

Optimization and Comprehensive Application of ORB Algorithm and CNN Combined in Feature Extraction

Kefei Wu

Fuzhou University, Fuzhou, Fujian, 350108, China
3131629727@qq.com

Abstract: In recent years, an increasing number of studies have attempted to combine ORB with CNN to achieve both efficiency and high accuracy. This survey, based on existing literature, systematically organizes typical fusion methods, application cases, optimization strategies, and challenges of ORB+CNN, and discusses future research trends in this direction. All literature used in this survey comes from published research results, including representative studies of ORB and CNN in fields such as iris recognition, clothing retrieval, vehicle classification, and SDN traffic classification.

Keywords: ORB algorithm; Convolutional Neural Network (CNN); Feature extraction; Fusion methods; Feature-level fusion; Decision-level fusion; Lightweight model.

1. Introduction

Feature extraction is a fundamental component of computer vision tasks. Traditional local features such as SIFT, SURF, and ORB have long dominated early vision systems, with ORB being widely adopted for its low computational overhead and rotation invariance. However, with the advancement of deep learning, CNNs have demonstrated superior abstraction capabilities in large-scale image recognition tasks, making it difficult for traditional algorithms alone to meet the accuracy requirements in complex scenarios [1].

Therefore, combining the lightweight nature of ORB with the representational power of CNN has become a research hotspot in recent years. The integration of ORB and CNN can preserve the efficiency and robustness of ORB keypoints while leveraging CNN for more fine-grained feature extraction, representing a path that balances accuracy and efficiency [2, 3].

- 1) The goals of this survey are:
- 2) To summarize the main fusion methods of ORB and CNN;
- 3) To review their practical performance in different tasks;
- 4) To outline the problems in current research;
- 5) To explore potential future research directions.

2. Principles and Comparative Characteristics of ORB and CNN

2.1. Composition and Source of Advantages of ORB

ORB uses FAST for keypoint detection and BRIEF for descriptor construction, with additional mechanisms for orientation and rotation invariance to compensate for the weaknesses of traditional BRIEF [1, 4]. Its core advantages include:

- 1) Extremely fast keypoint detection speed: The FAST detection process only requires local brightness comparisons.

- 2) High compressibility of binary descriptors: BRIEFs bit-string features only require Hamming distance calculations for matching.
- 3) Applicability to large-scale image matching or real-time systems: e.g., SLAM, online search, etc.
- 4) Very low computational resource requirements: Suitable for mobile and embedded platforms.

However, literature also emphasizes its defects—limited generalization ability, making it inefficient in handling complex lighting, occlusion variations, or special scenarios like medical imaging [2, 3].

2.2. Representational Capabilities and Limitations of CNN

CNNs utilize convolutional kernels to automatically learn features, enabling progressive abstraction from edges, textures, and shapes to semantic concepts. This capability is unmatched by traditional algorithms and is the fundamental reason for CNNs long-standing dominance in large-scale image tasks [1,5]. Defense presentations also highlight the clear hierarchical structure of CNNs for classification: from convolutional and pooling layers to fully connected layers, forming a stable feature learning mechanism.

However, typical issues of CNNs cannot be ignored:

- 1) Both training and inference rely on GPU/TPU hardware.
- 2) Unfriendly to small-sample scenarios.
- 3) Large model parameter count, unsuitable for energy-sensitive devices.
- 4) Difficulty in interpreting the semantic meaning of their features.

3. Typical Fusion Methods of ORB and CNN

The fusion of ORB+CNN can be mainly categorized into three types.

3.1. Feature-Level Fusion: Concatenation of ORB and CNN Feature Vectors

A typical case includes the iris recognition system Contact LensIris, where ORB features are filtered by EH-WOA (Elephant Herding Whale Optimization Algorithm) and then fused with CNN deep features, jointly participating in classification [2, 3]. This method uses ORBs local features to enhance CNNs global descriptive power, improving recognition rates under interference conditions like wearing contact lenses. The local texture information provided by ORB overcomes some lighting and occlusion effects, while CNN emphasizes overall structure, making the system more robust.

3.2. Decision-Level Fusion: CNN for Macro Classification, ORB for Local Supplement

In clothing retrieval tasks, Martinez et al. used CNN for clothing category classification and then used the number of ORB matching points and color difference for geometric verification of candidate results, thereby improving matching accuracy under cross-pose and cross-angle conditions [4]. This decision-level fusion can effectively reduce CNN misjudgments under pose variations, especially suitable for retrieval tasks of items like clothing and footwear that have appearance differences but similar texture structures.

3.3. Feeding ORB and CNN Features in Parallel into a Classification Network

In vehicle classification tasks, researchers first used YOLOv4 to detect vehicles, then simultaneously input ORB, SIFT, and CNN features for fine-grained classification [5]. Results showed that the local texture information provided by ORB effectively compensated for CNNs neglect of geometric details.

3.4. Using ORB as a Front-End Replacement for CNN, Forming an ORB + NN Hybrid Model

The ORB+NN system proposed by Yzzogi and Benaboud for traffic classification in SDN environments performed notably [6]:

- 1) Accuracy: 97.14%, close to CNNs 99.13%.
- 2) Parameter count and training time were only about 1/3 to 1/10 of CNNs.

- 3) More suitable for real-time, resource-constrained applications.

In SDN traffic image classification, features extracted by ORB combined with a shallow neural network achieved accuracy close to CNN with minimal computational power, while significantly reducing training speed and model parameters [6]. This indicates that in resource-constrained scenarios, ORB is not merely a supplement to CNN but can serve as an efficient alternative.

4. Application Scenario Review

4.1. Iris Recognition: Enhancing Robustness Against Interference Conditions

The fusion system can significantly improve performance under [2, 3]:

- 1) Wearing contact lenses
- 2) Significant lighting differences
- 3) Slight occlusion

ORB has advantages in scenes with obvious local texture and edge changes, while CNN can supplement more abstract features, making the overall system more robust.

4.2. Clothing Retrieval: Advantages of Geometric Verification for Cross-View Matching

CNN classification may ignore subtle differences in geometric patterns, while ORBs geometric matching is sensitive to these [4]. Their combination can:

- 1) Reduce mismatches caused by pose changes.
- 2) Provide more accurate similarity scores.

4.3. Aerial Vehicle Classification: Enhancing Small-Scale Target Recognition Ability

In aerial images, vehicles are small and backgrounds are complex. CNN may blur small targets, while ORBs local feature points can provide more structural information, improving classification performance [5].

4.4. Network Traffic Image Recognition: Lightweight Alternative for Resource-Constrained Scenarios

- 3) Online traffic monitoring systems and other energy-sensitive applications.

Table 1. Summary of ORB+CNN Fusion Applications

Application Domain	ORBs Role	CNNs Role	Summary of Results
<i>Iris Recognition</i>	Local features, fault tolerance enhancement	Deep representation	Significantly improves recognition accuracy under interference conditions
<i>Clothing Retrieval</i>	Geometric verification, fine-grained matching	Category classification	Effectively addresses recognition difficulties due to cross-pose variations
<i>Vehicle Classification (Aerial)</i>	Keypoint texture supplement	Global classification	Improves classification accuracy in small-target scenarios
<i>SDN Traffic Image Classification</i>	Efficient alternative to convolutional features	Shallow NN classifier	Significantly reduces inference time and computational costs

The ORB + NN combination shows that traditional features combined with shallow networks can ensure accuracy while drastically reducing training and inference costs, making it very suitable for [6] :(Table.1)

- 1) IoT networks
- 2) Edge devices

5. Current Challenges

5.1. Difficulty in Unified Representation of Multimodal Features

Binary ORB descriptors and CNN floating-point features have representational differences. Direct concatenation can lead to [2, 4]:

- 1) Inconsistent feature scales.
- 2) Unstable gradients.
- 3) Models requiring more complex feature fusion layers.

5.2. Significant Room for Improving ORB Robustness

In scenarios with strong light, motion blur, or non-rigid deformations—common in real-world applications—ORBs performance declines noticeably [2, 4-6].

5.3. Lack of a Unified Fusion Architecture

Current fusion methods are mostly task-specific, lacking a general framework. For example [2, 4-6]:

- 1) When to use ORB as the front-end?
- 2) What is the best way for feature concatenation?
- 3) How to dynamically select fusion strategies based on computing power?

5.4. Insufficient Standardization of Engineering Deployment Strategies

There is a lack of mature development norms for deployment on mobile devices, robotic systems, and embedded devices, limiting the industrial adoption of these methods [4-6].

6. Future Research Directions

First, learnable ORB is a very promising development path. In existing ORB, the point-pair set for BRIEF is predefined. However, different datasets and scenarios have varying requirements for feature points [2,3,6]. Future work could use reinforcement learning or online learning to automatically select optimal point pairs for different tasks, endowing ORB with adaptive capabilities beyond fixed rules.

Second, lightweight fusion for low-power and real-time systems will become a major trend. In many mobile and UAV environments, the computational consumption of deep models remains a serious issue [5, 6]. How to use ORB at the front-end to reduce input redundancy for CNN, or replace some convolutional layers with ORB, is an important direction for improving energy efficiency.

Third, end-to-end differentiable fusion of ORB and CNN remains worthy of exploration. Currently, key steps in ORB such as binarization and threshold comparison are non-

differentiable, making it difficult to directly integrate into deep networks [2,4-6]. If future research can propose differentiable alternative modules or construct new descriptor structures that maintain the ORB style while participating in gradient training, it could balance model performance and interpretability.

7. Conclusion

The combination of ORB and CNN represents the trend of integrating traditional features with deep learning, aiming to optimize both speed and accuracy. Through the classification and review of existing literature, it is evident that fused models significantly enhance system performance in iris recognition, clothing retrieval, vehicle classification, and network traffic recognition, demonstrating high research value. Despite remaining challenges such as inconsistent feature representation and insufficient robustness in extreme scenarios, future directions including learnable point-pair sets, end-to-end fusion architectures, and lightweight network design will further promote the widespread application of ORB+CNN technology in the vision field.

References

- [1] Chen, Y., Cheng, W., & Wang, Y. (2020). T-center: A novel feature extraction approach towards large-scale Iris recognition. *IEEE Access*, 8, 32365–32375. <https://doi.org/10.1109/ACCESS.2020.2973601>
- [2] Babu, G., & Khayum, P. A. (2022). Elephant herding with whale optimization enabled ORB features and CNN for Iris recognition. *Multimedia Tools and Applications*, 81, 5761–5794. <https://doi.org/10.1007/s11042-021-11821-z>
- [3] Kaur, B., & Saini, S. S. (2024). ContactLensIris: A robust iris recognition system based on integrated deep learning approach. *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*, 1–5. <https://doi.org/10.1109/ICEECT61758.2024.10739228>
- [4] Martinez, L., Rivera, M. M., Zezatti, A. O., Ponce, J., & Guzman, E. (2025). Web-based clothing searcher utilizing convolutional neural networks and dissimilarity rules for color classification in the HSL color space and ORB algorithm for garment characterization. *International Journal of Combinatorial Optimization Problems and Informatics*, 16(3), 118–132.
- [5] Yusuf, M. O., Hanzla, M., & Jalal, A. (2024). Vehicle detection and classification via YOLOv4 and CNN over aerial images. *2024 International Conference on Engineering & Computing Technologies (ICECT)*, 1–6. <https://doi.org/10.1109/ICECT61618.2024.10581252>
- [6] Yzzogh, H., & Benaboud, H. (2025). Evaluating ORB and SIFT with neural network as alternatives to CNN for traffic classification in SDN environments. *IEEE Access*, 13, 47986–48001. <https://doi.org/10.1109/ACCESS.2025.3553449>