A Traffic Information Awareness Approach Based on Video Data and Millimeter Wave Radar Data Fusion

Xuefan Yan¹*, Feng Ding²a, Changjie Liu³b, Dong Liu⁴c

¹ICS Big Group, Lian Yungang JARI Electronics Co., Ltd, Lian Yungang, China
²Traffic Police Detachment, Ri Zhao Municipal Public Security Bureau, Ri Zhao, China
³ICS Big Group, Lian Yungang JARI Electronics Co., Ltd, Lian Yungang, China
⁴ICS Big Group, Lian Yungang JARI Electronics Co., Ltd, Lian Yungang, China

* Corresponding author: Yanxuefan@jari.cn
a rzijd@163.com, b liuchangjie@jari.cn, c liudunhua1990@126.com

Abstract—In order to solve the problems of using a single radar or video sensor in the traffic information detection process, such as susceptibility to environmental influences and non-intuitive target reflections, we propose a detection method using target correlation matching, target tracking, and target data fusion, and we show how to adjust the detection weights of the video sensor and radar sensor by adjusting the noise matrix parameters, which forms a flexible, simple, and nimble method. The efficient architecture allows accurate results in challenging environments. We show the effect on radar detection data and video detection data before and after adjustment. We test the proposed method by building a hardware verification platform and finally demonstrate it on video images. The experimental results show that the proposed method can significantly reduce the testing time while providing high detection rate in more environments.

1. INTRODUCTION

In the process of organizing traffic operations, traffic information acquisition devices are widely used in urban intersections and road sections to collect environmental information at intersections, vehicle information, and thus help improve traffic organization. Traditional traffic data acquisition devices mostly use a single sensor, more commonly used are video sensors or millimeter-wave radar sensors, usually using video sensors to collect video images of intersections, extract the foreground target in the collected video images, and then the foreground target through the visual algorithm for feature extraction; or use millimeter-wave radar sensors to detect intersections, through the detection of the obtained radar The signal is processed and analyzed to determine the target feature information in the monitored intersection. However, on the one hand, the video image collected by the video sensor is susceptible to environmental influences, for example, in severe weather conditions such as fog, wind and rain, the video image collected by the video sensor will be unclear, and thus, the complete foreground target cannot be accurately extracted in the video image, and the features of the foreground target cannot be accurately extracted; on the other hand, the detection of the intersection using the millimeter wave radar sensor, Although the radar signal is not influenced by the environment, and it is
possible to obtain information about the speed and position of the target object based on the millimeter-wave radar signal, the radar signal does not reflect the intuitive characteristics of the detected target object, and thus, it is not possible to further classify and identify the detected object based on the radar signal.

With the development of automated driving technology, the fusion of video sensor and LIDAR data has been better applied in automated vehicles, however, the high cost of LIDAR and harsh conditions of use make it impossible to use in traditional traffic intersections, millimeter wave LIDAR sensors with low price, a wide range of use scenarios and other advantages to get more and more applications, but due to the video sensor and millimeter wave LIDAR respectively, the video sensor and the millimeter wave LIDAR are not suitable for use in traditional traffic intersections. The system needs to fuse multiple information from the same target collected by multiple sensors, which requires a huge amount of processing of the original data and aggravates the data processing work of the equipment. Therefore, this paper proposes an improved traffic information acquisition method based on the fusion of video data and millimeter wave radar data, which can fuse the strengths and weaknesses of both vision and radar sensor systems to achieve accurate perception of environmental information and solve the problem that existing technologies can hardly meet the accuracy and reliability requirements of the perception system.

2. MULTI-SOURCE DATA FUSION MODEL

A data fusion model based on video sensors and millimeter wave radar sensors can be viewed as consisting of four steps:

1) Data acquisition steps: acquisition of radar and vision system data;
2) Data association matching procedure: obtain a list of target object data detected by the two sensing systems and match and associate the target object data detected by the two sensing systems;
3) Target tracking steps: matching and tracking the target and updating the target's life cycle status;
4) Target data fusion step: The target data signals from the output of the radar system and the vision system are fused

The data fusion flow chart is shown in Fig. 1.
2.1 Data Association Matching Model
When matching data correlation, the coordinates of the two sensor systems are first identical by image coordinates, and the detected targets are filtered; then the target information of the front and back frames of either sensor system or the target data of both sensor systems corresponding to the same target in the target data list are correlated by using the point trail correlation and the wayfinding correlation methods. The point-trace correlation can be used to quickly correlate and match the targets, but when the targets are far away, the point-trace correlation will have a large error, so you can use the air-trace correlation method, which can compensate for the lack of point-trace correlation, so that you can correlate and match both distant and near targets.

Trace information in wayfinding correlation method includes object ID and life cycle status. Both video sensors and radar sensors track their detected targets and assign corresponding IDs to them in the system. When correlating frame data before and after, matching using track IDs and target life cycle conditions can increase the stability of the fusion result tracking and reduce the computation amount. When the target just enters the detection range of the two sensing systems, the correlation matrix can be determined by point-trace correlation first, and after the correlation is completed, the subsequent cycles can be correlated using the track to save the computational cost of calculating the correlation matrix. The data matching flow is shown in Fig. 2.

2.2 Target Tracking Model
The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

The target tracking step is mainly used to form a management of the target's life cycle, which facilitates the evaluation of the validity of the target, timely elimination of invalid or false targets, and improves the reliability and validity of the inspection.

Steps of target tracking include:

2.2.1 Apply Kalman filtering to the target data to form the predicted state value at the next moment and the predicted covariance corresponding to the state value; the Kalman filtering algorithm predicts the feasible region of the target for the next cycle based on the current relative motion state and makes
corrections based on the actual target information to achieve continuous tracking of the data before and after.

2.2.2. The matching results are obtained by correlating and matching the target data of the current moment with the target data of the previous moment.

2.2.3. For the perfectly matched target data, the prediction result is modified and updated with the current target data to get a new state value of the target and the same ID; for the target that does not exist at the current moment, the number of consecutive non-appearing frames is recorded; if the number of consecutive non-appearing frames is greater than the threshold number of frames, the target is judged to have disappeared; the threshold number of frames is based on the continuous existence of the target. The duration and target location data are adjusted; the longer the target has been present, the larger the frame threshold, the shorter the target has been present, the smaller the frame threshold, increasing the frame threshold if the target is within the target area, decreasing the frame threshold if the target is outside the target area.

In reality, due to sensor error, etc., there may be cases where the target disappears or the target itself is caused by an error, etc. Therefore, setting the frame threshold can avoid removing the target by mistake. If the target appears in the target area, such as directly in front of or very close to the sensor system, it represents a possible threat to the current system and should be focused on, and the frame threshold should be increased.

In the target tracking step, if a target has corresponding target data in only one sensor system, it is tracked continuously by the corresponding sensor system, and when the target enters the common detection range of both sensor systems, the target data is fused directly. In this process, the corresponding system ID of the target remains unchanged, and no correlation matching is required. This improves the processing speed, and at the same time allows the use and retention of more data information than existing techniques. The flow chart of the target tracking algorithm is shown in Figure 3.

![Figure 3. Target tracking algorithm flowchart.](image)

2.3. **Target object data fusion model**

The target tracking step is mainly used to form a management of the target's life cycle, which facilitates the evaluation of the validity of the target, timely elimination of invalid or false targets, and improves the reliability and validity of the inspection.
The target object data fusion step first predicts the predicted state value at the current moment based on the fusion result at the previous moment and the corresponding predicted covariance matrix, and then fuses the data from the two sensing systems using the following equation.

\[ K_1 = P H^T (H P H^T + R_1)^{-1} \]
\[ X_1 = X + K_1(Z_1 - HX) \]
\[ P_1 = (I - K_1 H)P \]
\[ K_2 = P_1 H^T (H P_1 H^T + R_2)^{-1} \]
\[ X_f = X_1 + K_2(Z_2 - HX_1) \]
\[ P_f = (I - K_2 H)P_1 \]

where \( X \) represents the predicted state value, \( P \) represents the predicted covariance matrix, \( Z_1 \) and \( Z_2 \) represent the observations of the two sensing systems on the same target at the corresponding time, \( R_1 \) and \( R_2 \) are the corresponding noise matrices, and \( X_f \) and \( P_f \) are the fused state value and covariance matrix.

3. EXPERIMENTAL RESULTS AND VALIDATION
This part consists of two parts: the matrix noise parameter influence results and the actual intersection test. In order to improve the target detection accuracy, the matrix noise parameters are adjusted, while the actual intersection test mainly compares the detection accuracy under different environments.

3.1. Target Integration Comparison
Since millimeter-wave radar systems are more accurate in detecting longitudinal information and speed information, and video systems are more accurate in detecting and identifying transverse information and category information, when performing data fusion, dynamically adjusting the noise matrix to adjust the degree of influence of the two sensing systems on the fusion results can complement the advantages of the two sensors, achieve full data utilization, and improve the fusion results.

By adjusting the parameters in the noise matrix, the degree of influence of the observations of the two sensing systems on the fusion result is adjusted. As shown in Fig. 4 and Fig. 5, where the target data includes transverse information, longitudinal information, velocity information, and category information, when the transverse information or category information of the target is obtained by fusion, the influence of the observation value of the video sensor is increased by adjusting the noise matrix; when the longitudinal information or velocity information of the target is obtained by fusion, the influence of the observation value of the millimeter wave radar system is increased by adjusting the noise matrix.
3.2. Practical testing of road junctions

In the actual test, four different scenarios were selected, in which the blue box on the vehicle in the picture shows the radar sensor detection data, the yellow box shows the video sensor detection data, and the green box shows the fusion data.

Fig. 6 shows the detection data under normal traffic condition on a clear day, which enables real-time detection of vehicles and pedestrians with an accuracy of 989 vehicles per 1000 vehicles.

Fig. 7 shows the detection data under normal traffic condition in light rain. The detection at the far end of the lens is mainly achieved by radar, and the detection accuracy is 984 vehicles per 1000 vehicles.

Fig. 8 shows the main detection cross-section of the stopping state in light rain, which enables real-time detection of motor vehicles, pedestrians, and motorcycles, especially when motorcycles are crossing, and the video sensor is a good complement to radar detection.

Fig. 9 shows the normal detection data under background glare, which enables real-time vehicle detection, especially in areas that cannot be detected by remote video, where the millimeter-wave radar sensor is a good complement to video detection.
Figure 6. Target data fusion for clear skies.

Figure 7. Target data fusion for rainy days.

Figure 8. Target data fusion for parking status.
4. CONCLUSION
We propose a detection method using target object association matching, target object tracking, and target object data fusion, and we show how to adjust the detection weights of video and radar sensors by adjusting the noise matrix parameters, which results in a flexible, simple, and efficient architecture for obtaining accurate results in challenging environments. We show the effect on radar detection data and video detection data before and after adjustment. In order to solve the problems of using a single radar or video sensor in the traffic information detection process, such as susceptibility to environmental influences and non-intuitive target reflection, we tested the proposed method by building a hardware verification platform, and finally demonstrated it on video images. The experimental results show that the proposed method can significantly reduce the testing time while providing high detection rate in more environments.

ACKNOWLEDGMENT
This research was supported by 716 grant CY20.007.

REFERENCES
[1] Bourne, Rupert RA, Seth R. Flaxman, Tasanee Braithwaite, Maria V. Cicinelli, Aditi Das, Jost B. Jonas, Jill Keeffe et al. "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis." The Lancet Global Health 5, 9: e888-e897 (2017).
[2] R. Cheng, K. Wang, K. Yang et al., “A ground and obstacle detection algorithm for the visually impaired,” 2015 IET International Conference on Biomedical Image and Signal Processing (ICBISP 2015).
[3] R. Cheng, K. Wang, K. Yang et al., “Real-time pedestrian crossing lights detection algorithm for the visually impaired,” Multimedia Tools and Applications. Papers 1-21, (2017).
[4] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik. Recognizing objects in range data using regional point descriptors. In European Conference on Computer Vision, pages 224–237, 2004.
[5] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In International Conference on Learning Representations, 2017.
[6] R. Klokov and V. Lempitsky. Escape from cells: Deep kd-networks for the recognition of 3D point cloud models. In Proceedings of the IEEE International Conference on Computer Vision, 2017.
[7] T. Linder and K. O. Arras, Multi-model hypothesis tracking of groups of people in RGB-D data, in IEEE Int. Conf. on Information Fusion (FUSION'14), 2014.

[8] S. H. Cen and P. Newman, Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions in Robotics and Automation, 2018. Proceedings, ICRA 2018.

[9] K. Yang, K. Wang, R. Cheng et al., “Detecting Traversable Area and Water Hazards for the Visually Impaired with a pRGB-D Sensor,” Sensors, 17(8), 1890 (2017).

[10] K. Yang, K. Wang, X. Zhao et al., “IR stereo RealSense: Decreasing minimum range of navigational assistance for visually impaired individuals,” Journal of Ambient Intelligence and Smart Environments, 9(6), 743-755(2017).

[11] R. Cheng, K. Wang, K. Yang et al., “Crosswalk navigation for people with visual impairments on a wearable device,” Journal of Electronic Imaging, 26(5), 1-12 (2017).

[12] Yang, Kailun, Kaiwei Wang, Weijian Hu, and Jian Bai. “Expanding the detection of traversable area with RealSense for the visually impaired.” Sensors 16(11), 1954 (2016).

[13] Zöllner, Michael, Stephan Huber, Hans-Christian Jetter, and Harald Reiterer. “NAVI—a proof-of-concept of a mobile navigational aid for visually impaired based on the Microsoft Kinect.” In IFIP Conference on Human-Computer Interaction, 584-587 (2011).