ITM Web of Conferences **45**, 01008 (2022) https://doi.org/10.1051/itmconf/20224501008
CSCNS2021

Information fusion structure of VTS sensors for multi-target tracking of vessels

Jingxue Chen[1,2], *Jie* Wang[1,2,*], and *Hua* Lu[3]

[1]School of Information Science and Engineering, Southeast University, China
[2]National Mobile Communications Research Laboratory, Nanjing, China
[3]Guangdong Communication and Networks Institute, Guangzhou, China

Abstract. Information fusion using VTS sensors of AIS, radar and camera are of great significance to the waterborne traffic supervision. Firstly, this paper invites shore-based CCTV cameras into detection and location of vessel targets combining with bounding boxes generated by deep-learning based detectors. Besides, this paper compares information fusion structure of central-level and track-level in simulated waterborne traffic scenario. Finally, this paper introduces a track selection method for sensors with large false alarm rate to obtain tracks with better performance on both fusion structure when fully considering strengths and weaknesses of all kinds of VTS sensors.

Keywords: VTS sensors, Information fusion, Central-level fusion, Track-level fusion.

1 Introduction

Recently years, waterborne traffic monitoring is mainly carried out by VTS sensors, including AIS, radar and CCTV cameras. However, each sensor has their own limitations. AIS is a passive acquisition system of vessel target information, which has high detection accuracy for detectable targets. However, it cannot work without shipboard equipment, and there is a risk of targets omission. According to [1], AIS is updated very slowly, from fastest 2S to slowest 6min. Although the information update rate of radar is higher than that of AIS, the former has large measurement error and false alarm rate, and even is affected by the inhomogeneity of water environment such as water wave, reverberation edge and extreme weather. CCTV camera has a high information update rate, which can collect abundant information, but it has a narrow coverage. With the development of computer vision, there appears some vessel datasets [2,3], and results show that vessel targets in the image can be detected quickly and accurately. If the camera has been calibrated, combined with the edge box, the camera can detect and locate vessel targets just like radar and AIS. Therefore, information fusion combined with camera is of great significance to maritime traffic supervision.

[*] Corresponding author: wangjie@seu.edu.cn

The existing research on VTS sensor information fusion only includes two types of sensors. These works can be divided into homogeneous fusion of radar and AIS in [4-9] when both sensors can provide the real position of target and heterogeneous fusion of radar and IR / EO sensor in [10] when the latter can only provide azimuth measurement. The fusion of AIS and radar can also be divided into central-level fusion [7-9] and track-level fusion [4-6] according to different fusion structures. The former directly associates and fuses detections from each sensor, and then track with fused detections. The latter generates local tracks from each sensor firstly, then these tracks are assigned and fused to update the center track. These two fusion structures have their own advantages, however, there exists no works to compare them in the case of fully considering the characteristics of VTS sensors.

According to the above analysis, AIS detections has a high reliability, while omission would occur due to the lack of shipborne equipment, radar and camera can detect without shipborne equipment, while false alarms would appear in their range of action. Therefore, when considering the false alarms and omission of detections of these three sensors, information association and fusion methods of radar and AIS will no longer be applicable. The structure of this paper is as follows. The second section introduces the method of detection and location using camera. The third section introduces the central-level fusion method. The fourth section introduces the track-level fusion method. The fifth section introduces the track-quality selection method. The sixth section gives the simulation results and analysis, and the seventh section is the conclusion.

2 Transformation of coordination for camera

Although binocular camera can perfectly locate targets, monocular camera is deployed in VTS and needs MSA staff to stare so far. In recent years, there appears some works on target detection of RGB images. [2] Put forward a dataset of vessels collected artificial and on-the-spot, and the results of [11] show that yolo-v3 can be applied to video surveillance to achieve real-time and accurate detection and classification of a variety of vessels. Detection task produces bounding boxes around vessel targets. The box contains a set of pixel position information which including the upper left corner, upper right corner, width and height of a detection. Combining with edge detection methods like Prewitt approximation, pixel coordinates on waterline of a vessel could be obtained to estimate location of vessel targets. Assume that (0, 0, h) is the real-world coordinates of a camera, the intrinsic matrix and extrinsic matrix is known. the position (X, Y) of a point on the surface of water can be calculated by:

$$\begin{bmatrix} x \\ y \\ \lambda \end{bmatrix} = H^{-1} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{h}{h_1} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{1}$$

$$(X, Y) = \left(\frac{x}{\lambda}, \frac{y}{\lambda}\right) \tag{2}$$

where (u,v) is its image on pixel coordinates system, h_1 denotes the distance between camera and the surface of water when image is taken considering that water level always changes and influences extrinsic matrix.

3 Central-level fusion

Central-level structure fuses detections from all the sensors at a certain time, and tracking with fused detections. The structure of central-level fusion could achieve best performance theoretically, for it can make fully use of all information contained in detections from each sensor. The structure of central-level fusion and tracking are shown in Fig.1.

Different from fusion detections from only AIS and radar, fusion detections from three sensors facing more complex situation. Considering characteristics of these three sensors, AIS detections with high credibility are selected as benchmark and detections from other sensors could not assigned to benchmark are put in fused detections. Cost of a detection assigned with a benchmark is given by:

$$c_{i,j} = \sqrt{(x_i - y_j)(x_i - y_j)'} \tag{3}$$

where vector x_i and y_j denotes location of a benchmark and a detection. Munkres algorithm is applied to obtain an optimal solution of min-cost matching.

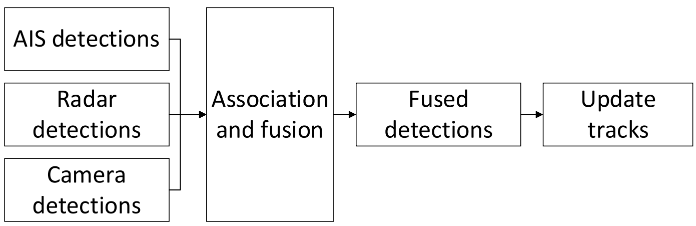

Fig. 1. Structure of Central-level fusion.

4 Track-level fusion

Track-level fusion which aims to fuse information by local tracks from each sensor. Although it may omit information when detections come from track without confirmation, requirement of communication bandwidth is lower than central-level fusion when only state and covariance of local tracks are sent to fuse. The structure of track-level fusion and tracking are shown in Fig.2.

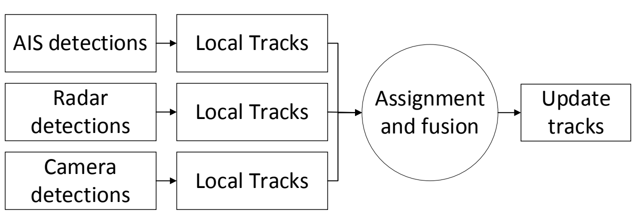

Fig. 2. Structure of track-level fusion.

4.1 Generation and fusion of local tracks

4.1.1 Generation of Local Tracks

Multi-target tracking in single sensor mainly includes four steps, separately are update of tracks, association of tracks and detections, initialization of tracks and detections, and deletion of tracks. Extended Kalman filter with motion mode of constant velocity is applied to update state estimate $\hat{X}(k|k)$ and covariance $\mathrm{P}(k|k)$ of local tracks over time step. Updating with EKF needs new observation Z. When dealing with multi-target tracking problems, Z_{k+1} of a track comes from D_{k+1} assigned to tracks by one-step predictions of

$\hat{X}(k+1|k)$ and $P(k+1|k)$, or just $\hat{X}(k+1|k)$ when a track is unassigned. Normalized distance is applied to measure the cost of a detection assigned to a track. normalized distance d_n is given by:

$$d_n = Z_{res}^T S^{-1} Z_{k+1} + log(|S|) \tag{4}$$

where

$$Z_{res} = Z_{k+1} - h(\hat{X}(k+1|k)) \tag{5}$$

$$S = R + HP(k+1|k)H^T \tag{6}$$

where h and H denotes defined measurement function and corresponding Jacobian of motion mode of constant velocity, and R denotes defined measurement noise covariance. Cost matrix for assignment is obtained by normalized distance between each pair of detection and track, Munkres algorithm is applied to obtain the optimal solution. According to assignment results, new tracks are initialized from unassociated detections and set to temporary in the beginning, and would be confirmed when they were associated more times than confirmation Threshold, and only confirmed tracks would be fused. Deletion occurs when tracks unassigned more than deletion threshold, and deleted tracks are no longer involved in updating and assignment.

4.1.2 Fusion of Local Tracks

Tracks from each sensor are called local tracks and fused tracks are called central tracks here. Only confirmed local tracks can be used to assign or initialize and then update a central track. Similar to tracking in single sensor, generation of center tracks can also be divided into four steps, including update of tracks, association between local tracks and central tracks, initialization and deletion of central tracks. In the update step, constant-velocity extended Kalman filter is used to predict state and covariance of central tracks. If local tracks from different sensors assigned to a central track at the same time, they are fused to update state estimate and covariance of the central track, otherwise the central track is update by a local track when there is only one assignment, or update by its prediction when it is unassigned. Covariance Intersection algorithm in [12] is adopted to fuse state and covariance from local tracks assuming that update rate of camera is faster than radar and AIS and considering that fusion situation is complex when there exists omission of local tracks. The fused state estimate and covariance is given by :

$$\hat{X}_F = \sum_{n=1}^{N} w_n P_F P_n^{-1} \hat{X}_n \tag{7}$$

$$P_F^{-1} = \sum_{n=1}^{N} w_n P_n^{-1} \tag{8}$$

where $\hat{X}_n$ and P_n are state estimate and covariance of sensor n, when they are $\hat{X}(k|k)$ and $P(k|k)$ of a sensor in time step k, and weight w_n is given by :

$$\min_{0 \le w_n \le 1} tr(P_F) = \min_{0 \le w_n \le 1} tr\{(\sum_{n=1}^{N} w_n P_n^{-1})^{-1}\}. \tag{9}$$

In the assignment step, cost matrix is calculated respectively between central tracks and state estimates of local tracks from each sensor using normalized distance, and Munkres algorithm is used to obtain solution of assignment problem. New temporary central track is initialized when a local track is unassigned and the normalized distance from other newly initialized central tracks by other sensors are larger than assignment threshold. If a central track continues unassigned more than deletion threshold, it would be deleted and no longer participate in updating and assignment.

4.2 Track-quality selection method

The results of [10] show that radar does produce some false tracks. Probability of false alarm of radar means each radar resolution cell has the same false alarm report rate after CFAR algorithm is applied, value of that is usually set to 1e-6, while in [13] is set to 1e-4. Different from tracks produced by detections of targets, fortunately confirmed false tracks experiences a smaller number of assignments during the existence time, and the normalized distance of history assignments would be larger and closer to assignment threshold. Therefore, quality Q of a track is calculated by:

$$Q = \sum_{m=sta}^{upd} q_m \tag{10}$$

$$q_m = \begin{cases} e^{-d} & when\ track\ is\ assigned \\ 0 & when\ track\ is\ not\ assigned \end{cases} \tag{11}$$

where sta denotes the time steps when the track initialized, upd denotes the latest time step, d denotes normalized distance between the track and its assignment at the time step m. A track is sent to fusion Only when its quality is large enough, and it would be suitable when the threshold is set to 1e-10 after numerous simulations.

5 Simulation results

The experimental simulation is carried out on the platform of MATLAB r2020b on Windows 10 Intel(R) Core (TM) i7-4790K CPU @4.00GHz, RAM 32.0 GB. A waterborne traffic scenario including three sensors which respectively are AIS, radar and camera is simulated. Assume that time alignment and coordinate transformation are completed before, Detection probability of detectable targets for AIS, radar and camera are set to 0.8, update rate of AIS and radar are set to 0.5 and that of camera is set to 1, false alarm probability of radar is set to 1e-4 according to [13]. Radar and camera are installed 20 meters away from the shore, field of view of radar is 180 degrees and that of camera is 100 degrees. The scenario includes 150 seconds, and the length of the waterborne on Y-axis is 1000 meters. According to waterborne traffic situation in Nanjing section of the Yangtse River, the width of waterborne on X-axis is set to 1000 meters and divided into main waterways, auxiliary waterways and medial divider, and large vessels should drive on waterways and keep to the right. There are 11 vessels set in the scenario, including 2 vessels in auxiliary waterway of [50,250] driving in opposite direction of Y-axis and one of which is undetectable to AIS, 2 vessels in main waterway of [250,450] driving in opposite direction of Y-axis, 3 vessels in main waterway of [550,750] driving in same direction of Y-axis, 2 vessels in auxiliary waterway of [750,950] driving in same direction of Y-axis and one of which is undetectable to AIS and camera, and 2 vessels driving across the scenario with the trajectory approximate to quarter arc and is undetectable to AIS. Ground truth of targets and annotations are shown in Fig.3.

OSPA proposed in [14] is adopted to measure the effect of tracking. Different from RMSE which is often used as indicator for tracking effect for single target and can only measure the error between track and ground Truth, OSPA is proposed for multi-target tracking task. OSPA fully considers the localization error, cardinality error response to omission of tracks and labelling error response to incorrect assignment, and is suitable for measuring the performance of tracking in the scenario. OSPA curve indicates the difference between tracks and ground truths in the scenario over time, and is closer to zero when tracks perfectly fit ground truth, which means lower OSPA would be better.

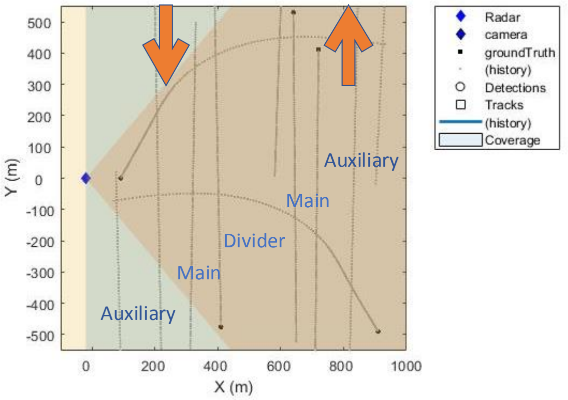

Fig. 3. Ground truth of vessel targets, division of waterways and driving direction.

5.1 Tracks from single sensor

Fig.4 shows simulation results of local tracks from AIS, radar, camera to show characteristics of these sensors. AIS tracks are closest to ground truth of detectable targets, while there are 4 vessels that is undetectable by AIS in total. Radar generates large amounts of false alarms with 1e-4 probability of false alarm, and produces 4 false tracks in total. Camera suffers from a narrow field of view and causes loss of tracks.

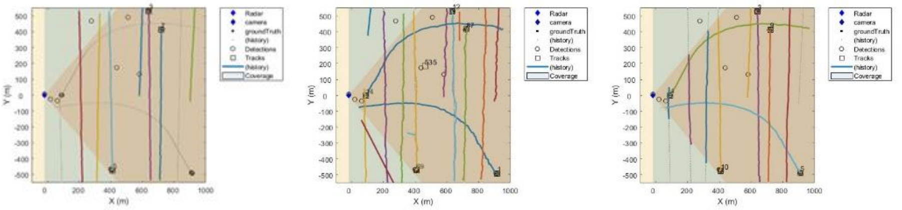

Fig. 4. Local tracks and corresponding OSPA curves from AIS, radar and camera. Left: AIS. Centre: radar. Right: camera.

5.2 Fusion without track-quality selection

Fig.5 shows simulation results of central-level fusion and track-level fusion by AIS, radar and camera without track selection method. Although OSPA of central-level fusion some times better than track-level fusion, the former does introduce more false tracks then the latter. The possible reason is that fused detections in central-level contain all false alarms from each sensor, and thus temporary track generated by false alarm from one sensor could be confirmed by false alarms from any sensor in follow time steps, and has the risk to generate larger amounts of false alarms.

5.3 Fusion with track-quality selection

Track-quality selection is applied on tracks in the case of central-level fusion, when it is applied on local tracks in the case of track-level fusion. Fig.6 shows simulation results of central-level fusion by AIS and radar, central-level fusion by AIS, radar and camera, track-level fusion by AIS and radar and track-level fusion by AIS, radar and camera with track selection method. OSPA curves show that fusion by AIS, radar and camera always achieve better results than fusion by the first two, thus camera would be applied as an important supplement to radar and AIS in waterborne traffic supervision. Besides, track-level fusion with track-quality on local sensor achieves best tracking results.

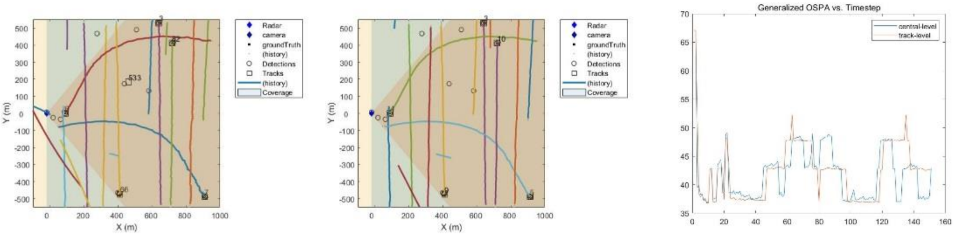

Fig. 5. Fusion without Track-quality selection of two structures. Left: central-level fusion. Centre: track-level fusion. Right: OSPA curve.

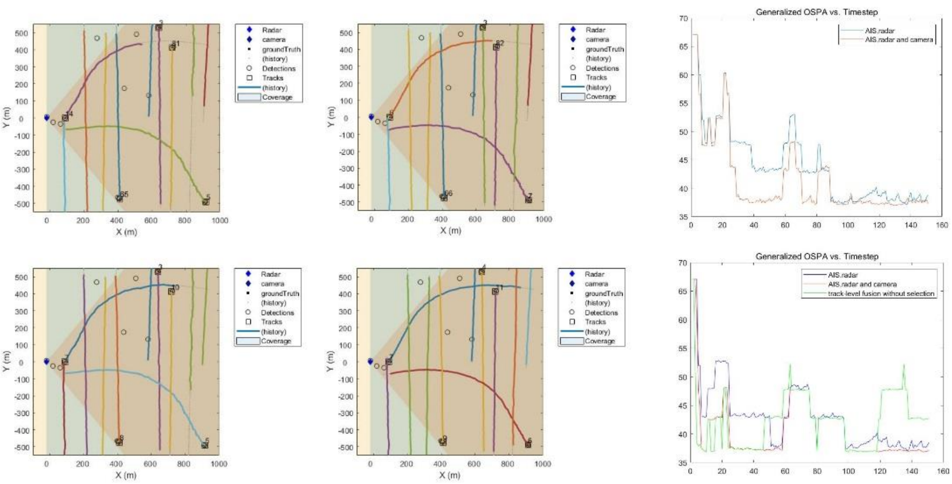

Fig. 6. Fusion with track-quality selection. First line: central-level fusion, left: AIS and radar, centre: AIS, radar and camera, right: corresponding OSPA curve. Second line: track-level fusion, left: AIS and radar, centre: AIS, radar and camera, right: corresponding OSPA curve.

6 Conclusion

This paper introduces the current research status of water traffic information fusion and the characteristics of VTS sensor, and introduces the target location method of CCTV camera combined with deep learning detector. Then this paper simulates central-level fusion and track-level fusion in a waterborne scenario. The simulation results show that fuse camera with AIS and radar can improve the tracking effect of vessel targets, and track-level tracking combined with track-quality selection can obtain the best tracking effect. There are two main directions for our future work. The first is the comprehensive application of the vessel detection and positioning method in the shore based closed-circuit television camera network. In addition, we are going to integrate the shore-based camera network with AIS and radar information, and combine the actual waterway and weather conditions to realize the real-time monitoring and navigation of vessel safety. Due to the lack of multi-source data, we have to wait for a comprehensive data set to continue our work in the future.

This work was supported by the Guangdong-Hong Kong-Macao Greater Bay Area next-generation communication and network application demonstration (No.: 2019YFB1804400), 2019 Key projects of Broadband communication and new network, National Key R&D Program of China.

References

1. ITU-R M.1371-5. "Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band." Feb 2014
2. Z. Shao, W. Wu, Z. Wang, W. Du and C. Li, "SeaShips: A Large-Scale Precisely Annotated Dataset for Ship Detection," in IEEE Transactions on Multimedia, vol. 20, no. 10, Oct. 2018, pp. 2593-2604.
3. D. Qiao, G. Liu, F. Dong, S. Jiang and L. Dai, "Marine Vessel Re-Identification: A Large-Scale Dataset and Global-and-Local Fusion-Based Discriminative Feature Learning," in IEEE Access, vol. 8, pp. 27744-27756, 2020, doi: 10.1109/ACCESS.2020.2969231.
4. X. Qin, M. Yan and D. Zhu, "Research on information fusion structure of radar and AIS," 2018 Chinese Control And Decision Conference (CCDC), 2018, pp. 3316-3322, doi: 10.1109/CCDC.2018.8407697.
5. D. Chen, P. Chen and C. Zhou, "Research on AIS and Radar Information Fusion Method Based on Distributed Kalman," 2019 5th International Conference on Transportation Information and Safety (ICTIS), 2019, pp. 1482-1486, doi: 10.1109/ICTIS.2019.8883594.
6. Zhenjie Xu, Jiacheng Li and Yunfang Chen, "Survey of track association of radar and AIS," 2017 2nd International Conference on Image, Vision and Computing (ICIVC), 2017, pp. 960-964, doi: 10.1109/ICIVC.2017.7984697.
7. XING Xu-feng, XIE Shi-yi, HUANG Miao-fen, YANG Guang-zhao and HUANG Shan. A Target Fusion Algorithm Based on offshore Radar and AIS sensors Wave-Measuring Sensor Buoy[J]. Journal of Ocean Technology, 2020, 39(03): 8-15.
8. Zhou Qiuhua. Study on AIS and Radar Information Fusion Based on Fuzzy Comprehensive Function [D]. Dalian Maritime University, 2013.
9. Hu Xiaorui. Research on Target Information Fusion of Radar and AIS Based on BP Neural Network[D]. Jimei University, 2012.
10. Ø. K. Helgesen, E. F. Brekke, H. H. Helgesen and Ø. Engelhardtsen, "Sensor Combinations in Heterogeneous Multi-sensor Fusion for Maritime Target Tracking," 2019 22th International Conference on Information Fusion (FUSION), 2019, pp. 1-9.
11. X. Nie, M. Yang and R. W. Liu, "Deep Neural Network-Based Robust Ship Detection Under Different Weather Conditions," 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 2019, pp. 47-52, doi: 10.1109/ITSC.2019.8917475.
12. S. Matzka and R. Altendorfer, "A comparison of track-to-track fusion algorithms for automotive sensor fusion," 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2008, pp. 189-194, doi: 10.1109/MFI.2008.4648063.
13. S. Maresca, A. Bogoni and P. Ghelfi, "CFAR Detection applied to MIMO Radar in a Simulated Maritime Surveillance Scenario," 2019 16th European Radar Conference (EuRAD), 2019, pp. 157-160.
14. B. Ristic, B. Vo, D. Clark and B. Vo, "A Metric for Performance Evaluation of Multi-Target Tracking Algorithms," in IEEE Transactions on Signal Processing, vol. 59, no. 7, pp. 3452-3457, July 2011, doi: 10.1109/TSP.2011.2140111.