugier, "Attentional pointnet for 3d-object detection in point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0-0.

[4] Y. Xiao, L. Daniel, and M. Gashinova, "Feature-based classification for image segmentation in automotive radar based on statistical distribution analysis," in *2020 IEEE Radar Conference (RadarConf20)*, 2020: IEEE, pp. 1-6.

[5] Y. Xiao, L. Daniel, and M. Gashinova, "Image segmentation and region classification in automotive high-resolution radar imagery," *IEEE Sensors Journal,* vol. 21, no. 5, pp. 6698-6711, 2020.

[6] G. Welch and G. Bishop, "An introduction to the Kalman filter," 1995.

[7] "Spatial FOG." Advanced navigation. https://www.advancednavigation.com/solutions/spatial-fog-dual/ (accessed at 05/05/2022).

[8] Y. Xiao, L. Daniel, and M. Gashinova, "The End-to-End Segmentation on Automotive Radar Imagery with MTI based on Frame-to-Frame Association," *In progress, intend to be submitted to IEEE Sensors Journal,* 2022.