Received: 5 April 2026 | Revised: 30 June 2026 | Accepted: 27 July 2026 | Published online: 24 August 2026

REVIEW

Medinformatics
2026, Vol. 00(00) 1–13
DOI: 10.47852/bonviewMEDIN62029908

BON VIEW PUBLISHING

Segmentation and Multimodal Fusion in Ocular Biometric Recognition: A Review

Yansuo Yu[1,†,*], Yongbin Qi[1,†], Shilin Zhao[1], Huanzhang Qi[1], Wendao Li[1], Haoqi Zhang[1], Da Teng[1] and Qiang Liu[1,*]

[1] *Academy of Artificial Intelligence, Beijing Institute of Petrochemical Technology, China*

Abstract: Accurate segmentation of ocular regions is a fundamental prerequisite for robust ocular biometric recognition, as it affects feature extraction, multimodal representation, and subsequent matching under unconstrained conditions. This review focuses on ocular region segmentation as an essential component of biometric recognition rather than an isolated semantic segmentation task. We survey major segmentation paradigms, including traditional image-processing methods, deep learning architectures, and emerging foundation-model-based approaches, and analyze their applications in delineating the iris, sclera, pupil, and periocular structures under challenges such as occlusion, illumination variation, specular reflection, and motion blur. We further categorize multimodal ocular fusion strategies into three levels—data-, feature-, and decision-level fusion—and compare their information preservation, alignment requirements, computational complexity, and deployment suitability. To facilitate reproducible evaluation, we review representative public datasets and benchmark protocols, summarizing their acquisition characteristics, annotation properties, and evaluation metrics. Finally, we discuss practical considerations for method selection and highlight persistent challenges, including image-quality degradation, cross-spectral and cross-sensor domain shifts, limited dataset diversity, and cross-modal misalignment. Future research directions are outlined toward standardized, quality-aware, efficient, and deployable ocular biometric recognition systems.

Keywords: ocular biometrics, multimodal fusion, feature segmentation, deep-learning-based segmentation, ocular biometric datasets

1. Introduction

With the rapid digitalization of society, conventional identity verification methods, including passwords, identity cards, and physical credentials, are increasingly inadequate because they are vulnerable to theft, forgery, and loss [1]. Biometrics offers a more secure alternative by enabling automated recognition from distinctive physiological or behavioral traits, thereby improving both security and usability in modern information systems. A central advantage of biometric authentication is its ability to reconcile security and convenience: unlike knowledge-based credentials that impose cognitive burden or token-based mechanisms that can be lost, biometric traits are intrinsically linked to the user and enable frictionless authentication [2]. These properties have supported widespread deployment in access control and mobile payment systems, where fingerprint recognition combined with liveness detection can mitigate fraud risks [3].

Among biometric modalities, ocular biometrics has attracted sustained attention because ocular traits are highly distinctive and relatively stable over time. Iris texture remains largely stable across the lifespan and is widely used for identity verification [4]; it has also been investigated in broader pattern-analysis settings, although biometric recognition remains the primary focus of this review [5]. Complementary ocular regions provide additional biometric evidence. The sclera contains vascular patterns that develop early and remain stable in adulthood, supporting recognition particularly when integrated with iris information [6]. The pupil exhibits dynamic responses to illumination and physiological state, providing auxiliary cues that may complement identity verification under specific acquisition settings [7]. The periocular region, encompassing eyelids, eyebrows, and surrounding skin, is comparatively less constrained during acquisition and can offer robust cues under both controlled and unconstrained conditions [8].

Despite their utility, individual ocular traits exhibit limitations when used in isolation. Iris recognition can be degraded by occlusions from eyelids or eyelashes [9]; pupil-based cues are sensitive to illumination variation [10]; and scleral vasculature may be difficult to capture reliably in unconstrained settings. Periocular appearance, while often available, can be influenced by non-genetic factors, including aging and environmental conditions [8]. Collectively, these factors can reduce discriminability and impair generalization in real-world deployments.

To improve reliability under challenging conditions, recent studies increasingly emphasize multimodal ocular biometrics that integrate complementary cues from multiple eye regions, including the iris, periocular region, sclera, and pupillary dynamics. Prior work reports consistent gains in recognition robustness and presentation-attack resistance: Rasheed et al. reviewed deep-learning-based approaches to iris segmentation and recognition

***Corresponding author:** Yansuo Yu, Academy of Artificial Intelligence, Beijing Institute of Petrochemical Technology, China. Email: yuyansuo@bipt.edu.cn. Qiang Liu, Academy of Artificial Intelligence, Beijing Institute of Petrochemical Technology, China. Email: liuq@bipt.edu.cn
†Yansuo Yu and Yongbin Qi contributed equally to this work

and indicated that these methods can improve the accuracy and reliability of iris-based biometric authentication by enhancing iris localization, feature extraction, and classification, although their performance remains affected by limited training data and noisy or low-quality ocular images [11], to improve reliability under challenging conditions, recent studies have increasingly emphasized multimodal ocular biometrics that integrate complementary cues from multiple eye regions, including the iris, periocular region, sclera, and pupillary dynamics. Such fusion strategies can compensate for modality-specific degradation and improve recognition robustness. In parallel, presentation-attack detection has become essential for strengthening the security of biometric systems. Shaheed et al. systematically reviewed deep-learning-based presentation-attack detection methods for multiple biometric modalities, including iris recognition, and highlighted their effectiveness in detecting spoofing and presentation attacks [12]. More broadly, fusion can compensate for modality-specific degradation by leveraging complementary evidence from other regions [13], and adaptable integration strategies can improve stability across diverse operating conditions [14].

Robust feature extraction remains a prerequisite for both unimodal and multimodal ocular systems. Semantic segmentation localizes key structures, including the iris, sclera, and pupil, while suppressing confounding factors, including eyelash occlusion and specular reflections; this, in turn, improves downstream representation learning and recognition performance [15]. Under non-ideal or pathological conditions, segmentation further supports adaptive selection of usable regions, mitigating the brittleness of traditional image-processing methods [15].

This review focuses on ocular region segmentation and multimodal fusion in the context of biometric recognition. Unlike general-purpose eye-image understanding or isolated semantic segmentation studies, the scope of this paper is limited to segmentation methods that support the extraction, normalization, and integration of biometric cues from the iris, sclera, pupil, and periocular region. In this setting, segmentation is not treated as an independent visual task but as a prerequisite for reliable feature representation, multimodal fusion, and recognition under unconstrained acquisition conditions. First, we provide a structured taxonomy of ocular region segmentation methods, covering traditional image-processing pipelines, modern deep learning architectures, and emerging promptable foundation-model-based approaches, and discuss their applicability to biometric-oriented ocular structures under occlusion, illumination variability, reflections, and motion blur. Second, we organize multimodal fusion strategies into three canonical levels—data-, feature-, and decision-level fusion—and summarize their trade-offs in terms of information preservation, alignment requirements, computational complexity, and deployment suitability. Third, we consolidate representative public datasets and benchmark practices, clarifying acquisition protocols, modality characteristics, annotation properties, and commonly adopted evaluation metrics. Finally, we synthesize persistent bottlenecks, including image-quality instability, cross-spectral and cross-sensor domain shifts, limited dataset diversity, and cross-modal misalignment, and outline future directions toward standardized, quality-aware, efficient, and deployable ocular biometric recognition pipelines.

2. Advances in Multimodal Ocular Feature Segmentation Techniques

Ocular biometric systems exploit multiple anatomical traits, including the iris, sclera, pupil, and periocular region. Accurate segmentation of these structures is a prerequisite for reliable feature extraction, multimodal fusion, and downstream recognition. Existing segmentation methods can be broadly divided into traditional image-processing approaches and deep-learning-based approaches. Traditional methods rely on handcrafted visual cues and are usually efficient and interpretable, making them useful under controlled acquisition conditions or resource-limited deployment scenarios. In contrast, deep learning methods learn hierarchical representations from data and generally provide stronger robustness under illumination variation, occlusion, reflections, and motion blur. However, this transition does not imply that traditional methods have become obsolete. The choice of segmentation paradigm should depend on image quality, annotation availability, computational budget, reproducibility, and deployment requirements. An overview of representative methodological streams is provided in Figures 1 and 2.

2.1. Traditional segmentation methodologies

In the early development of ocular biometrics, segmentation relied predominantly on handcrafted image-processing operations. Common methodological categories include thresholding [16], region-based partitioning [17], edge detection [18], clustering [19], and morphological processing [20]. Although these approaches can achieve satisfactory performance under constrained conditions, they often degrade on complex ocular images due to poor illumination, occlusion, specular reflections, and subject-specific appearance variation.

These limitations largely stem from dependence on low-level cues and sensitivity to hyperparameter settings. Thresholding is prone to failure under low contrast and noise [16], whereas region-based methods may exhibit over-segmentation when texture or illumination varies [17]. Gradient-based detectors can localize boundaries efficiently [21] but remain susceptible to noise and blurred edges [18]. Clustering-based strategies (including K-means [22]) can become unstable under non-uniform intensity distributions, a frequent characteristic of periocular imagery [19]. Morphological processing can isolate structures by exploiting shape priors, yet it typically requires careful parameter tuning and may incur substantial computational cost when applied iteratively [20]. As a result, traditional image-processing methods alone are often inadequate for modern multimodal ocular biometrics, which require robust segmentation across diverse real-world conditions.

2.2. Deep-learning-driven segmentation models

Deep learning has become the dominant paradigm for ocular feature segmentation, replacing handcrafted rules with hierarchical representation learning. Convolutional neural networks (CNNs) learn multilevel features that capture both local texture information and higher-order structural context, thereby improving robustness to illumination variation, noise, occlusion, and pose changes. This capability is particularly important for delineating fine-scale structures, including retinal vasculature, and for resolving ambiguous boundaries in periocular regions. In the following, we summarize three representative model families that are widely used in ocular segmentation—U-Net, DeepLab, and promptable foundation models, including the Segment Anything Model (SAM)—whose roles align conceptually with the unified framework in Figure 2.

2.2.1. U-Net architecture

U-Net is a canonical encoder–decoder architecture that has been widely adopted for medical image segmentation [23]. Its effectiveness stems from multi-scale feature extraction through

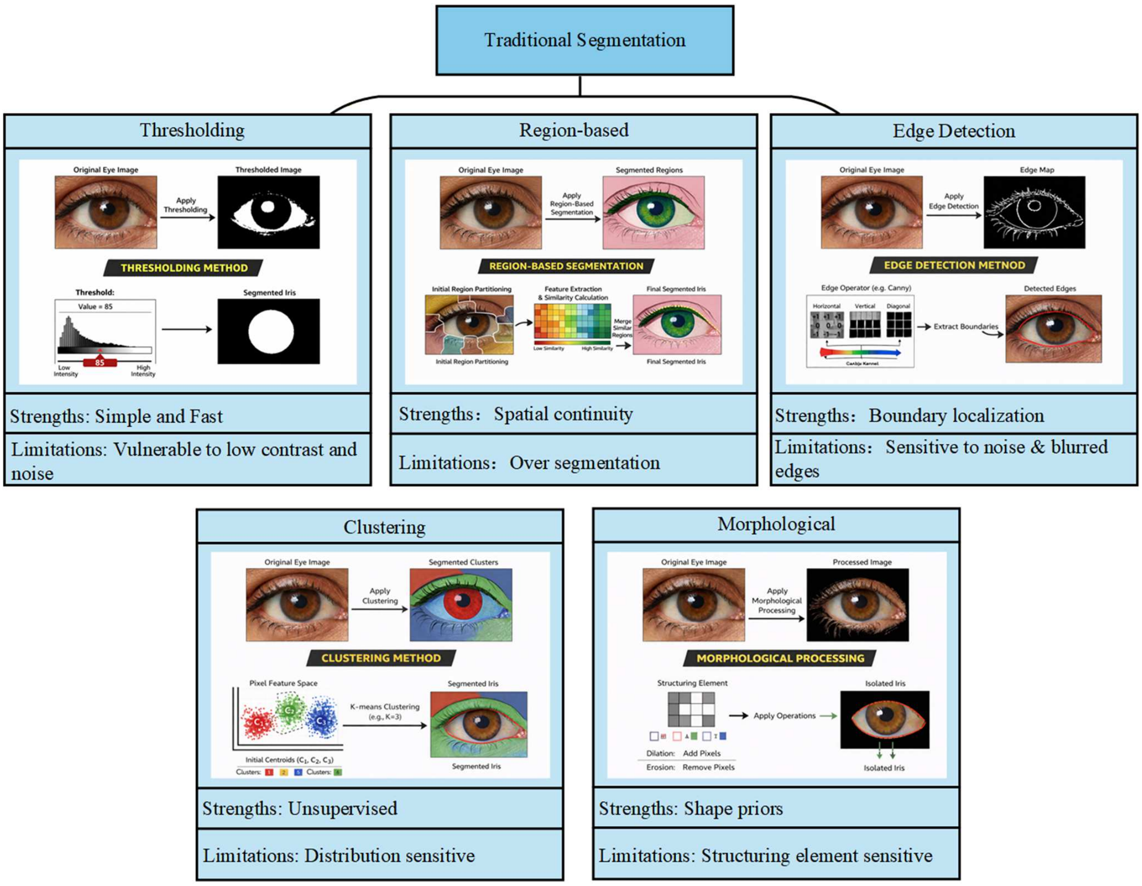

Figure 1. Traditional image-processing-based ocular region segmentation methods

downsampling and accurate localization through upsampling, while skip connections integrate encoder features with decoder representations to preserve spatial detail [24]. With appropriate data augmentation strategies, including elastic deformations, U-Net-style models can perform well even when training data are limited [25].

A substantial body of work has extended U-Net in several directions. Attention and gating mechanisms emphasize task-relevant structures [26, 27]; residual and dense connectivity facilitate optimization and feature reuse [28–31]; and hybrid CNN–Transformer designs model longer-range dependencies [32]. Collectively, these variants indicate that strengthened feature fusion and contextual modeling are critical for segmenting ocular regions characterized by subtle textures and ambiguous boundaries. The U-Net architecture is illustrated in Figure 2.

2.2.2. DeepLab framework

DeepLab advances semantic segmentation by leveraging dilated (atrous) convolutions and multi-scale context aggregation [33]. By expanding the receptive field without excessive downsampling, dilated convolutions preserve spatial resolution while capturing broader contextual information. DeepLabv3 introduced Atrous Spatial Pyramid Pooling to integrate features across multiple scales, which benefits ocular structures exhibiting substantial variability in size and shape [34]. DeepLabv3+ subsequently incorporated an encoder–decoder design to recover fine boundaries more effectively by integrating low-level features [35].

In ocular applications, subsequent studies have extended this model family to improve robustness and discriminability. Qi et al. proposed an ensemble framework built on DeepLabv3+ for sclera and vasculature segmentation [36]. Attention-augmented variants further strengthen region-focused representation learning under cluttered backgrounds [37]. Overall, the DeepLab lineage highlights the importance of multi-scale context modeling and boundary refinement; the DeepLabv3 architecture is illustrated in Figure 2.

2.2.3. Foundation models and promptable segmentation

The SAM is a foundation vision model that supports promptable segmentation by coupling a Vision Transformer (ViT)-based image encoder with a prompt encoder and a mask decoder [38]. Relative to task-specific networks, SAM exhibits strong transferability through flexible prompting, which has motivated its use in ocular segmentation, including iris segmentation [39], as well as broader adaptations in medical imaging [40]. To better accommodate ocular-domain characteristics and challenging acquisition conditions, including occlusion and complex illumination, SAM is frequently augmented with task-specific adaptation modules and multi-scale feature strategies [41]. Within multimodal ocular biometrics, promptable foundation models offer an alternative route to reduce manual design choices while preserving strong generalization potential (see Figure 2).

Foundation models are increasingly relevant to ocular biometrics because they provide transferable visual representations and promptable segmentation capabilities that may reduce the dependence on fully task-specific model design. In ocular biometric applications, their potential roles are not limited to direct mask prediction. They can also support interactive annotation, semi-automatic ground-truth generation, region-of-interest localization, cross-dataset adaptation, and preprocessing for

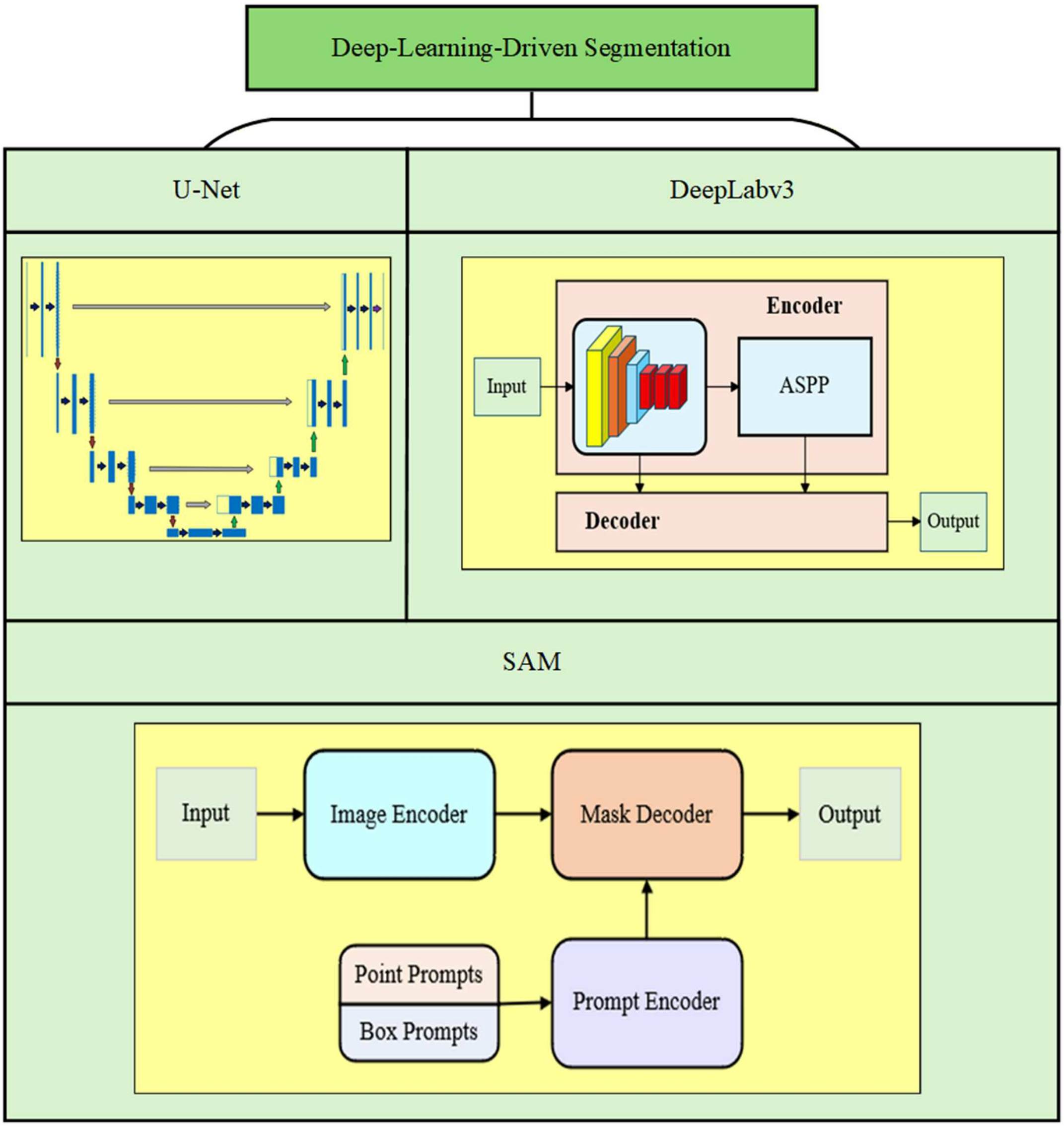

Figure 2. Deep-learning-driven ocular region segmentation methods

downstream iris, sclera, pupil, or periocular recognition. However, current foundation models are usually trained on broad natural-image or general medical-image data, and their direct application to ocular biometrics may still suffer from domain mismatch, insufficient boundary precision, limited sensitivity to thin scleral vessels, and high computational cost. Therefore, ocular-domain adaptation, prompt optimization, lightweight fine-tuning, and recognition-oriented evaluation remain necessary before foundation models can be reliably deployed in practical ocular biometric systems.

2.3. Comparative analysis and practical considerations

Although deep learning has become the dominant paradigm for ocular region segmentation, its advantages should be interpreted together with the practical constraints of ocular biometric applications. Traditional image-processing methods are simple, interpretable, and computationally efficient, and therefore remain useful in controlled imaging environments where ocular boundaries are clear and acquisition conditions are stable. However, their reliance on low-level cues, such as intensity, gradients, and morphology, makes them sensitive to illumination variation, specular reflections, occlusion, and parameter settings.

Deep learning methods, including CNN-based encoder–decoder networks, multi-scale segmentation frameworks, Transformer-based models, and promptable foundation models, provide stronger representation learning and are more robust to complex imaging conditions. Nevertheless, most of these methods require high-quality pixel-level annotations. In ocular biometrics, manual annotation is particularly costly because structures such as the iris boundary, pupil contour, eyelids, sclera, and scleral vasculature are often small, low-contrast, and affected by reflections or occlusions. In addition, complex models may introduce higher training and inference costs, increased code reproduction difficulty, and reduced suitability for real-time mobile or edge deployment.

Therefore, traditional and deep learning methods should not be viewed as mutually exclusive alternatives. Traditional methods are more suitable for constrained, low-cost, and interpretable segmentation pipelines, whereas deep learning methods are preferable for unconstrained scenarios with sufficient annotated data and computational resources. Foundation-model-based approaches such as SAM provide promising transferability, but their direct application to ocular biometrics still requires domain adaptation, prompt design, and efficiency optimization.

3. Multimodal Fusion Strategies for Ocular Biometrics

Multimodal fusion has become increasingly important for ocular biometric recognition because it integrates heterogeneous yet complementary cues, including iris texture, scleral vasculature, and pupil morphology. By jointly exploiting structural, textural, and boundary information, fusion improves both discriminability and robustness, particularly under unconstrained acquisition conditions characterized by illumination variation, motion blur, and occlusion. In ocular biometrics, fusion is commonly implemented at three levels—data-, feature-, and decision-level fusion—each involving a distinct trade-off among information preservation, computational cost, and implementation complexity. A unified overview of these paradigms is provided in Figure 3.

3.1. Data-level fusion

Data-level fusion combines raw observations from multiple modalities into a unified input for downstream processing [42]. Because fusion is performed prior to feature extraction, it can preserve fine-grained spatial or spectral detail and exploit cross-modal complementarity. Common implementations include pixel- or weight-based averaging and wavelet-based fusion [43]. In practice, data-level fusion is sensitive to imperfect spatial or temporal alignment; misregistration can introduce artifacts that propagate to subsequent stages. It may also increase computational and storage requirements due to the higher dimensionality of the fused input.

To mitigate cross-spectral and cross-resolution mismatch in iris recognition, Mostofa et al. [44] proposed an adversarial generative framework that maps visible-spectrum (VIS) images to the near-infrared (NIR) domain and applies a super-resolution module to restore fine texture detail. By reducing modality discrepancies at the data level, this framework improves recognition reliability. Wei et al. [45] proposed a Gabor Trident Network to estimate device-specific spectral components and generate spectral-invariant iris features, while adversarial learning and distribution-alignment objectives were employed to reduce the discrepancy between VIS and NIR iris images.

Therefore, data-level fusion is more suitable for controlled or semi-controlled ocular acquisition scenarios where multimodal samples are well registered, but it is less robust when cross-modal misalignment, motion blur, or sensor differences are severe.

3.2. Feature-level fusion

Feature-level fusion integrates descriptors extracted from different modalities into a unified representation [46]. Relative to data-level fusion, it is typically more efficient because it operates on compact feature vectors rather than raw signals and reduces reliance on strict pixel-wise alignment. Typical implementations include feature concatenation, weighted summation, and feature selection or dimensionality reduction to mitigate redundancy [47]. Nevertheless, effective feature-level fusion often requires careful normalization and compatibility design, as heterogeneous descriptors may differ in scale, noise characteristics, and statistical distributions.

Luo et al. [48] proposed an end-to-end deep feature fusion network that integrates spatial, channel, and co-attention mechanisms to adaptively combine iris and periocular representations, outperforming the corresponding unimodal biometric approaches on two publicly available datasets. Vyas [49] proposed a hybrid strategy that combines handcrafted descriptors, including local binary patterns and histograms of oriented gradients, with deep features, yielding more stable performance than single-source representations across multiple datasets.

In practical ocular biometric systems, feature-level fusion often provides a favorable compromise between discriminative

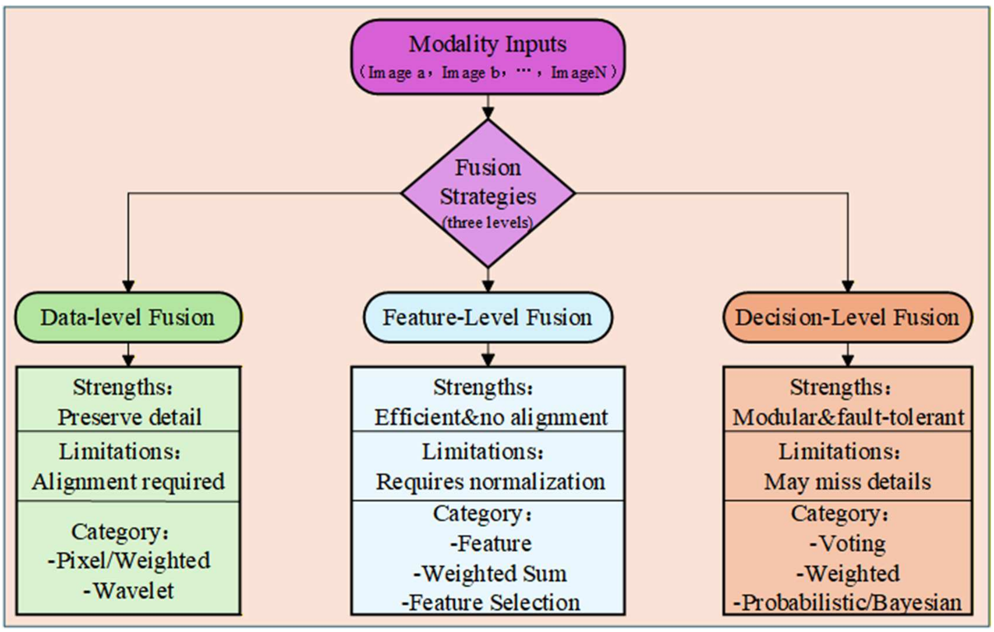

Figure 3. Overview of multimodal fusion strategies in ocular biometrics: data-level, feature-level, and decision-level fusion

complementarity and computational efficiency, especially when iris, sclera, pupil, and periocular features are extracted by different but compatible models.

3.3. Decision-level fusion

Decision-level fusion aggregates the outputs of modality-specific matchers into a final decision [50]. Typical rules include voting, weighted averaging, and probabilistic or Bayesian schemes [51]. This paradigm is modular and scalable because each modality can be processed independently, facilitating system extension and improving fault tolerance when one modality degrades. However, because fusion is performed after feature extraction or classification, decision-level fusion may not fully exploit fine-grained complementary information that is available at earlier stages.

Alonso-Fernandez et al. [52] reported that decision/score-level fusion can reduce the Equal Error Rate (EER) for cross-sensor periocular recognition by combining handcrafted and deep representations, particularly under practical covariates such as mask wearing. Overall, decision-level fusion is often preferred when heterogeneous sensors are difficult to align at the data level or when independent modality pipelines must be maintained for deployment constraints.

Thus, decision-level fusion is particularly suitable for modular systems, heterogeneous sensors, or situations in which some modalities may be unavailable, although careful score normalization and calibration are required.

3.4. Comparative analysis and selection guidance

As shown in Table 1, the three fusion strategies differ substantially in their assumptions and deployment suitability. Data-level fusion is advantageous when multimodal ocular images are accurately aligned, because it can preserve rich spatial and spectral information before feature extraction. However, its sensitivity to registration errors and increased computational burden limit its applicability in unconstrained environments. Feature-level fusion offers a more balanced solution for ocular biometric recognition. By integrating compact representations from the iris, sclera, pupil, and periocular region, it can exploit complementary information without requiring strict pixel-level alignment, although normalization and compatibility among heterogeneous descriptors remain necessary. Decision-level fusion is the most modular strategy and is suitable for systems involving heterogeneous sensors, independent matchers, or missing modalities. Nevertheless, because it operates on final scores or decisions, it may underuse fine-grained cross-modal relationships. Therefore, the choice of fusion strategy should depend on modality alignment, computational resources, feature compatibility, sensor heterogeneity, and deployment requirements.

4. Ocular Biometric Datasets: Construction and Applications

Ocular biometric datasets are commonly categorized by sensing modality and acquisition protocol. With respect to illumination, datasets are typically divided into VIS and NIR collections; with respect to capture conditions, they are often characterized as controlled or unconstrained. Datasets containing paired VIS and NIR samples are generally described as cross-spectral, whereas those acquired using different devices are referred to as cross-sensor. Representative datasets and their key attributes are summarized in Table 2; this section discusses how major dataset families support benchmarking for recognition, segmentation, and robustness evaluation.

4.1. Near-infrared (NIR) databases and applications

NIR imaging mitigates the effects of iris pigmentation and is therefore widely adopted in practical iris recognition. The CASIA Iris Image Database series [53] remains a canonical benchmark, having expanded to include more diverse subjects and increasingly

Table 1. Comparative analysis of multimodal fusion strategies for ocular biometric recognition

Fusion strategy	Fusion stage	Information preservation	Alignment requirement	Computational cost	Implementation complexity
Data level	Before feature extraction; raw images or signals are combined as a unified input	High; preserves pixel-level spatial, spectral, and texture information	High; requires accurate spatial or temporal registration between modalities	High; fused inputs may increase storage and processing burden	High; sensitive to preprocessing, calibration, and registration errors
Feature level	After feature extraction; descriptors from different modalities are integrated into a joint representation	Moderate to high; preserves complementary semantic and structural information while reducing raw-data redundancy	Moderate; does not require strict pixel-level alignment but needs feature compatibility	Moderate; operates on compact feature vectors rather than raw images	Moderate; requires feature normalization, weighting, selection, or dimensionality reduction
Decision level	After classification or matching; scores, probabilities, or decisions from different matchers are aggregated	Low to moderate; mainly uses high-level outputs and may lose fine-grained cross-modal correlations	Low; each modality can be processed independently	Low to moderate; depends on the number of matchers and fusion rules	Low to moderate; modular and easy to extend but requires score calibration

Table 2. Summary of representative ocular and Iris datasets with their main characteristics and applications

Dataset (year)	Modality	Setting	Scale (Imgs/Subj)	Resolution	Main features and applications
CASIA-Iris V1 (2003)	NIR (Iris)	Ctrl.	756/108	320 × 280	Early NIR baseline for iris benchmarking
CASIA-Iris V3 (2010)	NIR (Iris)	Ctrl.	22,034/700+	640 × 480	Interval/Lamp/Twins subsets for variability analysis
CASIA-Iris V4 (2010)	NIR (Iris)	Ctrl./Semi-ctrl.	54,601/2,800	640 × 480	Multiple subsets for robust segmentation and recognition
ND-IRIS-0405 (2009)	NIR (Iris)	Ctrl.	64,980/356	640 × 480	Large-scale NIR benchmark used in ICE 2005/2006
IIT Delhi Iris V1 (2010)	NIR (Iris)	Ctrl.	1,120/224	320 × 240	Lower-quality NIR sensing for robustness assessment
ND Contact Lens (2010)	NIR (Iris)	Ctrl.	21,700/211	640 × 480	Contact-lens interference evaluation dataset
ND Template Aging (2012)	NIR (Iris)	Ctrl. / Long.	22,156/322	640 × 480	Longitudinal template aging and stability analysis
UBIRIS.v1 (2005)	VIS (Iris)	Semi-ctrl./Unctrl.	1,877/241	800 × 600	Noisy VIS iris images with reflections and illumination changes
UBIRIS.v2 (2010)	VIS (Iris)	Unctrl.	11,102/261	400 × 300	Long-range motion blur and occlusion for NICE I/II
UTIRIS (2010)	VIS + NIR (Paired)	Ctrl.	1,540/79	2048 × 1360/1000 × 776	Early paired VIS–NIR protocol for cross-spectral studies
PolyU Cross-Spectral (2017)	VIS + NIR (Paired)	Ctrl.	12,540/209	640 × 480	Pixel-aligned VIS–NIR pairs for cross-spectral benchmarking
Cross-Eyed (2016)	VIS + NIR (Iris + Perioc.)	Semi-ctrl.	3,840/120	900 × 800/400 × 300	Cross-spectral iris/periocular data for fusion and adaptation
BiosecurID (2009)	Multimodal (Iris+)	Ctrl.	12,800/400	640 × 480	Multi-trait corpus for fusion benchmarking
BioSecure DS2 (2009)	Multimodal (Iris subset)	Ctrl.	5,336/667	640 × 480	Multimodal evaluations under office/web-style conditions
SBVPI (2020)	VIS (Sclera/Perioc.)	Ctrl.	1,858/55	3000 × 1700	High-resolution scleral vessel imagery for sclera recognition
OpenEDS (2019)	NIR (Eye/Perioc.)	Ctrl.	356,649/152	640 × 400	Semantic labels for iris/pupil/sclera segmentation benchmarks

Note: VIS, visible spectrum; NIR, near-infrared; Ctrl., controlled; Semi-ctrl., semi-controlled; Unctrl., uncontrolled; Long., longitudinal; Perioc., periocular; Paired, paired VIS–NIR samples. ICE, Iris Challenge Evaluation; NICE, Noisy Iris Challenge Evaluation.

challenging subsets, and is frequently used to evaluate segmentation accuracy, feature representation, and matching robustness. The ND-IRIS-0405 dataset [54] is another widely used NIR resource and has been employed in large-scale evaluations, including the Iris Challenge Evaluation (ICE) 2005 and 2006 campaigns. In addition, targeted Notre Dame datasets examine contact-lens interference [55] and longitudinal template aging [56]. Other controlled NIR datasets, including IIT Delhi Iris V1 [57], further support robustness analysis under reduced sensing quality and lens-related appearance changes.

4.2. Visible-spectrum (VIS) image databases

VIS iris images acquired in noncooperative scenarios typically exhibit larger appearance variation than NIR data, motivating unconstrained benchmarks. The University of Beira Interior Iris (UBIRIS) series [58, 59] progressed from mildly relaxed acquisition to long-range, motion-affected settings with frequent blur, glare, and occlusion and supported the Noisy Iris Challenge Evaluation (NICE) competitions emphasizing segmentation and noise handling. Competition-driven mobile resources further expanded VIS evaluation: Mobile Iris Challenge Evaluation (MICHE) I/II led to the MICHE database [60] for smartphone-based iris/periocular benchmarking, while the Visible Light Mobile Ocular Biometric (VISOB) challenge released the VISOB dataset [61], a larger-scale periocular corpus enabling studies of robustness to device variability, pose changes, and visible-light noise.

4.3. Cross-spectral databases

Cross-spectral datasets provide paired VIS and NIR samples to benchmark wavelength-induced domain shift. University of Tehran Iris (UTIRIS) [62] established early dual-spectral protocols, the PolyU Cross-Spectral Iris Database [63] expanded scale with improved alignment, and CROSS-EYED [64] further included both iris and periocular regions to support domain normalization and fusion studies.

4.4. Databases for sclera, pupil, and related modalities

Beyond iris texture, several dedicated datasets support research on alternative ocular traits and eye-region analysis. Sclera Blood Vessels, Periocular and Iris (SBVPI) [65] provides high-resolution VIS imagery of scleral vasculature to facilitate sclera-based recognition studies, whereas OpenEDS [66] offers large-scale NIR periocular data with semantic annotations, including the iris, pupil, and sclera, to support segmentation and eye-tracking benchmarking.

Public datasets and benchmark challenges provide a foundation for controlled evaluation (primarily NIR), robustness assessment under unconstrained capture (VIS and mobile settings), and analyses of domain shift in cross-spectral and cross-sensor scenarios. Multimodal resources further enable fusion benchmarking and complementary-evidence modeling; representative examples include BiosecurID and the BioSecure Multimodal Database [67, 68]. Together with Table 1, these resources facilitate reproducible cross-study comparisons.

4.5. Recent ocular region datasets from 2020 to 2026

As summarized in Table 3, recent ocular datasets from 2020 to 2026 indicate that ocular biometric research is moving beyond controlled iris acquisition toward more diverse and application-driven scenarios, including virtual reality/augmented reality (VR/AR) head-mounted displays (HMDs), mobile devices, outdoor ambient illumination, periocular biometrics, gaze-related eye-region analysis, and demographic diversity. OpenEDS2020 [69] and TEyeD [70] are particularly relevant to eye-region segmentation because they provide semantic annotations or landmark information for structures such as the pupil, iris, and eyelids. VRBiom [71] and mobile periocular datasets [72] further extend ocular biometrics to head-mounted and mobile acquisition environments, where non-frontal views, low resolution, glasses, gaze variation, and device heterogeneity become important factors. CASIA-Iris-Africa [73] contributes to demographic diversity in iris biometrics, while AmbientEye [74] highlights the emerging need for pupil segmentation under outdoor natural illumination. However, public datasets that simultaneously provide high-quality pixel-level annotations for the iris, pupil, sclera, scleral vessels, eyelids, and periocular structures remain limited. This limitation still restricts fair comparison of multi-region ocular segmentation methods and hinders the development of robust multimodal ocular biometric systems.

4.6. Evaluation metrics and their limitations

For ocular region segmentation, commonly used evaluation metrics include Dice coefficient, Intersection over Union (IoU), mean IoU, pixel accuracy, precision, and recall [75]. Let P and G denote the predicted segmentation mask and the ground-truth mask, respectively. Dice and IoU are commonly defined as:

$$\mathrm{Dice} = \frac{2|P \cap G|}{|P| + |G|}$$

$$\mathrm{IoU} = \frac{|P \cap G|}{|P \cup G|}$$

For pixel-wise classification, precision and recall can be computed as:

$$\mathrm{Precision} = \frac{TP}{TP + FP}, \quad \mathrm{Recall} = \frac{TP}{TP + FN}$$

For multi-class ocular segmentation, mean IoU can be further calculated by averaging the IoU values over all classes:

$$\mathrm{mIoU} = \frac{1}{C} \sum_{c=1}^{C} \frac{|P_c \cap G_c|}{|P_c \cup G_c|}$$

where C denotes the number of classes and P_c and G_c denote the predicted and ground-truth regions of class c, respectively. These overlap-based metrics are intuitive and effective for measuring the overall agreement between predicted masks and ground-truth annotations. However, they may be insufficient for evaluating fine-grained ocular structures. In iris segmentation, small deviations along the pupillary boundary or limbic boundary may only slightly affect Dice or IoU, but they can significantly influence iris normalization, texture encoding, and subsequent biometric matching. Similarly, thin scleral vessels, eyelid contours, and specular-reflection boundaries may be poorly localized even when the overall region overlap remains high.

Therefore, evaluation protocols for ocular biometric segmentation should combine region-overlap metrics with boundary-sensitive metrics. Hausdorff Distance, Average Symmetric Surface Distance [76], Boundary IoU, boundary F-score, and contour localization error [77] can provide complementary information about boundary accuracy and shape consistency. For iris-oriented applications, separate evaluation of inner and outer iris boundary localization is also recommended, because these boundaries directly determine the normalized iris texture used for recognition. In multimodal ocular biometrics, segmentation quality should be further interpreted together with downstream recognition metrics, such as EER, False Acceptance Rate, False Rejection Rate, and rank-based identification accuracy. This is important because a visually accurate or high-Dice segmentation mask does not always guarantee optimal biometric recognition performance.

In summary, existing ocular datasets provide important support for controlled iris recognition, unconstrained VIS acquisition, cross-spectral evaluation, periocular analysis, and eye-region segmentation. Recent datasets from 2020 to 2026 further extend the field toward VR/AR head-mounted devices, mobile

Table 3. Recent ocular region datasets from 2020 to 2026

Dataset (year)	Modality	Setting	Scale (Imgs/Subj)	Resolution	Main features and applications
OpenEDS2020 (2020)	NIR (Eye/Perioc.)	Ctrl. (VR-HMD)	Seg.: up to 29,500/80; Gaze: 550,400/80	Device-dependent	VR-based eye-image sequences with gaze vectors and semantic labels for eye segmentation; useful for temporal eye segmentation and gaze prediction
Mobile Periocular Dataset (2020)	VIS (Perioc.)	Unctrl. (Mobile)	1,122 subjects; 3 sessions; 196 devices	Variable	Large-scale mobile periocular dataset for unconstrained recognition and device-variability evaluation
TEyeD (2021)	Eye region (Pupil/Iris /Eyelids)	HMD/VR/AR	>20M images	Device-dependent	Large-scale HMD eye dataset with semantic segmentation, 2D/3D landmarks, eyeball annotations, gaze vectors, and eye movement labels
CASIA-Iris-Africa (2023)	NIR (Iris)	Ctrl.	28,717/1,023	Not specified	Large-scale African iris database with age, gender, and ethnicity attributes; useful for demographic diversity and bias evaluation
VRBiom (2024)	NIR (Iris/Perioc.)	HMD	900 videos/25; plus PA videos	400×400	HMD-based periocular videos for iris/periocular recognition, detection, semantic segmentation, and presentation-attack analysis
AmbientEye (2026)	Passive IR (Pupil)	Outdoor/Unctrl.	2,606,225/35	Variable	Outdoor passive-infrared pupil segmentation dataset with SAM2-assisted and human-refined pupil annotations

Note: VIS, visible spectrum; NIR, near-infrared; Passive IR, Passive Infrared; Ctrl., controlled; HMD, head-mounted display; Unctrl., uncontrolled; PA, presentation attack; Perioc., periocular.

acquisition, outdoor ambient illumination, demographic diversity, and more realistic eye-region analysis. Nevertheless, datasets with dense and consistent annotations for multiple ocular biometric regions remain scarce, particularly for joint segmentation of the iris, pupil, sclera, scleral vessels, eyelids, and periocular structures. In addition, current benchmarks still rely heavily on region-overlap metrics such as Dice and IoU, which may not fully reflect boundary precision or downstream recognition reliability. Future benchmark construction should therefore combine diverse acquisition protocols, fine-grained multi-region annotations, boundary-sensitive evaluation metrics, and recognition-oriented validation to better support deployable multimodal ocular biometric systems.

5. Challenges and Outlook

Despite substantial progress, unconstrained ocular biometrics remains constrained by variable acquisition quality, cross-spectral and cross-sensor domain shift, limited dataset diversity, and heterogeneity across ocular modalities. Addressing these limitations requires advances in data acquisition, robust modeling, fusion strategy design, and edge-efficient deployment.

In real-world settings, blinking, occlusions, motion blur, illumination variation, and off-axis gaze jointly impair segmentation and matching. Although attention-augmented encoder–decoder models improve robustness to partial occlusion [13], severe occlusion and large gaze deviations remain challenging. Future work should incorporate explicit quality and occlusion estimation, quality-aware training, and acquisition-aware operational strategies, including selective re-capture and adaptive thresholding.

Cross-spectral gaps (VIS vs NIR) and cross-sensor discrepancies introduce substantial differences in appearance and noise statistics, which hinder generalization. Learning-based alignment is promising [78], but reliable transfer will likely require stronger domain-adaptation objectives, uncertainty-aware alignment, and fusion schemes that tolerate sensor heterogeneity.

Dataset bias and limited covariate coverage remain major bottlenecks [79]. Large-scale annotations under unconstrained capture are still scarce. Synthetic data generation can broaden covariate coverage but requires careful validation to prevent artifacts [80]. Multimodal fusion is further constrained by cross-modal misalignment and scale discrepancies among iris, sclera, pupil, and periocular cues; quality-guided fusion and feature reconstruction provide initial directions [81], yet robust alignment under distortion and missing modalities remains an open problem. Overall, continued progress will depend on larger and more diverse datasets, principled alignment and fusion, and lightweight optimization—including pruning, quantization, and distillation—to support real-time deployment.

6. Conclusion

This review has summarized ocular region segmentation and multimodal fusion from the perspective of biometric recognition. The reviewed studies indicate that segmentation is not merely a preprocessing step but an important basis for extracting and integrating biometric cues from the iris, pupil, sclera, and periocular region. Traditional image-processing methods remain useful in controlled and resource-limited scenarios due to their simplicity and efficiency, whereas deep learning and foundation-model-based methods offer stronger representation capability but require annotated data, computational resources, and domain adaptation.

For multimodal ocular biometric recognition, data-level fusion is suitable for well-aligned multimodal inputs, feature-level fusion provides a balanced trade-off between complementary representation and computational efficiency, and decision-level fusion is more appropriate for modular systems with heterogeneous sensors or independent matchers. These differences suggest that segmentation and fusion strategies should be selected according to acquisition conditions, modality alignment, computational constraints, and deployment requirements.

Future research should further address the limitations of existing datasets and evaluation protocols. In particular, more diverse datasets with fine-grained multi-region annotations are needed, and overlap-based metrics such as Dice and IoU should be complemented by boundary-sensitive metrics and downstream recognition performance. Progress in quality-aware segmentation, cross-spectral and cross-sensor adaptation, missing-modality-tolerant fusion, and lightweight deployment will be essential for developing more robust and practical ocular biometric systems.

Funding

This work was supported by the National College Students Innovation and Entrepreneurship Training Program (No. 2026J00246) and the Beijing Municipal College Students Innovation and Entrepreneurship Training Program (No. 2026J0 0247).

Ethical Statement

This manuscript presents a comprehensive review of ocular biometric segmentation and multimodal fusion methods and does not report any original experiments involving human or animal subjects conducted by the authors. The datasets described in this review, including ocular image databases for iris, sclera, pupil, and periocular analysis, were obtained from publicly available resources released by the original dataset developers. According to the information provided by the original publications and dataset documentation, data collection procedures were conducted following the ethical requirements and institutional regulations applicable at the time of acquisition.

The authors did not access private participant information, recruit human subjects, collect new biometric samples, or perform additional analyses on identifiable human data. Therefore, no new ethical approval or informed consent was required for this review article.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

No new data were generated or collected in this review article. This study is based exclusively on previously published literature and publicly available ocular biometric datasets. The datasets mentioned in this review, including iris, sclera, pupil, and periocular image databases, can be accessed through the original publications or repositories maintained by their respective dataset providers. The authors did not create, modify, or redistribute any dataset during this study.

Author Contribution Statement

Yansuo Yu: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Yongbin Qi:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Shilin Zhao:** Validation, Formal analysis, Investigation, Resources, Data curation, Visualization. **Huanzhang Qi:** Validation, Formal analysis, Investigation, Resources, Data curation, Visualization. **Wendao Li:** Resources, Visualization. **Haoqi Zhang:** Resources, Visualization. **Da Teng:** Validation, Formal analysis, Resources, Supervision, Project administration. **Qiang Liu:** Supervision, Project administration, Funding acquisition.

References

[1] Xu, H. (2026). Multimodal identity recognition technology based on Deep Reinforcement Learning. *Discover Artificial Intelligence*. https://doi.org/10.1007/s44163-026-01457-z

[2] Alrawili, R., AlQahtani, A. A. S., & Khan, M. K. (2024). Comprehensive survey: Biometric user authentication application, evaluation, and discussion. *Computers and Electrical Engineering*, *119*, 109485. https://doi.org/10.1016/j.compeleceng.2024.109485

[3] Lokhande, M., Chopade, M., Desale, N., Khedkar, M., & Ghodke, P. (2024). Prevention of spoofing attack through fingerprint liveness detection system. In *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 574–579. https://doi.org/10.1109/ICSCSS60660.2024.10625540

[4] Nguyen, K., Proença, H., & Alonso-Fernandez, F. (2024). Deep learning for iris recognition: A survey. *ACM Computing Surveys*, *56*(9), 1–35. https://doi.org/10.1145/3651306

[5] Önal, M. N., Güraksin, G. E., & Duman, R. (2023). Convolutional neural network-based diabetes diagnostic system via iridology technique. *Multimedia Tools and Applications*, *82*(1), 173–194. https://doi.org/10.1007/s11042-022-13291-3

[6] Lee, S., Low, C. Y., Kim, J., & Teoh, A. B. J. (2022). Robust sclera recognition based on a local spherical structure. *Expert Systems with Applications*, *189*, 116081. https://doi.org/10.1016/j.eswa.2021.116081

[7] Park, S., Ha, J., & Kim, L. (2024). Event-related pupillary response-based authentication system using eye-tracker add-on augmented reality glasses for individual identification. *Frontiers in Physiology*, *15*, 1325784. https://doi.org/10.3389/fphys.2024.1325784

[8] Zhou, W., Ru, Y., Han, Y., Kong, L., Wang, Z., He, Y., & Sun, Z. (2025). Casia-pr-v1: A multi-ethnic, multi-device and cross-spectral dataset and a multiscale disentangled model for periocular recognition. *IEEE Transactions on Multimedia*, *27*, 7192–7204. https://doi.org/10.1109/TMM.2025.3599084

[9] El-Sofany, H., Bouallegue, B., & Abd El-Latif, Y. M. (2024). A proposed biometric authentication hybrid approach using iris recognition for improving cloud security. *Heliyon*, *10*(16), e36390. https://doi.org/10.1016/j.heliyon.2024.e36390

[10] Kolus, A. (2024). A systematic review on driver drowsiness detection using eye activity measures. *IEEE Access*, *12*, 97969–97993. https://doi.org/10.1109/ACCESS.2024.3424654

[11] Rasheed, H. H., Shamini, S. S., Mahmoud, M. A., & Alomari, M. A. (2023). Review of iris segmentation and recognition using deep learning to improve biometric application. *Journal of Intelligent Systems*, *32*(1), 20230139. https://doi.org/10.1515/jisys-2023-0139

[12] Shaheed, K., Szczuko, P., Kumar, M., Qureshi, I., Abbas, Q., & Ullah, I. (2024). Deep learning techniques for biometric security: A systematic review of presentation attack detection systems. *Engineering Applications of Artificial Intelligence*, *129*, 107569. https://doi.org/10.1016/j.engappai.2023.107569

[13] Tanna, R., Patel, T., Alotaibi, F. M., Jhaveri, R. H., & Gadekallu, T. R. (2026). OcclusionNetPlusPlus: A multi-scale similarity network with adaptive occlusion detection for robust iris recognition. *International Journal of Cognitive Computing in Engineering*, *7*, 74–85. https://doi.org/10.1016/j.ijcce.2025.09.002

[14] Ow Tiong, L. C., Sigmund, D., Chan, C.-H., & Jin Teoh, A. B. (2024). Flexible biometrics recognition: Bridging the multimodality gap through attention, alignment and prompt tuning. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 267–276. https://doi.org/10.1109/CVPR52733.2024.00033

[15] Ali Naqvi, R., Hussain, D., & Loh, W.-K. (2020). Artificial intelligence-based semantic segmentation of ocular regions for biometrics and healthcare applications. *Computers, Materials & Continua*, *66*(1), 715–732. https://doi.org/10.32604/cmc.2020.013249

[16] Abualigah, L., Almotairi, K. H., & Elaziz, M. A. (2023). Multilevel thresholding image segmentation using meta-heuristic optimization algorithms: Comparative analysis, open challenges and new trends. *Applied Intelligence*, *53*(10), 11654–11704. https://doi.org/10.1007/s10489-022-04064-4

[17] Ragunathan, R., & Sankaradass, V. (2026). A novel image segmentation framework using adaptive region-based unsupervised machine learning mechanism. *Journal of Clinical Ultrasound*, 1–32. https://doi.org/10.1002/jcu.70238

[18] Huang, Q., & Huang, J. (2025). Comprehensive review of edge and contour detection: From traditional methods to recent advances. *Neural Computing and Applications*, *37*(4), 2175–2209. https://doi.org/10.1007/s00521-024-10936-2

[19] Beevi, J. H., Jafar, O. A. M., & Shanavas, A. R. M. (2025). A survey of image segmentation based on evolutionary computation and clustering techniques. *International Journal of Computer Applications*, *186*(80), 11–18. https://doi.org/10.5120/ijca2025924739

[20] Sun, S., Huang, Y., Inoue, K., & Hara, K. (2023). Order space-based morphology for color image processing. *Journal of Imaging*, *9*(7), 139. https://doi.org/10.3390/jimaging9070139

[21] Ooi, A. Z. H., Embong, Z., Abd Hamid, A. I., Zainon, R., Wang, S. L., Ng, T. F., . . . , & Ibrahim, H. (2021). Interactive blood vessel segmentation from retinal fundus image based on Canny edge detector. *Sensors*, *21*(19), 6380. https://doi.org/10.3390/s21196380

[22] Shaikh, M., Khan, S., & Singh, K. (2022). An optimal cluster fusion and morphological segmentation technique for iris recognition. In *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)*, 1–6. https://doi.org/10.1109/CCET56606.2022.10080242

[23] Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, *9*, 82031–82057. https://doi.org/10.1109/ACCESS.2021.3086020

[24] Cui, Z., Wei, Y., Li, T., & Li, G. (2022). Image segmentation algorithm based on attention mechanism and jump connection. In *2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)*, 260–262. https://doi.org/10.1109/MLISE57402.2022.00058

[25] Du, S., Hao, K., Zhang, H., Tang, X., & Wei, B. (2022). Patch elastic deformation: An effective data augmentation method. *2022 China Automation Congress (CAC)*, 2079–2084. https://doi.org/10.1109/CAC57257.2022.10055297

[26] Zhang, Z., Zhao, T., Gay, H., Zhang, W., & Sun, B. (2021). Weaving attention U-net: A novel hybrid CNN and attention-based method for organs-at-risk segmentation in head and neck CT images. *Medical Physics*, *48*(11), 7052–7062. https://doi.org/10.1002/mp.15287

[27] Zhu, Y., Xu, X., Zhang, X., & Jiang, M. (2023). CCS-UNet: A cross-channel spatial attention model for accurate retinal vessel segmentation. *Biomedical Optics Express*, *14*(9), 4739. https://doi.org/10.1364/BOE.495766

[28] Jin, T., & Wang, Z. (2023). Multi-scale-ResUNet: An improve u-net with multi-scale attention and hybrid dilation for medical image segmentation. *Multimedia Tools and Applications*, *82*(18), 28473–28492. https://doi.org/10.1007/s11042-023-14404-2

[29] Zhang, X., Lin, Y., Li, L., Zeng, J., Lan, X., Zhang, X., . . . , & Zhu, H. (2024). MA-ResUNet: Multi-attention optic cup and optic disc segmentation based on improved U-Net. *IET Image Processing*, *18*(12), 3128–3142. https://doi.org/10.1049/ipr2.13160

[30] Yang, Z., Xu, P., Yang, Y., & Bao, B.-K. (2021). A densely connected network based on u-net for medical image segmentation. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *17*(3), 1–14. https://doi.org/10.1145/3446618

[31] Du, X., Wang, J., & Sun, W. (2021). Densely connected U-Net retinal vessel segmentation algorithm based on

multi-scale feature convolution extraction. *Medical Physics*, *48*(7), 3827–3841. https://doi.org/10.1002/mp.14944

[32] Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G., & Zhang, D. (2022). DS-TransUNet: Dual swin transformer U-Net for medical image segmentation. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–15. https://doi.org/10.1109/TIM.2022.3178991

[33] Hsu, C.-Y., Hu, R., Xiang, Y., Long, X., & Li, Z. (2022). Improving the DeepLabv3+ model with attention mechanisms applied to eye detection and segmentation. *Mathematics*, *10*(15), 2597. https://doi.org/10.3390/math10152597

[34] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

[35] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss. (Eds.), *Computer Vision – ECCV 2018* (*11211*, pp. 833–851). Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_49

[36] Qi, Y., Zhen, B., Wang, J., Zhao, S., Yu, Y., & Liu, Q. (2025). Deep learning for joint scleral and scleral vessel segmentation: A comparative study. *Medinformatics*, *3*(1), 10–21. https://doi.org/10.47852/bonviewMEDIN52025662

[37] Li, K., Zhang, L., Li, B., Li, S., & Ma, J. (2022). Attention-optimized DeepLab V3 + for automatic estimation of cucumber disease severity. *Plant Methods*, *18*(1), 109. https://doi.org/10.1186/s13007-022-00941-8

[38] Zhang, Y., Shen, Z., & Jiao, R. (2024). Segment anything model for medical image segmentation: Current applications and future directions. *Computers in Biology and Medicine*, *171*, 108238. https://doi.org/10.1016/j.compbiomed.2024.108238

[39] Farmanifard, P., & Ross, A. (2025). Iris-SAM: Iris segmentation using a foundation model. In C. Wallraven, C.-L. Liu, & A. Ross. (Eds.), *Pattern Recognition and Artificial Intelligence* (*14892*, pp. 394–409). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-8702-9_27

[40] Jin, X., Hu, J., Lin, J., Zhang, S., & Cao, L. (2025). U-SAM: Upgrade segment anything model with semantic-aware and memory-efficient. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. https://doi.org/10.1109/ICASSP49660.2025.10889270

[41] Zhou, H., He, Y., Cui, X., & Xie, Z. (2024). AGSAM: Agent-guided segment anything model for automatic segmentation in few-shot scenarios. *Bioengineering*, *11*(5), 447. https://doi.org/10.3390/bioengineering11050447

[42] Yong, Y., Zhao, S., Shuying, H., Weiguo, W., Wei, T., & Hangyuan, L. (2024). Deep-learning approaches for pixel-level pansharpening. *National Remote Sensing Bulletin*, *26*(12), 2411–2432. https://doi.org/10.11834/jrs.20211325

[43] Singh, S., Singh, H., Bueno, G., Deniz, O., Singh, S., Monga, H., . . . , & Pedraza, A. (2023). A review of image fusion: Methods, applications and performance metrics. *Digital Signal Processing*, *137*, 104020. https://doi.org/10.1016/j.dsp.2023.104020

[44] Mostofa, M., Mohamadi, S., Dawson, J., & Nasrabadi, N. M. (2021). Deep GAN-based cross-spectral cross-resolution iris recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *3*(4), 443–463. https://doi.org/10.1109/TBIOM.2021.3102736

[45] Wei, J., Wang, Y., Li, Y., He, R., & Sun, Z. (2022). Cross-spectral iris recognition by learning device-specific band. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(6), 3810–3824. https://doi.org/10.1109/TCSVT.2021.3117291

[46] Es-Sobbahi, H., Radouane, M., & Nafil, K. (2025). Multimodal biometrics: A review of handcrafted and AI–based fusion approaches. *IET Biometrics*, *2025*(1), 5055434. https://doi.org/10.1049/bme2/5055434

[47] Xiao, R., Yuan, S., Jin, K., Li, M., Tang, Y., & Shen, S. (2025). Cmfnet: A three-stage feature matching network with geometric consistency and attentional enhancement. *IET Image Processing*, *19*(1), e70050. https://doi.org/10.1049/ipr2.70050

[48] Luo, Z., Li, J., & Zhu, Y. (2021). A deep feature fusion network based on multiple attention mechanisms for joint iris-periocular biometric recognition. *IEEE Signal Processing Letters*, *28*, 1060–1064. https://doi.org/10.1109/LSP.2021.3079850

[49] Vyas, R. (2022). Enhanced near-infrared periocular recognition through collaborative rendering of hand crafted and deep features. *Multimedia Tools and Applications*, *81*(7), 9351–9365. https://doi.org/10.1007/s11042-021-11846-4

[50] Jian, X., & Yuan, L. (2023). Research and application of multimodal data fusion in medical images. In *2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 128–132. https://doi.org/10.1109/ITOEC57671.2023.10291374

[51] Xiao, G., Bavirisetti, D. P., Liu, G., & Zhang, X. (2020). Decision-level image fusion. In G. Xiao, D. P. Bavirisetti, G. Liu, & X. Zhang (Eds.), *Image Fusion* (pp. 149–170). Springer Singapore. https://doi.org/10.1007/978-981-15-4867-3_4

[52] Alonso-Fernandez, F., Raja, K. B., Raghavendra, R., Busch, C., Bigun, J., Vera-Rodriguez, R., & Fierrez, J. (2022). Cross-sensor periocular biometrics in a global pandemic: Comparative benchmark and novel multialgorithmic approach. *Information Fusion*, *83-84*, 110–130. https://doi.org/10.1016/j.inffus.2022.03.008

[53] Farmanifard, P., & Ross, A. (2024). ChatGPT meets iris biometrics. In *2024 IEEE International Joint Conference on Biometrics (IJCB)*, 1–10. https://doi.org/10.1109/IJCB62174.2024.10744525

[54] Bhuyian, R. A., Farmanifard, P., Sharma, R., Kuehlkamp, A., Boyd, A., Flynn, P. J., . . . , & Czajka, A. (2025). Beyond mortality: Advancements in post-mortem iris recognition through data collection and computer-aided forensic examination. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *7*(4), 808–823. https://doi.org/10.1109/TBIOM.2025.3585093

[55] Baker, S. E., Hentz, A., Bowyer, K. W., & Flynn, P. J. (2010). Degradation of iris recognition performance due to non-cosmetic prescription contact lenses. *Computer Vision and Image Understanding*, *114*(9), 1030–1044. https://doi.org/10.1016/j.cviu.2010.06.002

[56] Fenker, S. P., & Bowyer, K. W. (2012). Analysis of template aging in iris biometrics. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 45–51. https://doi.org/10.1109/CVPRW.2012.6239214

[57] Kumar, A., & Passi, A. (2010). Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recognition*, *43*(3), 1016–1026. https://doi.org/10.1016/j.patcog.2009.08.016
[58] Proença, H., & Alexandre, L. A. (2005). Ubiris: A noisy iris image database. In F. Roli & S. Vitulano. (Eds.), *Image Analysis and Processing – ICIAP 2005* (*3617*, pp. 970–977). Springer Berlin Heidelberg. https://doi.org/10.1007/11553595_119
[59] Proenca, H., Filipe, S., Santos, R., Oliveira, J., & Alexandre, L. A. (2010). The ubiris. V2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(8), 1529–1535. https://doi.org/10.1109/TPAMI.2009.66
[60] De Marsico, M., Nappi, M., Riccio, D., & Wechsler, H. (2015). Mobile Iris Challenge Evaluation (MICHE)-I, biometric iris dataset and protocols. *Pattern Recognition Letters*, *57*, 17–23. https://doi.org/10.1016/j.patrec.2015.02.009
[61] Rattani, A., Derakhshani, R., Saripalle, S. K., & Gottemukkula, V. (2016). ICIP 2016 competition on mobile ocular biometric recognition. In *2016 IEEE International Conference on Image Processing (ICIP)*, 320–324. https://doi.org/10.1109/ICIP.2016.7532371
[62] Hosseini, M. S., Araabi, B. N., & Soltanian-Zadeh, H. (2010). Pigment melanin: Pattern for iris recognition. *IEEE Transactions on Instrumentation and Measurement*, *59*(4), 792–804. https://doi.org/10.1109/TIM.2009.2037996
[63] Zhao, Z., & Kumar, A. (2017). Towards more accurate iris recognition using deeply learned spatially corresponding features. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 3829–3838. https://doi.org/10.1109/ICCV.2017.411
[64] Sequeira, A., Chen, L., Wild, P., Ferryman, J., Alonso-Fernandez, F., Raja, K. B., . . . , & Bigun, J. (2016). Cross-eyed—Cross-spectral iris/periocular recognition database and competition. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 1–5. https://doi.org/10.1109/BIOSIG.2016.7736915
[65] Das, S., De Ghosh, I., & Chattopadhyay, A. (2022). Sclera biometrics in restricted and unrestricted environment with cross dataset evaluation. *Displays*, *74*, 102257. https://doi.org/10.1016/j.displa.2022.102257
[66] Peng, Z., Xu, J., Li, S., Ji, J., Huang, Y., Zhang, J., & Ma, L. (2025). EyeSeg: An uncertainty-aware eye segmentation framework for AR/VR. *arXiv Preprint: 2507.09649.*
[67] Fierrez, J., Galbally, J., Ortega-Garcia, J., Freire, M. R., Alonso-Fernandez, F., Ramos, D., . . . , & Gracia-Roche, J. J. (2010). BiosecurID: A multimodal biometric database. *Pattern Analysis and Applications*, *13*(2), 235–246. https://doi.org/10.1007/s10044-009-0151-4
[68] Ortega-Garcia, J., Fierrez, J., Alonso-Fernandez, F., Galbally, J., Freire, M. R., Gonzalez-Rodriguez, J., . . . , & Savran, A. (2009). The multiscenario multienvironment biosecure multimodal database (BMDB). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(6), 1097–1111. https://doi.org/10.1109/TPAMI.2009.76
[69] Palmero, C., Sharma, A., Behrendt, K., Krishnakumar, K., Komogortsev, O. V., & Talathi, S. S. (2021). OpenEDS2020 challenge on gaze tracking for VR: Dataset and results. *Sensors*, *21*(14), 4769. https://doi.org/10.3390/s21144769
[70] Lohr, D., Aziz, S., Friedman, L., & Komogortsev, O. V. (2023). GazeBaseVR, a large-scale, longitudinal, binocular eye-tracking dataset collected in virtual reality. *Scientific Data*, *10*(1), 177. https://doi.org/10.1038/s41597-023-02075-5
[71] Kotwal, K., Ulucan, I., Özbulak, G., Selliah, J., & Marcel, S. (2025). VRBiom: A new periocular dataset for biometric applications of head-mounted display. *Electronics*, *14*(9), 1835. https://doi.org/10.3390/electronics14091835
[72] Zanlorensi, L. A., Laroca, R., Lucio, D. R., Santos, L. R., Britto Jr, A. S., & Menotti, D. (2022). A new periocular dataset collected by mobile devices in unconstrained scenarios. *Scientific Reports*, *12*(1), 17989. https://doi.org/10.1038/s41598-022-22811-y
[73] Muhammad, J., Wang, Y., Hu, J., Zhang, K., & Sun, Z. (2024). CASIA-iris-africa: A large-scale African iris image database. *Machine Intelligence Research*, *21*(2), 383–399. https://doi.org/10.1007/s11633-022-1402-8
[74] Chen, S., & Epps, J. (2023). A high-quality landmarked infrared eye video dataset (IREye4Task): Eye behaviors, insights and benchmarks for wearable mental state analysis. *IEEE Transactions on Affective Computing*, *14*(4), 3078–3093. https://doi.org/10.1109/TAFFC.2023.3258915
[75] Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, *15*(1), 210. https://doi.org/10.1186/s13104-022-06096-y
[76] Podobnik, G., & Vrtovec, T. (2024). Understanding implementation pitfalls of distance-based metrics for image segmentation. *arXiv Preprint: 2410.02630.*
[77] Cheng, B., Girshick, R., Dollar, P., Berg, A. C., & Kirillov, A. (2021). Boundary IoU: Improving object-centric image segmentation evaluation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15329–15337. https://doi.org/10.1109/CVPR46437.2021.01508
[78] Tapia, J. E., Gonzalez, S., & Busch, C. (2021). Iris liveness detection using a cascade of dedicated deep learning networks. *IEEE Transactions on Information Forensics and Security*, *17*, 42–52. https://doi.org/10.1109/TIFS.2021.3132582
[79] Sumi, M. R., Das, P., Hossain, A., Dey, S., & Schuckers, S. (2024). A comprehensive evaluation of iris segmentation on benchmarking datasets. *Sensors*, *24*(21), 7079. https://doi.org/10.3390/s24217079
[80] Xun, S., Li, D., Zhu, H., Chen, M., Wang, J., Li, J., . . . , & Huang, P. (2022). Generative adversarial networks in medical image segmentation: A review. *Computers in Biology and Medicine*, *140*, 105063. https://doi.org/10.1016/j.compbiomed.2021.105063
[81] Yan, Z., Wang, Y., Zhang, K., Sun, Z., & He, L. (2024). Boosting multi-modal ocular recognition via spatial feature reconstruction and unsupervised image quality estimation. *Machine Intelligence Research*, *21*(1), 197–214. https://doi.org/10.1007/s11633-023-1415-y

How to Cite: Yu, Y., Qi, Y., Zhao, S., Qi, H., Li, W., Zhang, H., …, & Liu, Q. (2026). Segmentation and Multimodal Fusion in Ocular Biometric Recognition: A Review. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN62029908