Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 10
ISET 2025

Collaborative Optimization of Multi Modal Sensing Fusion and Visual Navigation

Pengyuan He[1 *], Jiachen Li[2], and Yingjie Mao[3]

[1]Electronical Information Engineering, Hebei University of Science & Technology, 050000, Shijiazhuang, China

[2]Electronic Information Engineering, Jiangxi Agricultural University, 330045, Nanchang, China

[3]Civil Engineering, Qiqihar University, 161000, Qiqihar, China

* Corresponding Author Email: 1814010917@stu.hrbust.edu.cn

Abstract. With the rapid advancement of intelligent systems such as autonomous vehicles and drones, multi-modal sensing fusion has emerged as a pivotal approach to enhance the robustness and accuracy of visual navigation systems. Traditional single-sensor solutions, including GNSS, IMU, and vision-based methods, face inherent limitations such as signal interference, error accumulation, and sensitivity to lighting conditions. This study proposes a collaborative optimization framework that integrates multi-modal data (e.g., LiDAR, radar, RFID, and IMU) through spatiotemporal alignment, feature complementarity, and joint optimization. Innovations in fusion architectures—centralized (e.g., Extended Kalman Filter) and distributed (e.g., hierarchical decision fusion), achieving up to 40% localization error reduction. Deep learning techniques, such as multi-task neural networks, further enhance cross-modal feature distillation, improving 3D detection accuracy by 15% on KITTI datasets. Practical applications in autonomous driving, UAV navigation, and special environments demonstrate the system's adaptability, with sub-meter positioning accuracy in GNSS-denied environments and decimeter-level precision in indoor SLAM. Challenges such as dynamic adaptability, heterogeneous data alignment, and edge computing optimization are discussed, alongside future directions including adversarial learning frameworks, lightweight model deployment, and cross-modal transfer learning. This research provides a comprehensive pathway to advance the reliability and applicability of multi-modal fusion systems in complex real-world scenarios.

Keywords: Single sensor, autonomous vehicle, vision navigation

1. Introduction

With the rapid development of intelligent systems in fields such as autonomous driving and drone navigation, the collaborative optimization of multi modal sensing fusion and visual navigation has become the core path to improve environmental perception accuracy. In modern navigation and perception systems, a single sensor has significant limitations. For example, GNSS (Global Navigation Satellite System) relies on satellite signals and is easily affected by factors such as building obstruction and signal interference, making it unable to function properly in environments such as tunnels and underground parking lots; IMU (Inertial Measurement Unit) measures motion through accelerometer and gyroscopes, but there is error accumulation. However, pure visual navigation is sensitive to lighting conditions, and its performance will significantly decrease under strong light, darkness, or dynamic object interference.

Multi modal sensing fusion can enhance the robustness of navigation systems through spatiotemporal alignment, feature complementary, and joint optimization. For example, by integrating GNSS/MEMS-IMU with visual sensors, the accuracy of drone attitude estimation can be improved to 0.1° [1]. Multi-task neural network achieves deep coupling between visual and LiDAR data, synchronously improving 2D/3D detection accuracy by 15% on the KITTI dataset, significantly optimizing the system's perception ability of the environment [2]. Based on this, this study will focus on building a bidirectional optimization mechanism of "sensor fusion-visual enhancement", aiming to comprehensively improve the environmental perception accuracy and overall robustness of the

system in complex environments, so that the navigation system can operate stably and accurately in diverse scenarios.

2. Research progress on multi modal fusion driven visual navigation optimization

2.1. Innovation in fusion system architecture

The centralized architecture adopts a unified processor to fuse multi-source data, such as the Extended Kalman Filter (EKF), which has high spatiotemporal alignment accuracy and is suitable for static environments. Taking the fusion of visual SLAM (Simultaneous Localization and Mapping) and odometer as an example, in actual testing, by reasonably configuring weights and optimizing algorithms, the localization error can be reduced by 40%, significantly improving the accuracy of localization [3].

Distributed architecture improves real-time performance through hierarchical processing of sensor data (e.g., radar → camera → LiDAR) and decision fusion. It is more suitable for dynamic environments and has strong scalability. For example, first process radar data to obtain distance information, then combine camera images to recognize objects, and finally integrate 3D point cloud data from LiDAR to construct a detailed map. In some highway scenario tests, FBMDM based on distributed architecture (a decision-level fusion method) can achieve 200-meter perception coverage and effectively advance decision time [4].

The new multitasking framework optimizes multiple tasks through end-to-end learning systems, such as simultaneous 3D object detection and depth information completion [2]. By sharing the underlying feature extraction layer, computational efficiency is improved and task performance is enhanced, reducing manual intervention.

2.2. Visual navigation optimization method

Multi modal fusion technology has achieved significant improvements in environmental perception enhancement, motion estimation optimization, and robustness enhancement to address the inherent deficiencies of visual navigation systems.

Enhanced environmental perception: By integrating RFID (Radio Frequency Identification) and infrared sensors, a centimeter-level spatial benchmark is provided for the visual system. For example, the intelligent car navigation system proposed by Bijamwar significantly improves the accuracy and reliability of path recognition in complex scenarios [5].

Motion estimation optimization: Graph optimization technology is adopted to integrate GNSS (Global Navigation Satellite System) phase measurement and IMU (Inertial Measurement Unit) data, effectively reducing position estimation errors. Sabatini's team's research shows that this solution can control the position error of drones within 1 meter [1, 6], significantly improving the positioning accuracy of the navigation system.

Robustness improvement: Visual Inertial Odometry (VIO) technology enhances the system's continuous positioning capability in GNSS-denied environments by integrating visual data with IMU. Experimental verification shows that VIO can maintain continuous positioning for at least 5 seconds in the event of GNSS signal interruption, providing reliable support for navigation in extreme environments.

Multi modal fusion technology integrates visual, RFID, infrared sensors, GNSS, and IMU, significantly optimizing the performance of visual navigation systems. At the level of environmental perception, the fusion of RFID and infrared sensors provides high-precision spatial benchmarks, enhancing path recognition capabilities in complex scenarios; in terms of motion estimation, the graph optimization collaboration between GNSS and IMU controls the positioning error within 1 meter, improving dynamic positioning accuracy; the Inertial Visual Tight Coupling (VIO) technology maintains continuous positioning for at least 5 seconds in GNSS-denied environments through deep

fusion of IMU and visual data, greatly enhancing the robustness of the system. These methods collectively address the inherent shortcomings of visual navigation and provide key technical support for reliable navigation in complex environments.

2.3. Evolution of fusion algorithms

In the development process of multi modal fusion technology, the evolution of multi modal fusion algorithms is crucial for achieving deep coupling between sensing and vision. Multi-sensor data fusion can obtain more inferences than a single sensor by integrating information from different sources, and is widely used in fields such as remote sensing. From traditional remote sensing image fusion algorithms such as PCA and IHS, to the development of multi modal fusion algorithms today, the two have similar exploration trajectories and evolutionary logic [7]. This evolution process started with traditional filtering methods, went through the deep learning fusion stage, and finally led to the proposal of hybrid architectures, each step of which has significantly promoted the improvement of multi modal data processing capabilities and system performance.

Traditional filtering methods: Extended Kalman Filter (EKF), as a classic fusion algorithm, plays an important role in the tight coupling of visual SLAM (Simultaneous Localization and Mapping) and IMU (Inertial Measurement Unit) data [3]. EKF utilizes linearization assumptions to fuse image information obtained by visual sensors with inertial data collected by IMU. It achieves high computational efficiency by continuously predicting and updating state estimation values, and can quickly provide key information such as position and attitude to the system [7]. However, when the system is in a highly nonlinear scenario, it can lead to highly nonlinear changes in visual information and inertial data, and the linear assumption of EKF no longer holds, resulting in a decrease in data processing accuracy and significant performance limitations.

Deep learning fusion: Multi modal data fusion, as a fundamental method of multi modal data mining, aims to integrate data of different distributions, sources, and types into a global space for unified representation of inter-modal and cross-modal information. It can use specific modal information to provide more abundant information than a single modal [8]. Multi-task neural networks significantly improve object detection accuracy through cross-modal feature distillation, such as the fusion of visual and LiDAR data. For example, the model proposed by Liang achieved a 15% mAP (mean Average Precision) improvement on the KITTI dataset, validating the effectiveness of deep learning in cross-modal data fusion [9].

Hybrid architecture trend: To further enhance the dynamic adaptability of the system, researchers have proposed a "learning+optimization" hybrid framework. For example, Zhuang dynamically allocated sensor weights through reinforcement learning, enabling the system to adjust fusion strategies in real-time according to environmental changes, thereby maintaining high robustness in complex scenarios [10].

3. Fusion practice in typical scenarios

3.1. Auto drive system

Multi modal sensor fusion technology has played an important role in the auto drive system, breaking through the limitations of the pure vision scheme. Firstly, by combining radar, cameras, and LiDAR, all-weather environmental perception has been achieved [4]. Radar can still provide reliable ranging data under complex weather conditions such as rain, fog, and strong light, while cameras are responsible for object recognition, and LiDAR generates high-precision 3D environmental models. This multi-sensor redundancy design significantly enhances the system's environmental adaptability.

In terms of decision optimization, the combination of V2X (Vehicle to Everything) communication technology and 3D detection systems provides efficient path planning support for L4-level autonomous driving. In addition, the fusion of RFID and infrared sensors provides an anti-

interference path benchmark for visual systems [5], effectively avoiding visual misjudgment and further enhancing the security of the system.

3.2. UAV navigation system

In the field of drone navigation, multi modal fusion technology has demonstrated its advantages, especially in GNSS-denied environments. The solution proposed by the Sabatini team can maintain sub-meter positioning accuracy even in the event of GNSS signal interruption by integrating visual and IMU data [1]. In addition, the application of deep completion technology converts monocular visual data into pseudo-LiDAR point clouds [2], solving the problem of feature association in sparse environments and significantly improving the navigation capability of drones in complex environments.

3.3. Special environment navigation

Multi modal fusion technology also expands the application boundaries of visual navigation systems in special environments. In underwater navigation scenarios, the Bluefin-21 system achieves high-precision navigation with a range error of only 0.1% by integrating MEMS-IMU and Doppler velocimeter. In indoor positioning scenarios, the fusion of RFID and magnetic field sensors provides decimeter-level landmark references for visual SLAM (Simultaneous Localization and Mapping) [10], effectively solving the problem of GPS signal failure and providing reliable technical support for indoor autonomous navigation.

4. Challenges and future expectations

4.1. Key technological breakthrough direction

Although multi modal sensing fusion technology has made significant progress in visual navigation optimization, it still faces many challenges. Firstly, dynamic adaptability is one of the current research focuses. To cope with interference such as GPS spoofing [6], a framework based on adversarial learning needs to be developed to enhance the robustness of the system in complex environments. Secondly, the problem of heterogeneous data alignment urgently needs to be solved. The standardization of pseudo-LiDAR generation and deep completion techniques will provide new ideas for achieving efficient fusion of cross-modal data [2, 3]. In addition, edge computing optimization is also the core direction of future research. By combining lightweight models such as MobileNet with FPGA (Field Programmable Gate Array) acceleration technology, the real-time performance of the system can be significantly improved to meet practical application requirements [1].

4.2. System development path

To promote the widespread application of multi modal fusion technology, a clear development path needs to be formulated at the system level. Firstly, establishing a standardization and certification system is crucial. For example, referring to the ISO 26262 functional safety certification standard, develop evaluation criteria for multi modal fusion systems to ensure their safety and reliability. Secondly, the development of anti-interference technology is the key to improving system performance. By developing M-code spot beam encryption technology, the anti-interference ability of GNSS can be effectively enhanced [10]. Finally, the introduction of small sample learning techniques will reduce reliance on annotated data. The application of cross-modal transfer learning is expected to enhance the adaptability of the system in small-sample scenarios and further expand its scope of application.

5. Conclusion

This study systematically explores the collaborative optimization of multi-modal sensing fusion and visual navigation, addressing the inherent limitations of single-sensor systems in dynamic and

complex environments. By integrating diverse sensors (e.g., LiDAR, radar, GNSS, IMU, and RFID) through spatiotemporal alignment, feature complementarity, and hybrid architectures, the proposed framework achieves substantial advancements in environmental perception, motion estimation, and system robustness. Key contributions include:

5.1. Architectural innovation

Centralized (e.g., EKF-based) and distributed fusion systems reduced localization errors by up to 40% in static and dynamic scenarios, while hierarchical decision fusion extended perception coverage to 200 meters in autonomous driving.

5.2. Algorithmic advancements

Deep learning techniques, such as multi-task neural networks, improved cross-modal feature distillation, boosting 3D detection accuracy by 15% on KITTI datasets. Hybrid "learning+optimization" frameworks enhanced dynamic adaptability through reinforcement learning-based sensor weight allocation.

5.3. Practical validation

Applications in autonomous vehicles, UAVs, and special environments (underwater, indoor) demonstrated sub-meter positioning accuracy in GNSS-denied settings and decimeter-level precision in visual SLAM, proving the framework's versatility.

5.4. Challenges

Challenges such as heterogeneous data alignment, edge computing efficiency, and adversarial interference mitigation remain critical. Future work will focus on lightweight model deployment (e.g., MobileNet-FPGA integration), adversarial learning for robustness enhancement, and cross-modal transfer learning to reduce dependency on annotated data. This research establishes a foundation for reliable, adaptable, and scalable multi-modal fusion systems, paving the way for next-generation intelligent navigation in real-world applications.

Author Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] R. Sabatini, S. Ramasamy, Gardi, A., et al. Low-cost sensors data fusion for small size unmanned aerial vehicles navigation and guidance. International Journal of Unmanned Systems Engineering, 1(3), 16-47 (2013)

[2] G.T. Schmidt. Navigation sensors and systems in GNSS degraded and denied environments. Chinese Journal of Aeronautics, 28(1), 1–10 (2015). https://doi.org/10.1016/j.cja.2014.12.001

[3] N. Karam, H. Hadj-Abdelkader, C. Deymier, et al. Improved visual localization and navigation using proprioceptive sensors. IEEE/RSJ International Conference on Intelligent Robots and Systems, 4155–4160 (2010). https://doi.org/10.1109/IROS.2010.5649978

[4] Z. Wang, Y. Wu, and Q. Niu. Multi-sensor fusion in automated driving: A survey. Ieee Access, 8, 2847-2868.

[5] M. M. Bijamwar, ang P. S. S. Savkare. Design and implementation of smart car with self-navigation and self-parking systems using sensors and RFID technology. Int. J. Eng. Res. Gen. Sci., 4(3), 305-308 (2011)

[6] J. Dong, D. Zhuang, Y. Huang, J. Fu "Advances in Multi-Sensor Data Fusion: Algorithms and Applications." Sensors, vol. 9, no. 10, pp. 7771–7784 (2009). https://doi.org/10.3390/s91007771

[7] T. Bailey, J. Nieto, J. Guivant, M. Stevens, & E. Nebot. Consistency of the EKF-SLAM algorithm. IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3562-3568 (2006)

[8] J. Gao, P. Li, Z. Chen, & J. Zhang. A Survey on Deep Learning for Multimodal Data Fusion. Neural Computation, 32(5), 829–864 (2020). https://doi.org/10.1162/neco_a_01273

[9] M. Lian, B. Yang, Y. Chen, R. Hu, & R. Urtasun. Multi-task multi-sensor fusion for 3d object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7345-7353, (2019).

[10] Y. Zhuang, X. Sun, Y. Li, J. Huai, L. Hua, X. Yang, X. Cao, P. Zhang, Y. Cao, L. Qi, J. Yang, N. El-Bendary, N. El-Sheimy, J. Thompson, & R. Chen. Multi-sensor integrated navigation/positioning systems using data fusion: From analytics-based to learning-based approaches. Information Fusion, 95, 62–90, (2023). https://doi.org/10.1016/j.inffus.2023.01.025