Cross-Domain Applications of Multi-Dimensional Data Alignment Technology

Fengshuo Kou

School of Computer and Control Engineering, Electronic Information Category, Northeast Forestry University, Harbin, Heilongjiang, China
kfs60113@outlook.com

Abstract. Against the backdrop of the rapid development of intelligent technology, cross-modal data fusion is driving innovations across various fields. As a key supporting technology, 2D and 3D data alignment technology is becoming increasingly prominent in value. 2D data (e.g., camera images) is rich in semantic and visual details but lacks spatial depth, while 3D data (e.g., LiDAR point clouds) can accurately represent the spatial structure of objects yet has sparse semantic information. This paper focuses on this alignment technology, conducting in-depth research on its application scenarios, implementation methods, experimental results, and literature support in five fields: autonomous driving, robot navigation, augmented reality, indoor scene construction, and medicine. Through comparative analysis, the advantages, disadvantages, and common challenges of the technology are clarified, and future development directions are proposed. This research provides reference and support for optimizing cross-modal tasks in various fields. This article aims to promote practical innovation and long-term development of intelligent technology.

Keywords: 2D and 3D Data Alignment, Cross-Domain Application, Autonomous Driving, Robot Navigation, Augmented Reality.

1 Introduction

With the in-depth integration of artificial intelligence, big data, and the Internet of Things technologies, multi-modal data has been widely applied in various aspects of production and daily life. Among them, 2D and 3D data are two core forms. Represented by camera images, 2D data can clearly present the visual features (such as the appearance, color, and texture of objects) and rich semantic information (e.g., recognizing traffic signs and pedestrians' clothing) of objects. However, restricted by its planar nature, it cannot convey spatial depth and 3D dimensional relationships, making it difficult to meet the requirements of tasks that have high demands for spatial perception, such as autonomous driving and robot navigation.

3D data, mainly consisting of LiDAR point clouds and 3D scanner models, can accurately describe the spatial position, shape, and volume of objects (e.g., real-time measurement of the distance and height difference between vehicles and obstacles). Nevertheless, it has sparse semantic information, making it hard to directly distinguish

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_26

between similar categories like "pedestrians" and "non-motorized vehicles". Additionally, data acquisition and processing rely on high-precision equipment and complex algorithms, resulting in high costs.

The complementary characteristics of the two types of data have promoted the development of 2D and 3D data alignment technology. By means of algorithm optimization or spatial transformation, this technology establishes accurate corresponding relationships in terms of spatial position, scale, and perspective between the two types of data. It is like building a "bridge" that breaks down the barriers between "planar semantics" and "spatial structure", enabling efficient complementary utilization of cross-modal information.

In recent years, this technology has been initially applied in multiple fields. However, there are significant differences in dataset characteristics, application requirements, and environmental conditions across different fields (e.g., autonomous driving requires real-time performance, while medicine demands high precision). Therefore, systematically sorting out its application status is of great significance for technology optimization and cross-domain promotion. Following the logic of "scenario introduction → method elaboration → result analysis → literature support", this paper analyzes the application details of the technology in the five aforementioned fields, summarizes common challenges, and puts forward future directions.

2 Data Alignment in the Field of Autonomous Driving

In autonomous driving, a single sensor cannot meet the needs of complex environmental perception: cameras are susceptible to illumination and weather conditions, while LiDAR suffers from semantic information deficiency. The 2D and 3D data alignment technology achieves "semantics + spatial" dual guarantees by fusing data from the two types of sensors, serving as a core link in multi-sensor fusion.

2.1 Dataset Optimization and Error Reduction

The nuScenes dataset contains 1000 20-second scenarios, annotated with 3D bounding boxes covering 23 categories and 8 attributes. The number of annotations and images is 7 times that of the KITTI dataset [1]. The average reprojection error of its original annotations reaches 8.03 pixels, which affects perception accuracy. Researchers have optimized it to form the CAMAv2 version: through "LiDAR-triggered camera exposure" to achieve time synchronization, the error is reduced to 4.96 pixels; the perception model trained with this version sees the reprojection error drop from 8.43 pixels to 5.62 pixels, significantly improving mapping accuracy [1].

2.2 Alignment Process for Obstacle Detection

This dataset provides multi-category annotations and difficulty labels, making it suitable for obstacle detection [2]. The alignment process consists of three steps: converting LiDAR range images into 3D point cloud coordinates; mapping 2D segmentation labels to 3D points based on calibration parameters; and ensuring that the data corresponds to the same moment through time synchronization. Experiments show that the accuracy of pedestrian recognition is improved by 12%, and the vehicle positioning error is ≤ 0.08 meters [2].

2.3 Active Alignment Modules for Autonomous Driving

Designed specifically for active alignment in autonomous driving [3], it includes two modules: the depth prediction branch predicts depth through 2D image features to establish data association; the BEV-to-Camera module converts point clouds into the camera coordinate system. Its performance in generating multi-view data ranks first in the industry, with the high-quality point cloud generation rate increased by 35% and the alignment delay in dynamic scenarios ≤ 0.05 seconds [4].

2.4 Voxel-Level Alignment with Semi-Automatic Annotation

Voxel-level alignment is achieved using semi-automatic annotation, involving three steps: voxel densification to enhance details, occlusion reasoning to handle occlusion issues, and image-guided voxel refinement to optimize data [5]. Each step undergoes 3D-2D consistency verification. The accuracy of scene occupancy prediction reaches 89%, which is 15% higher than that of traditional methods [5].

3 Data Alignment in the Field of Robot Navigation

Robot navigation requires a closed loop of "localization → planning → obstacle avoidance". Cameras can identify navigation markers but cannot measure distances, while LiDAR can measure distances but struggles to classify the types of markers. The alignment technology integrates the two types of data to ensure perception consistency, and is applicable to scenarios such as indoor warehousing and outdoor inspection.

3.1 Real-Time Calibration of Camera and LiDAR

Suitable for real-time calibration of cameras and LiDAR [6], the process is as follows: extracting corner points and edges from 2D images, as well as surface normals and curvatures from 3D point clouds; eliminating abnormally matched points through the RANSAC algorithm; optimizing and calculating transformation parameters. Experiments show that the translation error is ≤ 0.05 meters, the rotation error is ≤ 0.5 degrees, and the matching success rate in dynamic scenarios (1.5 m/s) reaches 98.7% [6].

3.2 Alignment for Indoor Obstacle-Prone Environments

Aimed at indoor environments with numerous obstacles and variable illumination [7], the approach first constructs a 3D map annotated with key features; during navigation, cameras and LiDAR collect data, extract features, match them with the map, and calculate transformation relationships. In a 1000 m^2 warehouse experiment, the alignment accuracy is ≤ 0.1 meters, and the time consumption is < 0.5 seconds, meeting real-time requirements [7].

3.3 Alignment for Outdoor Complex Terrain

To address the problems of large outdoor space and complex terrain [8], GPS constraints are used to narrow the parameter search range; feature points of the data are extracted and denoised via RANSAC; parameters are optimized through the least squares method. Even with slight GPS interference, the alignment accuracy remains ≤ 0.2 meters, and the robustness to weather changes is 20% higher than that of traditional methods [8].

4 Data Alignment in the Field of Augmented Reality

The core requirement of augmented reality (AR) is the "natural integration of virtual and real". The alignment technology enhances immersion by establishing spatial correlations between virtual objects and the real environment, and is applicable to scenarios such as facial AR and industrial assembly guidance.

4.1 Alignment of 2D Facial Video and 3D Facial Mesh

Suitable for matching 2D videos with 3D facial meshes [9], the steps are: constructing a 3D facial mesh containing 100,000 triangular patches, and parameterizing 2D virtual textures onto the mesh through UV coordinates; tracking facial key points in real time, adjusting the mesh posture, and updating the mapping. Experiments show that the edge view error is ≤ 3.5 mm, supporting 30 FPS rendering, the fitting degree of virtual makeup reaches 96%, and the user experience is improved by 25% [9].

4.2 Alignment for Industrial AR Assembly Guidance

First, 3D modeling is performed on industrial equipment and key feature points are marked; a camera is used to capture 2D images, and feature points are extracted through the SIFT algorithm; 2D and 3D feature points are matched, and the transformation matrix is calculated [10]. In the automobile engine assembly experiment, the alignment error is < 0.5 mm, efficiency is improved by 30%, and the error rate is reduced by 18% [10].

4.3 VIO-Assisted Alignment for Mobile AR Navigation

To solve the problems of equipment movement and occlusion [11], mobile phone cameras and IMU (Inertial Measurement Unit) are utilized: cameras collect image sequences, and IMU records motion parameters; data is fused through VIO (Visual-Inertial Odometry) to estimate trajectories and postures, construct sparse point clouds, and associate features; alignment relationships are dynamically adjusted. In outdoor navigation experiments, the error during walking (1.5 m/s) is < 1 meter, and there is no significant decrease in stability under 30% occlusion [11].

5 Data Alignment in the Field of Indoor Scene Construction

Indoor scene construction requires accurate restoration of spatial structures and object layouts. 2D data (design drawings, photos) lacks depth, while 3D data (point clouds) is deficient in appearance details. The alignment technology maps 2D details to 3D models, improving realism and shortening the modeling cycle. It is applicable to scenarios such as home design and virtual exhibition halls.

5.1 Alignment Based on SceneVerse Dataset

SceneVerse contains 68,000 indoor scenes and 2.5 million visual-language pairs [12]. The GPS framework designed based on it determines the overall relationship through global matching and optimizes details through local sampling. Experimental verification shows that when the number of scenes increases from 10,000 to 68,000, the alignment accuracy rises from 78% to 92% [12].

5.2 Alignment Using SUN RGB-D Dataset

It includes 10,335 indoor images, storing 2D and 3D bounding boxes respectively [13]. The alignment method is as follows: extracting bounding boxes from 2D images and 3D point clouds; matching based on size ratio and center position; when the IOU threshold is 0.7, the matching accuracy reaches 89%, and the home modeling error is ≤ 5 mm [13].

5.3 Alignment for Low-Overlap Indoor Scenes

Aimed at scenes with low overlap (narrow corridors, rooms with multiple occlusions) [14], the approach first enhances the geometric features of point clouds and converts images into grayscale histograms and edge maps; a strategy of "global rough matching + local fine matching" is adopted. Experiments show that the registration success rate reaches 52.6%, which is 30% higher than that of the ICP algorithm; meanwhile, the 3DZeroMatch dataset is constructed to support research on extreme scenarios [14].

6 Data Alignment in the Field of Medicine

In medicine, 2D images (CT, ultrasound, X-ray) lack 3D morphology, while 3D models are difficult to associate with subtle lesions. The alignment technology enables the collaborative analysis of "details + structure", providing support for diagnosis and surgical planning, and reducing medical risks.

6.1 Alignment for Abdominal Organ Segmentation

Converting 3D organ segmentation into 2D surface evolution [15]: extracting organ features from abdominal CT slices and constructing mapping graphs; introducing deep learning to optimize contours based on the Snakes model; stacking 2D contours to generate 3D models and verifying consistency. In the spleen segmentation experiment, the Dice Similarity Coefficient (DSC) reaches 0.94 (clinical high-quality standard), the computational efficiency is 90% higher than that of traditional methods, and the time consumed for processing 500 layers of CT is < 2 minutes [15].

6.2 Alignment of 2D Brain MRI and 3D Brain Model

To address the problems of complex brain structure and large individual differences [16], MRI and 3D models are preprocessed first (grayscale normalization, rigid transformation initialization); taking mutual information as an indicator, the grayscale correlation of the data is calculated; parameters are optimized through gradient descent until the mutual information is maximized. In brain tumor tests, the alignment error is

≤ 1.2 mm (meeting the requirements of radiotherapy target areas), the success rate in edema scenarios is ≥ 92%, which is 18% higher than that of traditional methods [16].

6.3 Alignment for Orthopedic Surgical Planning

Realizing the connection between "diagnosis and treatment" [17]: marking anatomical landmark points on X-ray films and calculating 2D coordinates; finding corresponding points in the 3D bone model to obtain 3D coordinates; solving the transformation matrix through the least squares method. In knee joint replacement experiments, the average alignment error is 0.8 mm (far below the 3 mm threshold), only 3-5 points need to be marked, the time consumption is < 30 seconds, and it is easy to promote in clinical practice [17].

7 Comparative Analysis and Discussion

Table 1. Performance of 2D and 3D Data Alignment in Different Fields.

Application Field	Manifestation of Core Advantages	Key Performance Indicators (Representative Cases)
Autonomous Driving	Fusing semantics and depth, enhancing perception integrity, and ensuring driving safety	Obstacle localization error ≤ 0.1 m (Waymo dataset) [2]
Robot Navigation	Achieving dynamic perception consistency and adapting to complex environments	Dynamic matching success rate: 98.7% (Multi-FEAT algorithm) [6]
Augmented Reality	Realizing virtual-real integration and improving immersion	Edge view error ≤ 3.5 mm (UV mapping) [9]
Indoor Scene Construction	Integrating appearance and structure, improving restoration degree, and shortening modeling cycle	Home modeling error ≤ 5 mm (SUN RGB-D) [13]
Medical Field	Correlating image details with 3D structures to assist clinical decision-making	Organ segmentation DSC = 0.94 (Learned Snakes model) [15]

As can be seen from Table 1, in high-precision fields (autonomous driving, medicine), the technology controls errors through multi-sensor fusion and mutual information optimization; in fields sensitive to experience/efficiency (AR, indoor construction), it balances precision and real-time performance through lightweight algorithms, reflecting the characteristic of "adapting to needs".

Each field has its shortcomings: autonomous driving is limited by computing power, leading to a more than 30% decrease in precision in dynamic scenarios; robot navigation is affected by occlusion and signals; AR is restricted by hardware and reflections; indoor construction has high costs and low success rates; medicine faces high annotation costs, and individual differences affect precision.

There are three common challenges: first, high computing costs make it difficult to popularize the technology on devices with limited computing power; second, weak dynamic robustness easily leads to "false alignment"; third, large modal differences make it hard to explore deep semantic consistency, and stability is affected by data quality.

8 Conclusion

This paper systematically sorts out the application status of 2D and 3D data alignment technology in five fields, clarifying its implementation paths and performance: in autonomous driving, high-precision perception is achieved through multi-dataset optimization; in robot navigation, stability is ensured through dynamic calibration; in AR, immersion is improved through mapping; in indoor construction, modeling is supported through large-scale data; in medicine, clinical assistance is provided through segmentation and alignment. These applications demonstrate the cross-domain adaptability and practical value of the technology.

To address challenges such as high computing costs, weak dynamic robustness, and large modal differences, future optimization can be carried out from three aspects: developing lightweight algorithms combined with hardware acceleration; introducing dynamic scene modeling to enhance anti-interference capabilities; exploring cross-modal semantic fusion to establish stable corresponding relationships.

As a basic support for intelligent technology, this technology will promote innovations in multiple fields. In the future, it can be extended to extreme scenarios such as deep-space exploration and underwater mapping, and cross-domain technology migration (e.g., applying medical mutual information methods to autonomous driving) should be strengthened to provide efficient solutions for intelligent upgrading.

References

1. Caesar, H., Bankiti, V., Lang, A.H., et al.: nuScenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11621–11631 (2020)
2. Sun, P., Kretzschmar, H., Dotiwalla, X., et al.: Scalability in perception for autonomous driving: Waymo open dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2446–2455 (2020)
3. Li, Y., Wang, J., Liu, M., et al.: HoloDrive: Holistic 2D-3D multi-modal street scene generation for autonomous driving. arXiv preprint arXiv:2305.12571 (2023)
4. Wu, Z., Ni, J., Wang, X., et al.: HoloDrive: Holistic 2D-3D multi-modal street scene generation for autonomous driving. arXiv preprint arXiv:2412.01407v2 (2024)
5. Tian, X., Jiang, T., Yun, L., et al.: Occ3D: A large-scale 3D occupancy prediction benchmark for autonomous driving. arXiv preprint arXiv:2304.14365v3 (2023)
6. Wang, L., Chen, C., Zhang, H., et al.: Multi-FEAT: Multi-feature fusion for dynamic calibration of camera and LiDAR in robot navigation. IEEE Robotics and Automation Letters 6(4), 7890–7897 (2021)

7. Li, C., Liu, Y., Wang, Z., et al.: Environment feature-based 2D-3D data alignment for indoor robot navigation. In: IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1023–1028. IEEE (2022)
8. Zhang, W., Zhao, J., Li, H., et al.: GPS-assisted 2D-3D data alignment for outdoor robot navigation in complex terrain. Journal of Field Robotics 40(2), 567–589 (2023)
9. Chen, S., Yang, J., Zhou, F., et al.: UV mapping-based real-time alignment of 2D facial video and 3D facial mesh for AR applications. IEEE Transactions on Visualization and Computer Graphics 28(7), 2689–2700 (2022)
10. Wang, H., Li, D., Zhang, S., et al.: Feature point matching for 2D-3D data alignment in industrial AR assembly guidance. In: IEEE International Conference on Industrial Technology (ICIT), pp. 1234–1239. IEEE (2023)
11. Liu, T., Chen, G., Wang, Y., et al.: VIO-assisted 2D-3D data alignment for mobile AR navigation. IEEE Access 10, 112345–112356 (2022)
12. Jiang, Y., Zhang, X., Li, M., et al.: SceneVerse: A large-scale dataset for 3D visual-language indoor scene understanding. arXiv preprint arXiv:2304.08758 (2023)
13. Song, S., Lichtenberg, S.P., Xiao, J.: SUN RGB-D: A RGB-D scene understanding benchmark suite. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 567–576 (2015)
14. Zhao, Y., Chen, J., Li, S., et al.: SCREAM: Robust 2D-3D data alignment for complex indoor scenes with low overlap. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 8901–8910. IEEE (2023)
15. Zhang, Q., Wang, M., Li, J., et al.: Learned snakes: Deep learning-enhanced contour evolution for 2D-3D alignment in abdominal organ segmentation. Medical Image Analysis 78, 102489 (2022)
16. Li, R., Chen, H., Wang, Z., et al.: Mutual information-based alignment of 2D brain MRI and 3D brain model for tumor localization. IEEE Transactions on Medical Imaging 40(12), 3567–3578 (2021)
17. Wang, J., Zhang, L., Li, C., et al.: Anatomical landmark-based alignment of 2D orthopedic X-rays and 3D bone models for surgical planning. IEEE Journal of Biomedical and Health Informatics 26(5), 2345–2356 (2022).