Image and Vision Computing 108 (2021) 104109

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

A survey of iris datasets

Lubos Omelina [a,b,c,*], Jozef Goga [b], Jarmila Pavlovicova [b], Milos Oravec [b], Bart Jansen [a,c]

[a] Department of Electronics and Informatics, Vrije Universiteit Brussel, Pleinlaan 2 1050, Brussels, Belgium
[b] Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Ilkovicova 3, 812 19 Bratislava, Slovakia
[c] imec, Kapeldreef 75, B-3001, Leuven, Belgium

ARTICLE INFO

Article history:
Received 30 December 2019
Received in revised form 10 October 2020
Accepted 6 January 2021
Available online 23 January 2021

Keywords:
Biometrics
Iris recognition
Iris datasets
Human iris

ABSTRACT

Research on human eye image processing and iris recognition has grown steadily over the last few decades. It is important for researchers interested in this discipline to know the relevant datasets in this area to (i) be able to compare their results and (ii) speed up their research using existing datasets rather than creating custom datasets. In this paper, we provide a comprehensive overview of the existing publicly available datasets and their popularity in the research community using a bibliometric approach. We reviewed 158 different iris datasets referenced from the 689 most relevant research articles indexed by the *Web of Science* online library. We categorized the datasets and described the properties important for performing relevant research. We provide an overview of the databases per category to help investigators conducting research in the domain of iris recognition to identify relevant datasets.

1. Introduction

Publicly available datasets of human iris images play a major role in research into iris recognition. Most of the available datasets share a substantial number of properties (e.g., near-infrared imaging) and meet the requirements of the widespread and de facto standard recognition method introduced by John Daugman [1]. With the recent popularity of mobile computing and deep learning in biometrics, new databases have been introduced, containing more challenging images of irises. It can be difficult for newcomers to the iris recognition field to identify the major and appropriate databases suitable for their research topics. When entering or conducting research in biometrics, researchers typically go through an extensive amount of published work to identify the state-of-the-art methods, data sources, and benchmark datasets. However, many of the datasets, either very popular benchmarks or niche datasets, are not available, despite the claims of the authors, due to a variety of reasons. Although there are public search engines[1] providing access to freely available research datasets, biometric datasets are typically not included. Due to the personal nature of the data, the dataset providers typically allow their use only for noncommercial research purposes. In addition, authors typically follow the access carefully and require a signature from the researcher or a legal representative of the research institution. This adds additional constraints that limit the popularity of certain datasets among researchers.

The main purpose of this work is to help navigate among the 158 databases used in iris recognition research that are often declared to be publicly available or not explicitly stated otherwise. In this paper, we review existing and publicly available (for research purposes) datasets of human irises. We categorized the datasets based on the research areas for which they are suitable. In particular, we focus on the imaging process with respect to current trends in imaging. We created a list of the databases by (i) reviewing the relevant journal papers indexed by the *Web of Science* library and (ii) by searching through online search engines. We also analyze the popularity of the databases and, based on that analysis, we discuss trends in iris imaging. Within the analysis, we also critically reviewed the databases to understand their suitability for particular iris recognition research tasks.

In addition to reviewing the datasets, we also attempted to identify the original reference (the first publication, if it exists) introducing the dataset as well as the earliest research performed and published using the dataset. We also report the number of classes and iris images contained in each dataset, as these are often the most important properties (for data-driven research, e.g., machine-learning approaches).

We reviewed 689 papers on iris recognition or related research from the most relevant journals. Based on the review, we aim to answer the following research questions (RQs):

1. RQ 1: What are the existing and available databases?
2. RQ 2: What are the most popular iris databases?
3. RQ 3: What are the differences, downsides, and commonalities among the databases?

* Corresponding author.
E-mail addresses: lomelina@etrovub.be (L. Omelina), jozef.goga@stuba.sk (J. Goga), jarmila.pavlovicova@stuba.sk (J. Pavlovicova), milos.oravec@stuba.sk (M. Oravec), bjansen@etrovub.be (B. Jansen).

[1] An example of a dataset search engine is *Google Dataset Search*: https://toolbox.google.com/datasetsearch

https://doi.org/10.1016/j.imavis.2021.104109
0262-8856/

4. RQ 4: What are the common properties of the popular databases?
5. RQ 5: What areas in the field of iris recognition lack an available database?
6. RQ 6: What are appropriate recommendations for creating an iris database?

The rest of this article is organized as follows. Section 2 provides a description of related work, mainly other reviews or surveys related to iris recognition. In section 3, we present the results of a bibliometric analysis where we answer RQ 1 and RQ 2. Section 4, reviewing existing databases and answering RQ 3, presents a critical review and comparison of existing iris databases. In section 5, we describe common attributes of the popular databases, answering RQ 4. In section 6, we discuss limitations and areas with underdeveloped datasets (answering RQ 5) and formulate general recommendations for creating new datasets (RQ 6) to reach scientific relevance (section 8).

2. Related work

There are multiple review papers on the topic of iris recognition and eye image processing. These surveys focus on processing and recognition and devote only a limited space to discussing available datasets.

Nguyen et al. [2] provided an extensive survey of long-range iris recognition. They discuss existing systems and their limitations; however, they only briefly discuss three publicly available datasets, MBGC, UBIRIS V2.0, and CASIA-Iris-Distance. The authors discuss limitations of the recognition methods. However, it is not clear whether the presented datasets are sufficient for future research (e.g., due to limitations in hardware, mainly sensors and optics).

Alonso-Fernandez & Bigun [3] reviewed research related to periocular biometrics. The authors briefly describe five publicly available iris datasets and four periocular datasets (often also used for iris recognition research, although the captured iris is typically very small, 50 pixels). The authors also point out the limitations of sensors and see a future for imaging at a distance, but do not explain why the selected databases would have future perspectives.

While Farmanullah [4] provided an extensive review of segmentation methods for non-ideal and unconstrained biometric iris systems, together with results on 21 datasets, he omits any description of the datasets used or reasons for selecting them.

De Marsico et al. [5] provided a survey of machine-learning methods for iris recognition. While the authors provide results on 11 publicly available datasets, actual comparisons and descriptions of the datasets are absent. While most of the results are performed on the CASIA Iris Dataset v.1, they concluded that more extensive experimentation on the later datasets is needed.

In their survey on understanding iris images, Bowyer et al. [6] selected and described 10 datasets. However, one of them is no longer available (BATH), and two datasets, ICE2005 and ICE2006, are available only within a much larger dataset, ND-IRIS-0405, where it is not clear which files correspond to which particular datasets.[2]

Neves et al. [7] described biometric recognition in surveillance scenarios. The authors refer to four iris datasets with limited descriptions. They refer to the datasets as the main biometric datasets; however, as we show later (see Section 3), there are other datasets with higher significance (in terms of impact) for the research community.

Rattani and Derakhshani [8] provided a survey of methods for ocular recognition in the visible spectrum. The authors describe seven datasets collected in the visible spectrum. However, some datasets mentioned by the authors are no longer available, despite the claims of the original authors of the datasets.[3]

[2] The ND-IRIS-0405 dataset contains a description of the files that were used in the ICE 2005 challenge. However, ICE 2006 files are not explicitly described.

[3] For example, despite multiple attempts to obtain the VSSIRIS dataset [9] and contacting the authors, we were not able to download it.

There are other reviews [10,11] on the topic of iris segmentation and recognition; however, the authors do not discuss available datasets.

3. Popularity of databases

Using the most appropriate dataset for a given problem is a basic assumption for the successful validation of any scientific method. To the best of our knowledge, currently there is no extensive review of available iris image datasets; therefore, it is difficult to select the appropriate dataset. The selection process involves extensive and time-consuming research, which is often a reason for creating a custom dataset instead. Authors typically justify the new database by reviewing a few (up to five) empirically chosen popular databases that are often unrelated to the present research (e.g., [12,13]).

In this section, we discuss the popularity of iris image databases based on bibliometric research and provide statistical information about the available ones. The popularity of databases helps identify research areas that receive low attention. The quantified popularity of iris databases is also useful as an indicator of research trends and underdeveloped areas. We define popularity as the citation rank within the selected library. Next, we describe our systematic review of the literature.

Owing to space constraints, the complete list of identified databases (with details) and the list of reviewed publications can be found in the supplementary materials.

3.1. Identification of relevant studies

There are many sources of literature relevant to iris recognition. For the purpose of this survey, we focused on the most scientifically relevant papers, that is, studies that have the highest impact on the research community. Therefore, we selected the *Web of Science* online library, as many of the most-cited publications in iris recognition known to the authors of this study, are indexed by this library.

We selected 1,012 journal articles listed in the *Web of Science* online library. We searched for the simultaneous presence of two keywords: “iris” and “recognition.” We limited the search to (i) English language, (ii) Science Citation Index Expanded, and (iii) only the research areas: COMPUTER SCIENCE, ENGINEERING, IMAGING SCIENCE PHOTOGRAPHIC TECHNOLOGY, OPTICS, TELECOMMUNICATIONS, MATHEMATICAL COMPUTATIONAL BIOLOGY, AUTOMATION CONTROL SYSTEMS, MATHEMATICS, SCIENCE TECHNOLOGY OTHER TOPICS, and ROBOTICS.

3.2. Primary study selection and study quality assessment

From the 1,012 studies, we selected 689 relevant articles, that is, the articles where authors report the use of at least one iris image database in their research. The relevant articles were selected based on the abstract, description of the experiment (most of the time an experimental results section) and conclusion. We excluded:

- articles that appeared in the search results but are not related to research of the human iris (e.g. studies that referred to the popular *iris flower data set*),
- duplicates or the same studies published in different sources, and
- studies that are related to human iris, but do not use any iris database in the research (review studies, iris enrollment methods, etc.).

Some iris databases (e.g., the CASIA Iris Database) are continuously evolving, and the number of included images is increasing. Despite the existence of release versions of the CASIA database, some reviewed papers have used intermediate versions that do not correspond to any version published online. In such cases, we classified the database to the closest release version available in terms of the number of images.

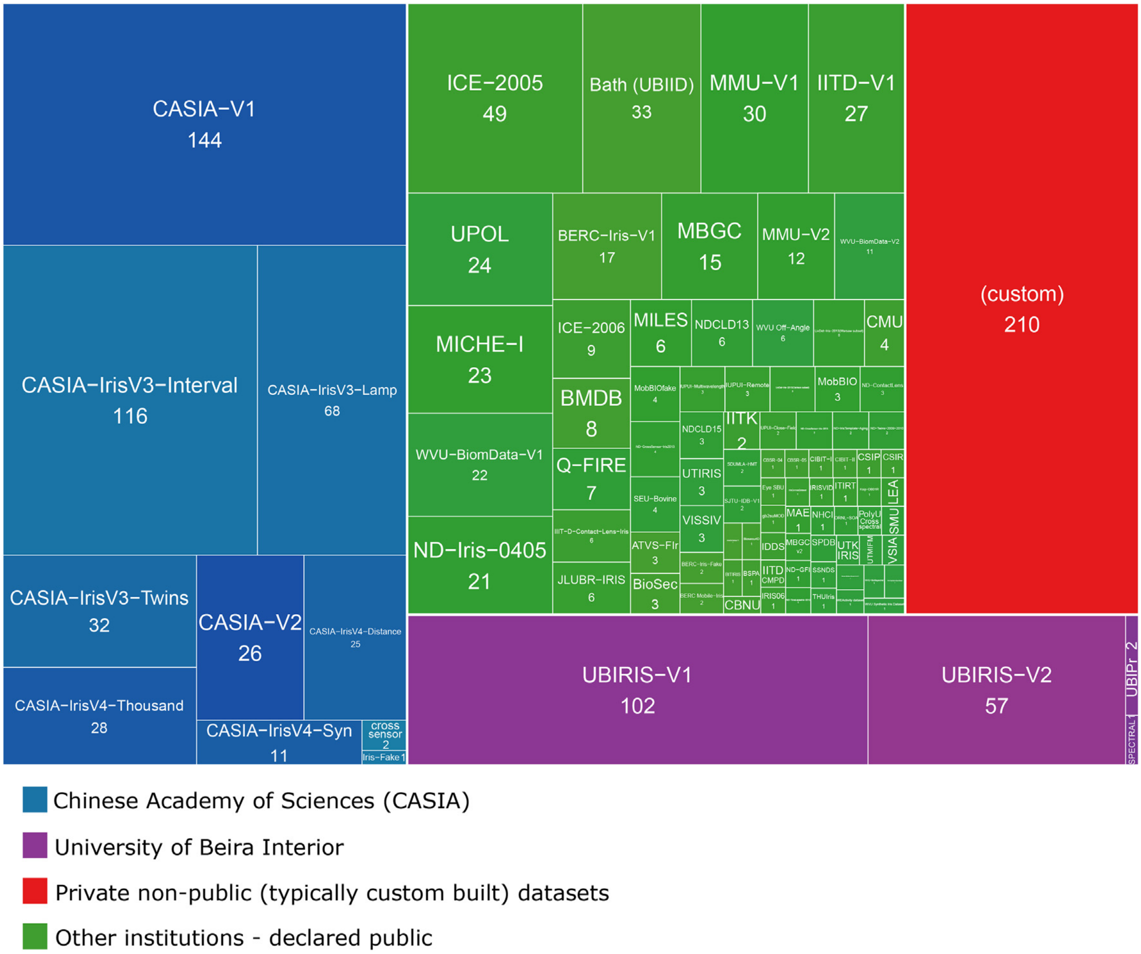

Fig. 1. Popularity of iris image databases. The treemap chart shows the cited databases with the number of citations among the reviewed publications.

3.3. Review results and databases used

Based on our research, we found that 57.37% (397) of the publications evaluated their method only on one database, and the average number of databases used was 1.87 per publication (see Fig. 2 for more details). 30.35% (210) of publications were evaluated on custom (non-published) databases, where 22.54% (156) publications used only a custom database.

The overall popularity (in terms of uses and citations) of the databases is illustrated in Fig. 1. The iris databases produced by the Chinese Academy of Sciences (CASIA) are the most popular and have been used 453 times in 305 articles, or in more than 44% of the studies. Actually, the CASIA Iris Database v. 1, although being obsolete [14], is still the most popular and most cited iris image database (used mostly as a benchmark in 144 (21%) studies).

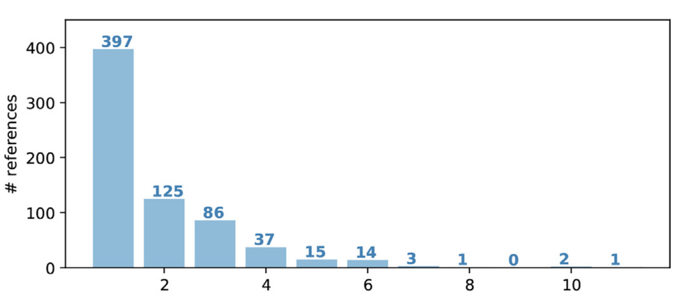

Fig. 2. Histogram of database usage per publication. The average number of databases used in a publication was 1.87.

To help navigate the existing publicly available databases, we provide descriptions and compare several characteristics, focusing on the suitability of the databases for particular research topics. The definition of "publicly available" is rather loose in this domain and is especially sensitive due to the nature of the data. Most of the databases require signing a license agreement to obtain a copy. However, this is understandable, as it provides legal protection for the people whose data is contained in the database as well as for the authors of the databases. In addition, many authors put additional constrains, mainly:

- the placement of an acknowledgment in the published studies that use the database
- a citation reference to the paper introducing the database
- a citation reference to other papers

We searched the publications and online resources for publicly available databases, where we identified 158 existing databases used in the research. Out of 158, 57% 78 of the databases had at least two citations within the studies we reviewed. We attempted to obtain a copy of each database to verify its availability. The results are summarized in Fig. 3. We could access 81 of the 158 identified databases free of charge. We sent a minimum of two requests by e-mail via different institutions; where possible, we also reached the authors via phone. For 29 databases, we received no response; 34 databases are private or not available anymore; 7 databases are available for a fee and one database was available for joint research only. Because joint research usually includes sharing of the intellectual property rights, we did not consider these databases as publicly available.

From the descriptions of the databases, we studied the spectrum in which the images were captured (see Fig. 4). 102 databases were captured (out of 158) in the near-infrared spectrum; 35 databases were captured in the visible spectrum, and 16 databases were captured in both the visible spectrum and near-infrared spectrum. We verified the number of samples in the available databases and the spectral domain in which the images were taken (see Fig. 5). There were 1,378,867 iris images in the available databases combined, where the number of near-infrared spectrum images is substantially larger (1,257,468) than the number of images captured in the visible spectrum (138,989).

4. Databases

Early research on iris recognition was performed on datasets that were not publicly available, but were collected by each research group separately, often even for a specific study or paper. One of the first consistently reported results was obtained by John Daugman when using the UAE database (publicly unavailable). The first publicly available dataset was CASIA v. 1 [15], introduced in 2003. Since then, the CASIA database (including its updated versions) has become the most popular benchmark for the evaluation of iris recognition methods.

With the popularity of iris recognition in various use-cases, there has been a need for additional benchmarks, and therefore, new publicly available databases of human iris images. Each new database typically exploits one or more properties of iris recognition. These properties can be split into two groups: *intrinsic properties* and *extrinsic properties*.

- *Intrinsic properties* are inherent to the technology or the acquisition process. For instance, the spectrum at which the iris is captured (near infrared (NIR) or visible), the size of the iris in the image, and the sensor type (dedicated or integrated in a mobile device).
- *Extrinsic properties* are not related to technology but are typically related to the use case. For instance, the influence of aging, images captured under unconstrained conditions (influence of glasses, specular reflections, outdoor imaging, etc.), and the possibility of spoofing with an artificial iris.

In this section, we provide descriptions of the identified research databases. We identified 158 databases that are mostly declared to be publicly available for research purposes. We divided the databases into six groups: *(i)* databases collected in a controlled environment, *(ii)* synthetic databases, *(iii)* multimodal databases, *(iv)* databases exploring intrinsic properties, *(v)* databases exploring extrinsic properties, and *(vi)* uncommon research databases.

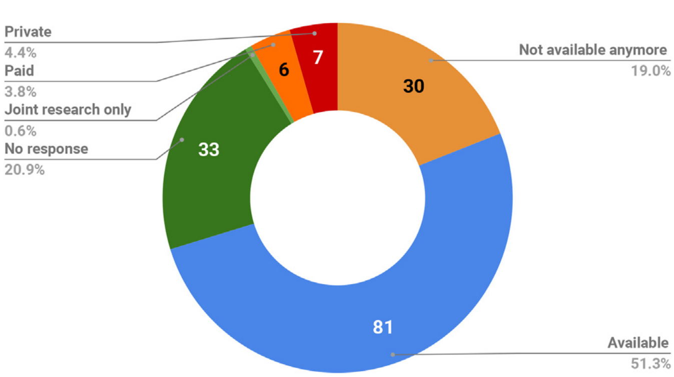

Fig. 3. Availability of iris databases.

4.1. Controlled environment

Early publicly available datasets were captured in a controlled environment where the different properties, both extrinsic and intrinsic, were kept constant. The aim of these databases is to perform fundamental research and to evaluate the actual variability of irises and the repeatability of the capture over time, where changes in the environment could negatively influence the evaluation. A list of the databases collected for this purpose is summarized in Table 2 and sample images from these databases are shown in Fig. 6. Because early research on iris recognition has focused on the development of fundamental aspects of recognition methods, the capture process and environment were constrained to iris imaging. This also includes the number of sessions, typically two, in which the iris images were collected. In addition, sessions are typically organized within a short time span and use a single sensor. Data collection organized in more than two sessions is an exception. We identified that only eight databases (out of 158) were collected in three or more sessions (see Table 1).

The first publicly available database was the CASIA Iris Database v.1 [26] collected by the National Laboratory of Pattern Recognition, Institute of Automation, CASIA. The database was captured by a custom-made NIR camera, and the authors of this database manually processed the images by replacing the pupil area (and the specular reflections) with a constant intensity value. Because the manual intervention made the problem artificially simple, it is not recommended to use this database in iris biometrics research [14] as results obtained from this database might be misleading. Since the initial release of CASIA v.1, the database was continuously updated to the current CASIA v.4 [27]. In addition, the structuring introduced in the newer version helps to explore the influence of different effects, for example, intra-class variations (CASIA-Iris-Lamp), correlations in twins (CASIA-Iris-Twins), influence of the capture distance (CASIA-Iris-Distance), cross-sensor compatibility, influence of aging, and unconstrained capture from mobile devices (CASIA-Iris-Mobile-V1.0).

CASIA-IrisV4-Interval [27] contains 2,639 images (395 classes) captured by a custom NIR camera in an indoor environment. The database is well-suited to study the detailed features of iris texture.

CASIA-IrisV4-Lamp [27] contains 16,212 images (819 classes) captured by a dedicated iris scanner (OKI Irispass-H). The images were captured in an indoor environment with lamps (visible light illumination of the rooms) both on and off.

CASIA-IrisV4-Thousand [27] contains 20,000 images (2,000 classes) captured by a dedicated iris scanner (Irisking IKEMB-100). Similar to the CASIA-IrisV4-Lamp, the images were captured in an indoor environment with lamps (visible light illumination of the rooms) both on and off. This database was the first publicly available iris database with more than 1,000 subjects.

In 2007, the IIT Delhi Iris Database (IITD-V1) [28,29] was introduced. It contains 1,120 NIR images (224 classes) captured in a constrained environment and is limited to Indian subjects.

The CUHK Iris Image Dataset [30,31] contains 254 images (36 classes) captured in the NIR spectrum. This database was among the first publicly available iris databases but is rather small in size.

4.2. Synthetic iris databases

Collecting a large - scale iris database publicly available (for research purposes) is legally and logistically a difficult task. Due to possible constraints (logistic, privacy, scale, etc.) during the collection of real biometric data, databases with synthetic samples offer a possible alternative. With the evolution of computer graphics it is possible to generate synthetic iris images that have properties similar to real iris scans. In addition, almost all factors influencing data collection (e.g., noise, eye rotation, reflections, iris structure) can be controlled better than during real data collection. The protocol is typically replaced by methods and steps with which the samples are produced. Although

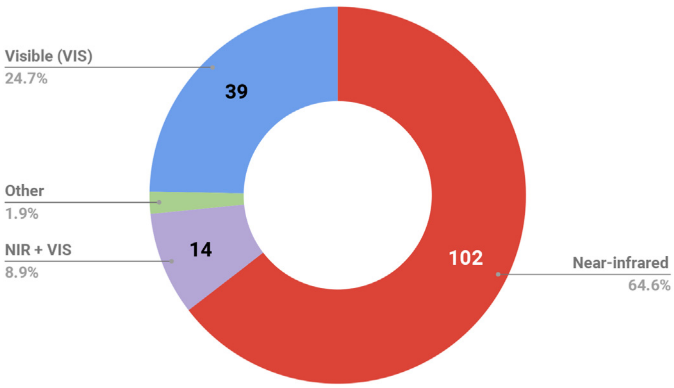

Fig. 4. Databases divided according to the spectrum in which they were captured. “Other” means non-visible and non-near-infrared spectrum, e.g., near-ultraviolet or thermal spectra.

the synthetic samples can be better controlled, the databases with real biometric data are still the ultimate benchmark. Existing datasets containing synthetic iris images are summarized in Table 3 and example images from these databases are shown in Fig. 7.

The WVU Synthetic Texture Based Iris Dataset [36] contains 7,000 synthetic iris NIR images of 1,000 classes using a texture-based approach [35]. The database was created using a computational model that generated iris textures representing the global iris appearance.

The WVU Synthetic Model Based Iris Dataset [34] contains 160,000 synthetic NIR iris images in 10,000 classes generated using a model-based and anatomy - based approach [33]. Forty controllable parameters were used (e.g., iris size, pupil size, and fiber size) to generate these images, resulting in a diverse database simulating effects like noise, rotation, blur, motion blur, low contrast, and specular reflections.

The CASIA-IrisV4-Syn [27] database contains 10,000 images in 1,000 classes created by processing the images of real irises and applying patch-based sampling to create a new prototype from which the new images were created [32]. The database contains images generated using the CASIA-IrisV1 database [26].

Although existing databases of synthetic iris images contain substantially more images than real iris image databases, their popularity remains limited. We speculate that the reasons are twofold: (i) there is uncertainty regarding the degree to which synthetic iris images share the properties and quality of real iris scans, and (ii) despite no personal information being involved, the databases are available only on requests, requiring signing a written license agreement. In addition, all synthetic databases are designed to match the properties of real iris codes when the traditional Daugman's method is used. More recently,

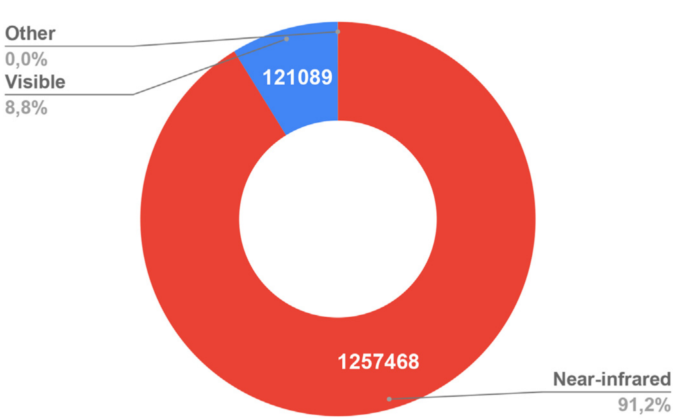

Fig. 5. Number of publicly available iris images (from the available databases reviewed) divided based on the spectrum in which they were taken.

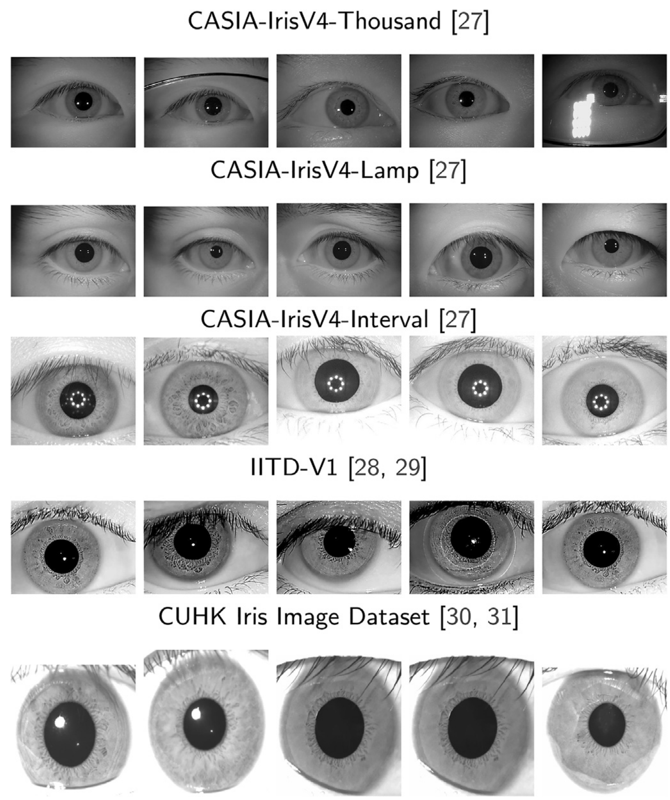

Fig. 6. Iris images from datasets captured in controlled environments.

Table 1
Available databases collected in 3 or more sessions.

Database name	Number of sessions	Year	Number of images (classes)
ND-CrossSensor-Iris 2012 [16]	27	2012	117,503 (1352)
ND-CrossSensor-Iris 2013 [17]	27	2013	116,564 (1352)
WVU Twins Day Dataset 2010–2015 [18–20]	5	2010	N/A (152+)
CASIA-IrisV1 Aging [21,22]	4	2014	36,240 (100)
ND-TimeLapse-Iris 2012 [23,24]	4	2012	6797 (46)
BiosecurID	4	2009	3200 (800)
ND-IrisTemplate-Aging 2008–2010 [25]	3	2012	11,776 (176 U 314 U 362)

Table 2
Databases of iris images captured in controlled environments.

Database name	Number of images (classes)
CASIA-IrisV4-Thousand [27]	20,000 (2,000)
CASIA-IrisV4-Lamp [27]	16,212 (819)
CASIA-IrisV4-Interval [27]	2,639 (395)
IITD-V1 [28,29]	1,120 (224)
CUHK Iris Image Dataset [30,31]	254 (36)

Table 3
Synthetic iris databases.

Database name	Generation method	Number of images (classes)
WVU Synthetic Model Based Iris Dataset [33,34]	3D model	160,000 (10,000)
CASIA-IrisV4-Syn [27,32]	texture	10,000 (1,000)
WVU Synthetic Texture Based Iris Dataset [35,36]	texture	7,000 (1,000)

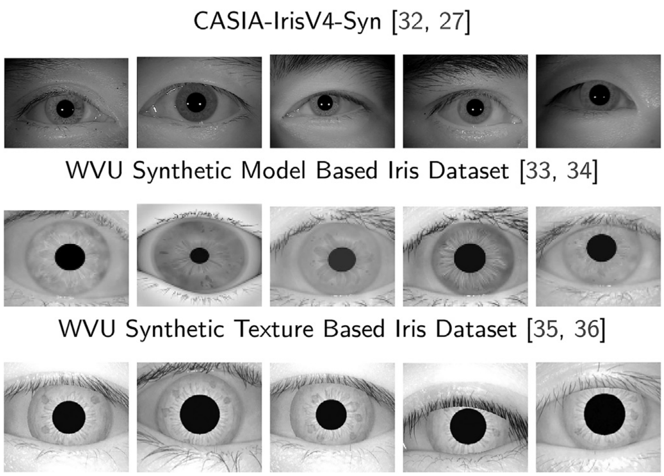

Fig. 7. Synthetic iris images.

machine-learning-based approaches (e.g., convolutional neural networks) could learn incorrect features or introduce bias into the trained model. However, recent research also includes liveness detection, where the goal is to detect artificially generated irises with a specific code to prevent potential attacks.

4.3. Multimodal databases

Using multiple biometric characteristics simultaneously in a multimodal system can significantly improve the identification accuracy [37]. Authors often opt for creating a multimodal database by combining existing datasets and assuming statistical independence of the combined characteristics. To verify and overcome this assumption, multiple multimodal databases have been collected (see Table 4 and sample images in Fig. 8).

The Center for Identification Technology Research[4] organized several biometric collections, including a collection of iris images. The Multimodal Biometric Dataset Collections (MBDC) were captured at two institutes, West Virginia University and Clarkson University, resulting in three multimodal databases that included iris modality. The irises were scanned using the OKI IRISPASS-h iris scanner and included in the three databases. The face data in these databases are not available in combination with other modalities (only as a single subset), to prevent identifying individuals and protect their privacy.

The first release of the West Virginia University MBDC database (a.k.a. BIOMDATA-V1) [38] contains 3,043 iris images of 231 subjects (462 classes). The iris modality is combined with five other modalities (face, voice, fingerprint, hand geometry, and palmprint) and additional soft-biometric information, such as height, weight, age, ethnicity, and gender. The number of sessions per subject varied across the database. The second release from the West Virginia University MBDC database (BIOMDATA-V2) [38] contains 763 iris images of 72 subjects (144 classes) captured by the same iris scanner. The main difference is that BIOMDATA-V2 also contains facial videos with voice.

The Clarkson University MBDC database contains 7629 iris images of 247 subjects (494 classes) and contains face, voice, fingerprint, and palmprint modalities. Unlike the BIOMDATA-V1 and BIOMDATA-V2 databases, it does not contain hand geometry modality, and the fingerprints were scanned using four different scanners.

The Quality-Face/Iris Research Ensemble (Q-FIRE) database [39,40], also collected at Clarkson University, is a multimodal database containing 2,800 iris images (390 classes) scanned with a dedicated iris scanner. In addition, it contains NIR iris videos (approximately 3,000 frames per subject, 586,560 iris frames combined) and facial images captured by standard color cameras. The subjects were captured at various distances, making the dataset interesting for research into iris imaging at a distance.

[4] https://citer.clarkson.edu/

Table 4
Multimodal databases containing iris modalities.

Database name	Modalities other than iris	Number of iris images (classes)
Clarkson University MBDC [38]	Face, Voice, Fingerprint, Palmprint	7,629 (494)
West Virginia University MBDC, release 1 [38]	Face, Voice, Fingerprint, Hand Geometry, Palmprint	3,043 (462)
Q-FIRE [39,40]	Face	1,800 (390) + video sequences
SDUMLA-HMT [43,44]	Face, Iris, Finger Vein, Fingerprint, Gait	1,060 (212)
Iris Cornea Dataset [41,42]	Iris texture, Corneal shape	780 (78)
West Virginia University MBDC, release 2 [38]	Face, Face video with voice, Fingerprint, Hand geometry, Palmprint	763 (144)

The Iris Cornea Dataset [41,42], created at the Université de Montréal, is a multimodal database that contains iris images (in the visible domain) combined with the three-dimensional topographical shapes of the corneas. The images were scanned with two different devices. A simple Plugable USB 2.0 Microscope webcam was used for the iris, and a Pentacam HR topographer was used for scanning the cornea. The database contains 780 iris images of 39 subjects (78 classes).

The Homologous Multi-modal Traits Database [43,44] was developed at Shandong University, in the Machine Learning and Applications Group (SDUMLA-HMT) and contains 1,060 iris images (212 classes). The database was captured in the NIR spectrum by a custom capturing device.

4.4. Intrinsic properties

Intrinsic properties cover properties within the technological setup and execution during the iris acquisition process. These properties influence the design of the recognition system and typically do not change after deployment. Active research also covers (but not only):

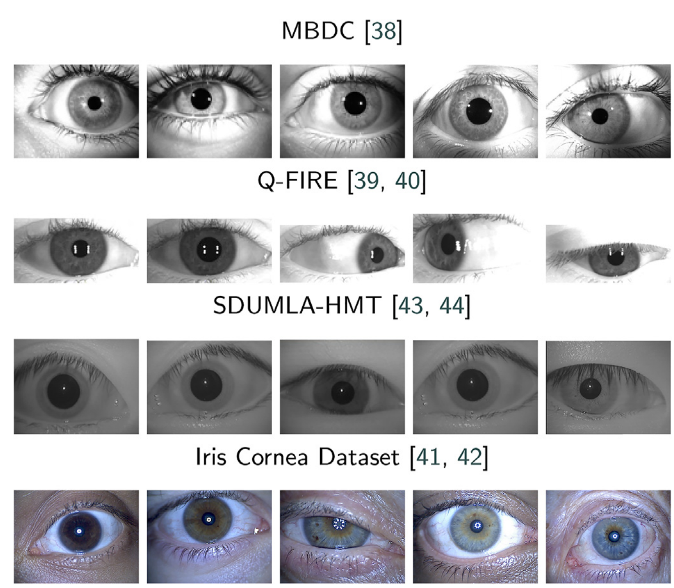

Fig. 8. Iris images from multimodal databases. The subsets of the MBDC database (from Clarkson University and West Virginia University) were collected using the same sensor and at the same event and contain similar types of images.

- spectrum in which the iris was imaged
- size of the iris in the acquired image
- cross-sensor compatibility
- image acquisition distance

Although the ideal size of the iris in an image has been well studied, many of the reported results were performed on simulated or downsampled images. The downsampling process does not result in identical images, owing to different sensor sizes and iris distances (e.g., due to noise or sensor filters) and therefore does not fully correspond to realistic environments. Daugman recommended a minimum iris radius of 70 pixels, but 100 to 140 pixels is more typical in field trials [45]. The National Institute of Standards and Technology (NIST) standard and reports follow prior research and refer to an iris radius of at least 100 pixels. While theoretical limitations can point towards a specific size of the iris, the ideal size is influenced by multiple variables, such as noise, quality of the optics (mainly diffraction, which limits spatial resolution), and sensor type and sensitivity.

4.4.1. Iris imaging at a distance

One of the most active topics in iris recognition is the problem of iris imaging at a distance. Traditional iris recognition systems achieve near-perfect *false match rates* and *false non-match rates*; however, they require collaboration of the subject, constrained illumination, and a short imaging distance. Iris imaging at a distance refers to a group of several problems related to image acquisition and their solutions, such that traditional recognition methods can be applied. Table 5 lists the available databases of iris images captured at a distance (See Fig. 9 for the sample images).

Acquiring iris images at a distance comes with the following issues:

- *Iris size* - with increasing distance between the subject and camera, larger focal lengths are used for imaging.
- *Amount of light* reflected from the iris - It is a challenging engineering problem to develop artificial lighting at larger distances that is safe and convenient, therefore many studies focus on imaging under natural lighting, frequently in the visible spectrum. Additionally, because an iris is a relatively small object compared to the imaging distance, researchers chose optical systems with large apertures to increase the amount of incoming light.
- *Focus* - optical systems with a larger aperture have a smaller depth of focus. This means that the optical system must perform a focusing operation and maintain the image in focus while imaging.
- *Motion blur* - the problem of imaging an iris at a distance usually includes non-cooperative imaging, i.e., subjects are prone to be in motion, producing motion blur in the resulting image. To overcome this issue, a mechanical system of mirrors can be used for efficient tracking of the subject.

We identified several publicly available iris databases designed to study iris imaging at a distance (Table 2).

The UBIRIS-V2 [46] database contains 11,102 images of 261 subjects (522 classes). The images in the database were captured under visible light (using a Canon EOS 5D DSLR camera) and simulated less-constrained or non-constrained environments, that is, occlusion of the iris texture while the subject was walking towards the camera from 8 to 4 m away. The UBIPr [47] database contains 10,950 images of 261 subjects (522 classes). The UBIPr database is a version of the UBIRIS-V2 database with images cropped wider to contain the ocular area for periocular recognition research. The CASIA-V4-Distance [27,48] database contains 2,567 images of 142 subjects (284 classes). The images were captured using a self-developed multimodal long-range acquisition system at a distance of 3 m (the exact distance per image is not clear).

Table 5
Databases of iris images captured at a distance.

Database name	Capturing distance (m)	Number of images (classes)
UBIRIS-V2 [46]	4–8	11,102 (522)
UBIPr [47]	4,5,6,7,8	10,950 (522)
CASIA-V4-Distance [27,48]	≥ 3	2,567 (284)

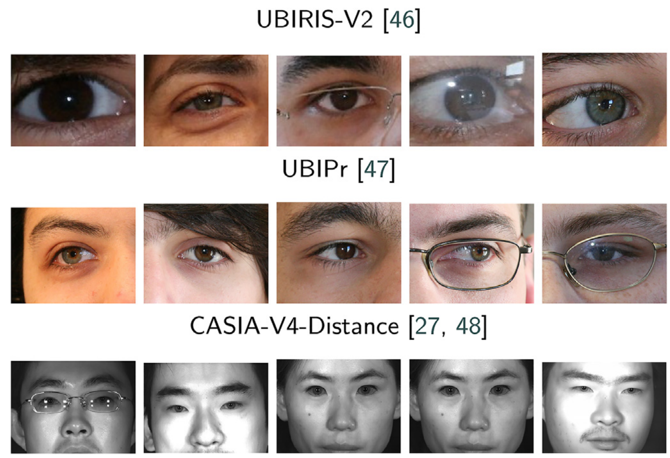

Fig. 9. Iris images captured at a distance.

4.4.2. Smartphone/Mobile databases

With the spread of smartphones, many researchers have started to work on iris recognition in the mobile environment. Because these devices already have built-in, rather high-resolution cameras, a relevant research question is whether it is possible to use it for iris recognition tasks. Using built-in cameras means using the visible spectrum instead of NIR. Recognizing the iris in the visible spectrum poses additional challenges, for instance, a wider variety of environmental conditions, wide-angle optical systems, and passive lighting. Because the use of built-in color cameras is a challenging task, some research groups have developed extension modules with an NIR camera and active lighting. With the research on iris recognition on smartphones, smartphone-captured datasets have also been introduced. Table 6 summarizes the available databases of iris images captured on mobile devices, and Fig. 10 shows sample images from these databases.

CASIA Iris M1 (mobile) [51,52] is a database containing three subsets: S1 [53], S2 [61], and S3 [53]. The database was collected with a custom-made NIR iris imaging module (S1 and S2 with an improved version of the module) and a mobile phone with an integrated NIR iris-scanning sensor (S3). All scanning devices used active lighting. The database contains 11,000 images (the three datasets combined) from 630 subjects. All three datasets contain data in JPEG format. The capture distance varied from 20 to 30 cm, although it is not clear how accurate the distance measurements were.

CASIA also publicly released a training database within the BTAS competition [54], which contains 4,500 images of 150 subjects (150 classes). They used a second version of a custom-built NIR mobile camera module to collect the images. The database includes variations in distance, non-linear deformations, eyeglasses, and specular reflections. It is not clear whether the testing part of the dataset is available as we received no response from the authors.

MICHE DB [55,56,60] is a database created within the Mobile Iris Challenge Evaluation I/II, organized at the University of Salerno. It was collected using three state-of-the-art devices[5] without additional hardware, and the images were therefore captured in the visible spectrum using a built-in - in camera application. In contrast to CASIA Iris M1

[5] iPhone 5, Samsung Galaxy 4, and Samsung Galaxy Tab

Table 6
Iris image databases created using smartphone devices[a].

Database name	Devices	Spectrum	Number of images (classes)
VISOB [49,50]	Frontal cameras of: iPhone 5 s, Samsung Note 4, Oppo N1	VIS	75,428 (1,100)
CASIA Iris M1 (mobile) [51–53]		NIR	11,000 (1,260)
subset S1	(custom) CASIA module v.1	NIR	1,400 (140)
subset S2	(custom) CASIA module v.2	NIR	6,000 (400)
subset S3	a phone with iris NIR camera	NIR	3,600 (720)
CASIA BTAS [54]	(custom) CASIA module v.2	NIR	4,500 (300)
MICHE DB [55–57]	iPhone 5, Samsung Galaxy (IV + Tablet II)	VIS	3,732 (184)
CSIP [58,59]	Xperia Arc S, iPhone 4, THL W200, Huawei Ideos X3	VIS	2,004 (100)

[a] The authors do not disclose whether the subjects in the individual subsets of the CASIA Iris M1 (mobile) database overlap. The sum of classes from the three subsets, 1260, assumes no overlap. If there is overlap, it can be fewer

and BERC Mobile-Iris, MICHE was collected in a relatively unconstrained environment. The database contains 3,732 images from 92 subjects (184 classes).

CSIP [58,59] is a database collected at Universidade da Beira Interior (UBI) and was designed to study the fusion of iris and periocular biometrics. The 2,004 images (100 classes) were captured by 4 different smartphones (10 mobile setups) to enable cross-sensor and varying acquisition scenarios. In addition to the images, an iris segmentation mask and annotations (participant ID and information on the acquisition conditions) are also provided.

The Visible Light Mobile Ocular Biometric Database (VISOB), [49,50] is focused on ocular recognition from mobile devices developed within the International Conference on Image Processing (ICIP) 2016 challenge, the Competition on Mobile Ocular Biometric Recognition (CMOBR). The database contains 75,428 images from 550 adult volunteers (1,100 classes) captured by 3 different smartphones (iPhone 5 s, Samsung Note 4, and Oppo N1). Although this database was designed for ocular recognition, the quality of many images could allow for research on unconstrained iris recognition. The main downside we identified is the downscaling of the images to a relatively small size, as they contain only cropped regions resized to 240 × 160 pixels.

Fig. 10. Iris images captured using smartphones.

Table 7
Databases of iris images captured in visible light.

Database name	Sensor type	Number of images (classes)
UBIRIS-V2 [46]	Monochrome	11,102 (522)
UBIRIS-V1 [62]	Monochrome	1,877 (241)
MILES [63,64]	RGB	832 (50)
UPOL [65,66]	RGB	384 (128)
Eye SBU [67,68]	RGB	70 (18)

4.4.3. Visible light

Despite NIR being the de-facto standard in iris imaging, much research has focused on visible light imaging. The three most popular databases in this domain are UBIRIS-V1 [62], UBIRIS-V2 [46], and UPOL [65,66] (see Table 7 and Fig. 11). These databases together serve as a benchmark and are referenced by 24.1% of all reviewed iris recognition papers (UBIRIS-V1: 14.7%, UBIRIS-V2: 8.2%, and UPOL: 3.5%). The UPOL database contains 384 high-quality iris images of 64 people captured in an unusually well-constrained environment. The UBIRIS databases contain noisy images captured in a less constrained environment, in the case of UBIRIS-V2, captured at a distance and on the move. UBIRIS-V1 contains 1,877 images of 241 subjects (241 classes), and UBIRIS-V2 contains 11,102 images of 261 subjects (522 classes).

The MILES database[6] [63,64] contains 832 high resolution images (50 classes) of various pigmentations captured using commercial Miles cameras designed for capturing iris and sclera.

The Eye SBU [67,68] database was collected at the Shahid Beheshti University. It contains 70 images (18 classes) captured by a consumer color camera under various conditions.

4.4.4. Multi-spectral imaging

Multi-spectral imaging involves capturing the iris in multiple spectral bands of light. In addition to the NIR spectrum, researchers often analyze the combination with visible light. Existing databases typically combine the NIR images with color images (captured by a consumer color camera with a Bayer filter on the sensor) containing 3 different spectral bands[7] (red, green, and blue). Databases developed for multi-spectral iris recognition research are summarized in Table 8, and sample images are shown in Fig. 12.

The PolyU Cross-spectral Iris Image Database [69,70] is a bi-spectral iris image dataset (a combination of NIR and VIS images) developed to study cross-spectral iris recognition. It contains 12,540 iris images of 209 subjects. Although the database is presented as bi-spectral, it can

[6] Images provided by Miles Research: www.milesresearch.com

[7] Despite capturing three different color bands separately, authors often present visible light images as a single spectrum. Capturing the visible light spectrum in a single band would require a monochromatic camera with a visible light filter equally distributed across all pixels.

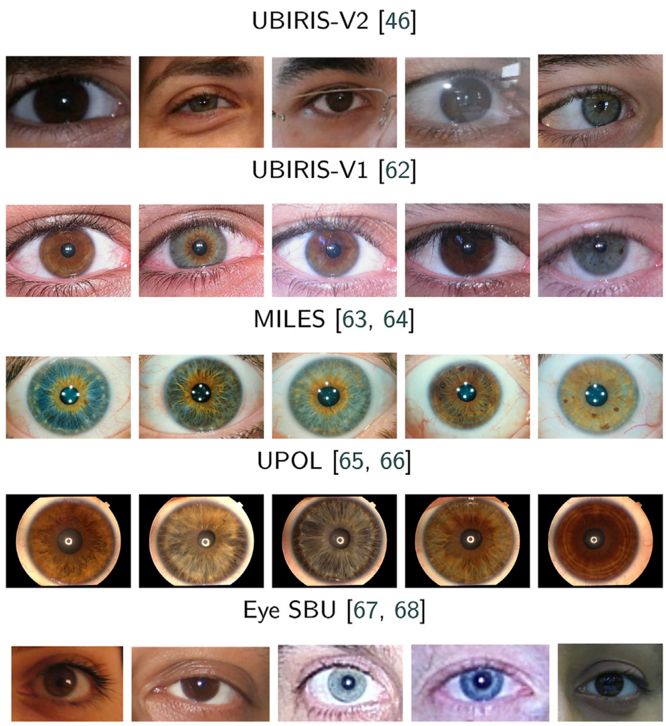

Fig. 11. Iris images from databases compiled for research into iris recognition in visible light.

be argued that there are actually four spectral bands captured, as color cameras capture three colors (red, green, and blue) separately.

The CROSS-EYED [71,72] database created for the purpose of the CROSS-EYED competition [72] contains 11,520 images (240 classes) in both visible (VIS/RGB) and NIR spectra. This database was captured by a dual-spectrum sensor acquiring the images synchronously, eliminating any possible influence of pupil dilation on the reported results. The images were captured at a distance of 1.5–2 m and contain the ocular region, making it also suitable for periocular recognition.

The IIITD Multi-spectral Periocular Database [73,77] contains 1,240 ocular images captured in three spectral bands (visible, night vision, and NIR). The database contains 5 images per eye for each of 62 subjects (124 classes irises) and for each spectrum (1,860 images in total). Images for the visible and night vision spectra contain the whole ocular area; therefore, two eyes are in each image. Most of the images are, unfortunately, too out-of-focus for iris recognition, and the spectral bands for the three spectra are not defined by the authors. For instance, it is unclear whether the images labeled as captured in night vision contain images in the NIR band or closer to the ultraviolet band, as both of them are often used in night vision systems. Because the visible color spectrum corresponds to regular color images, it can be argued that color images contain three different bands of visible spectra.

The University of Tehran IRIS (UTIRIS) database [74,75] contains 1,540 (770 per spectrum) images from 79 subjects (158 classes) captured in the visible spectrum (using a consumer RGB camera) and near-infrared spectrum. The main goal of the database is to identify features created by melanin pigment in the visible spectrum that are not available in the near-infrared spectrum [74].

Table 8
Databases of iris images captured in multiple-spectra.

Database name	Spectra	Number of images (classes)
PolyU Cross-spectral [69,70]	NIR, VIS	12,540 (418)
CROSS-EYED [71,72]	NIR, VIS	11,520 (240)
IIITD Multi-spectral Periocular [73]	NIR, VIS, Night Vision	1,860 (124)
UTIRIS [74,75]	NIR, VIS	1,540 (158)
Off Axis/Angle Iris Dataset [76]	NIR, VIS	865 (38 ∪ 146)

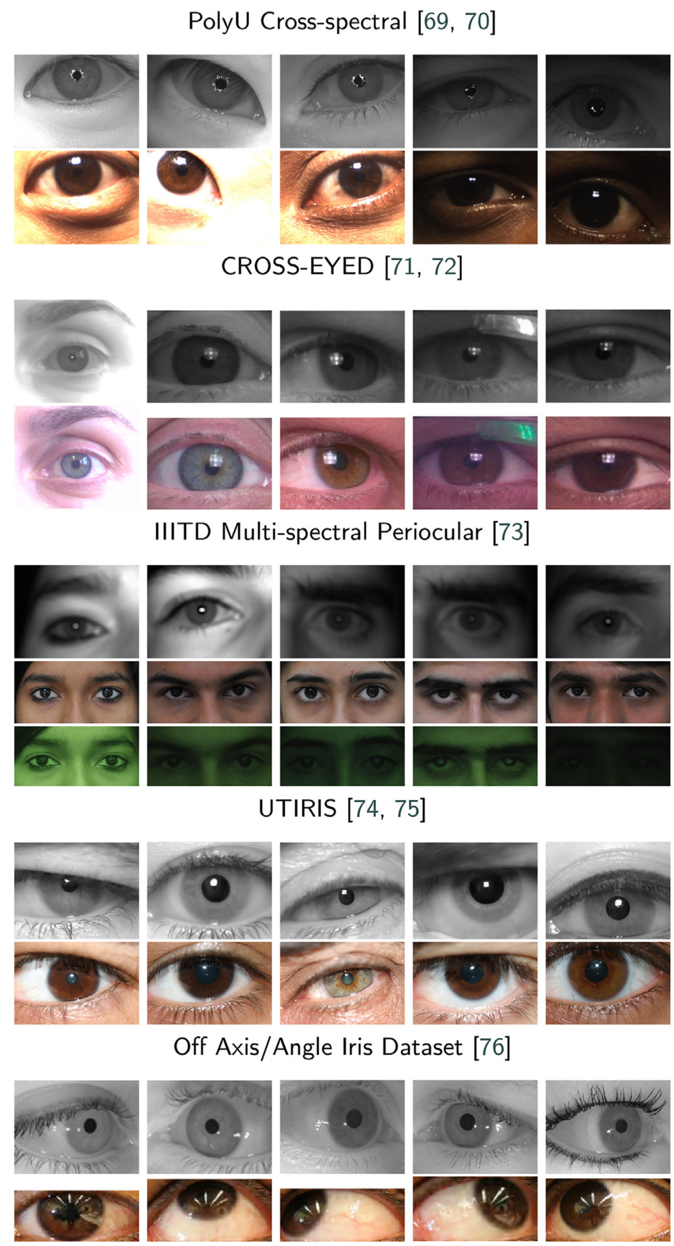

Fig. 12. Iris images captured in multiple spectra. (The license of the CROSS-EYED database does not allow us to display the images.)

The Off Axis/Angle Iris Dataset, Release 1 [76] collected at West Virginia University contains 865 images ((146 + 38 classes) captured by a color camera[8] in NIR mode and 597 images (146 classes) captured by a monochrome camera). The database contains images captured under different gaze directions (0, 15, and 30 degrees of rotation with respect to the camera).

4.4.5. Cross-sensor imaging

The wide and global deployment of iris recognition systems requires the use of various sensors, often produced by different manufacturers. Differences in sensor quality and image capturing processes raise questions about the influence of this variability on iris recognition rates. To

[8] The authors used a Sony CyberShot DSC F717 set to night-vision mode.

analyze these influences, several cross-sensor iris databases were introduced. Table 9 summarizes the existing cross-sensor databases together with sensors used for iris capture (see examples in Fig. 13). The existing databases use only a few different sensors (two or three different camera brands) for capturing the images. Therefore, for a larger evaluation, a database with more sensors could be introduced. In addition, the description of the databases does not specify whether multiple different physical devices of one brand were used (and how many) or only a single device was used. To evaluate the cause of the observed differences, it would be beneficial to describe the spectral sensitivity, amount of incoming light, and properties of the optical system used. The IIITD-WVU Mobile Iris Spoofing Dataset [78] (a subset of the LivDet Iris 2017 competition [79]) was designed as a cross-database evaluation, cross-sensor training and testing database where the acquisition environments also varied. The training part of the IIITD-WVU dataset is composed of 2,250 real and 1,000 textured contact lens iris images from the IIIT-Delhi Contact Lens Iris (CLI) database (captured by *(i)* a Cogent dual iris sensor (CIS 202) and *(ii)* a VistaFA2E), and 3,000 print attack images are selected from the IIITD Iris Spoofing (IIS) database. The testing part of the IIITD-WVU dataset is a multi-session iris spoofing dataset containing 4,209 iris images (702 classes).

The ND-CrossSensor-Iris-2012 (BTAS-2012) database [16] contains 29,939 images from an LG4000 sensor and 117,503 images from an LG2200 sensor, where every subject occurs at least twice (1,352 classes in total). Images captured with the LG2200 sensor are also available in a post-processed form, given the non-unit pixel aspect ratio (i.e., non-square pixels) of the sensor.

The ND-CrossSensor-Iris-2013 (BTAS-2013) database [17,80] contains 29,939 images from an LG4000 sensor and 117,503 images from an LG2200 sensor (1,352 classes in total). This dataset is very similar to the ND-CrossSensor-Iris-2012 and contains post-processed images (due to non-square pixels) from the LG2200 sensor.

Within the ICB Competition on Cross-sensor Iris Recognition 2015 (CSIR 2015) [52,81], another publicly available dataset was introduced. However, only the training dataset containing 8,000 images was introduced. It contains iris images of 200 eyes from 100 subjects, 4,000 images captured by an IKEMB-220 sensor, and 4,000 images captured by an EyeGuard AD100 sensor.

The CSIP database [58,59] described in Section 4.4.2 is a database captured by four mass-market smartphone devices.

4.5. Extrinsic properties

Extrinsic properties are properties related to the use case and environment that are not technology related. These properties increase the variability of the iris features, and this variability cannot be, or can be only with difficulty, technologically reduced. Active research also covers (but not only):

- influence of aging
- influence of occlusions or contact lenses
- spoofing/liveness detection
- genetic relationship

While some of the databases have been created as a benchmark to overcome a certain problem (e.g., influence of aging), other databases serve to verify the quality of existing methods. For instance, although some iris features are provably genetically related (e.g., iris color), the features suggested in Daugman's method [45] have shown statistical independence, even in genetically related irises. Methods such as mapping or learning the variability presented in the data are typically used to overcome extrinsic factors. Most of the methods are therefore based on machine learning rather than being hand-crafted.

Table 9
Cross-sensor iris databases.

Database name	Capturing sensor	Number of images (classes)
ND-CrossSensor-Iris-2012 (BTAS-2012) [16]	LG 2200 EOU, LG iCam 4000	147,442 (1,352)
ND-CrossSensor-Iris-2013 (BTAS-2013) [17,80]	LG 2200 EOU, LG iCam 4000	146,550 (1,352)
IIITD-WVU Mobile Iris Spoofing [78,79]	Cogent dual iris sensor (CIS 202), VistaFA2E, IriShield MK2120U	10,459 (702)
CB CSIR 2015 (training dataset)[a] [81]	IKEMB-220, EyeGuard AD100	8,000 (200)
CSIP [58,59]	XPERIA ARC S, iPhone 4, THL W200, HUAWEI U8510[b]	2,004 (100)

[a] The provided information is for the publicly available training dataset of the ICB Competition on Cross-sensor Iris Recognition. The testing dataset containing an additional 24,000 images is not publicly available.
[b] The Xperia Arc S uses the Exmor R CMOS sensor; other devices do not have the sensor specified by the manufacturers

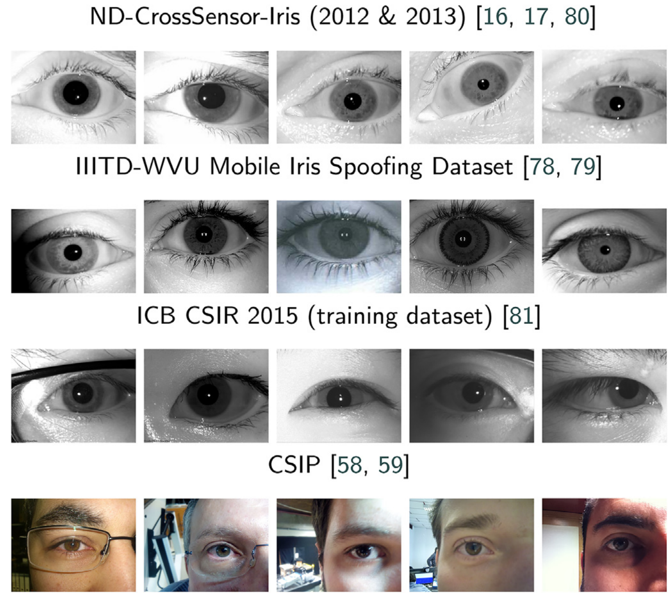

Fig. 13. Iris images captured by multiple sensors.

4.5.1. Contact lenses

Contact lenses present a challenge to iris recognition as they deform or interfere with iris patterns. In general, there are three types of contact lenses: *(i)* transparent (soft), *(ii)* color cosmetic (textured), and *(iii)* functional. The available databases of iris images captured with different contact lenses are summarized in Table 10, and examples are shown in Fig. 14.

The largest available database of irises captured with contact lenses, published by the University of Notre Dame, is the ND Iris Contact Lenses 2010 database [82]. It contains 12,003 images of 87 subjects (176 classes) wearing non-cosmetic transparent prescription contact lenses and 9,697 images of 124 subjects (248 classes) wearing no contact lenses. The images were captured with dedicated iris recognition NIR sensors.

Table 10
Databases of images of irises with contact lenses.

Database name	Number of images (classes)
ND Contact Lens 2010 [82]	21,700 (248 ∪ 176)
ND Contact Lens 2015 [83]	7,300 (N/A)
IIITD Contact Lens [84]	6,570 (202)
ND Contact Lens 2013 [85]	5,100 (N/A)

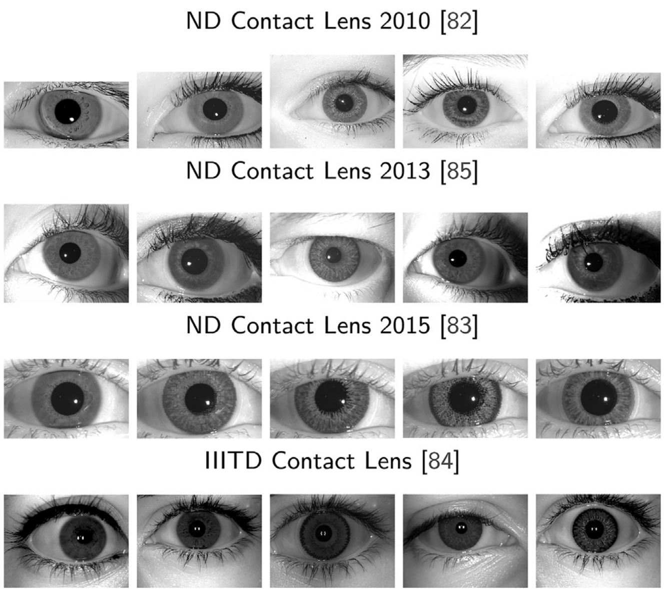

Fig. 14. Iris images capturing different types of contact lenses.

The IIITD Contact Lens Iris Database [84] also contains iris images with transparent (soft) contact lenses as well as textured lenses in four colors. There are 6,570 images of 101 subjects (202 classes). The database was also collected using dedicated iris sensors operating in the NIR.

Another database from the University of Notre Dame is the ND Contact Lens Detection 2013 Database [85]. It contains 5,100 images divided into two datasets, where each is further divided into a training set and a verification set.

The most recent database from the University of Notre Dame is the ND Contact Lens Dataset 2015 [83]. This database of 7,300 images was constructed to evaluate contact lens detection under various experimental scenarios.

There are other datasets that contain images from previous databases. The Notre Dame LivDet-Iris 2017 Subset was created for the LivDet-Iris 2017 competition.[9] In 2016, the Iris Combined Spoofing Database [86,87] was introduced, which is a combination of multiple publicly available spoofing databases, namely the IIIT-Delhi CLI Database, the IIS Database, and the Multi-sensor Iris Database.

4.5.2. Iris spoofing and liveness detection

Iris spoofing is a mechanism by which one can obfuscate or impersonate the identity of an individual [88]. With the development of spoofing and anti-spoofing methods, the need for research and benchmark databases arose (see Table 11 and Fig. 15).

The IIITD Iris Spoofing (IIS) database [88] contains 4,848 images of 101 subjects (202 classes). The authors used the IIIT-Delhi CLI database and used a Cogent CIS 202 dual eye iris scanner and HP flatbed optical scanner to create print attack scenarios. Currently, it is not available separately, only as a part of the Iris Combined Spoofing Database [89].

The IIITD Combined Spoofing database [86,87] consists of images from multiple publicly available spoofing databases (including the IIS). It combines 9,325 authentic iris images (of 1,872 subjects, 3,744 classes) with 11,368 spoofed images from real-world scenarios of a variety of iris spoofing attacks.

The LivDet (Liveness Detection Competitions) series of competitions also have a large impact on the research into iris spoofing and liveness detection. The competitions typically provide a composite database consisting of multiple images from the databases mentioned above that is available to participants. There have been 5 editions focused on iris recognition, in 2009, 2011, 2013 [90], 2015, [91] and 2017 [79]. Authors using composite databases typically refer to them as competition databases, but we want to highlight that central access is not being maintained. This means that the actual databases for the past editions of the competitions must be obtained from the original authors and composed by the receiver.[10]

The aforementioned IIITD-WVU Mobile Iris Spoofing Dataset [78,79] (the training subset) is a composite database containing textured contact lenses, print attack images (printouts of the textured contact lenses), and real iris images (without contact lenses). The testing subset is a separately created dataset that contains iris images captured using the IriShield MK2120U mobile iris sensor. It contains 4,209 iris images acquired with textured or patterned contact lenses and without any lenses (real iris images).

The ATVS-FIr database [92,93], created at the Universidad Autonoma de Madrid, contains 800 images from 50 genuine subjects (100 classes). In addition, it contains 800 fake iris images from 50 fake identities (100 classes). The fake images were captured from high-quality printed images created through the process described in [94].

The Sclera Liveness Dataset [95,96] contains 500 images (50 classes) of genuine eyes captured in the visible spectrum by a mobile device. In addition to genuine eyes, the database contains 400 spoofing images acquired from a digital screen and 100 print attack images.

The *Eye Tracker Print-Attack Database v. 1* (ETPADv1) [97,98] created at Texas State University contains 200 iris images (left eye only) collected from 100 genuine subjects (100 classes). The database also contains 600 eye movement recordings, 200 recordings from genuine subjects, and 2 × 400 recordings captured during a print-attack. The *Eye Tracker Print-Attack Database v. 2* (ETPADv2) [97,99] is the successor to the ETPADv1 database and contains 400 iris images (also the left eye only) collected from 200 genuine subjects (200 classes). The database contains 400 eye movement recordings from genuine subjects and 2 × 400 recordings from 2 different types of print-attacks. Both ETPADv1 and ETPADv2 were created to use gaze estimation for iris spoofing/liveness detection.

4.5.3. Standard noise factors

After iris recognition was demonstrated to work reliably in controlled environments, new more challenging databases were introduced (see Table 12 and Fig. 16). The nature of the challenge came from extrinsic properties, for example, occlusion of the iris by eyelashes, presence of glasses, or off-angle images.

The Multimedia University Iris Database - version 1 (MMU v.1) [105,110] contains 450 images from 45 individuals (90 classes) captured by a dedicated iris scanning sensor. The second version of the database, the Multimedia University Iris Database - version 2 (MMU v.2) [105,106], introduced one year later, contains 995 images from 100 individuals (200 classes). Both versions of the database were unique in introducing eyelash obstruction and eye rotation iris images.

Another more extensive database was released within the Iris Challenge Evaluation (ICE) 2005 [103] by NIST. The ICE 2005 database contained 2,953 images from 132 subjects (264 classes), and was created for large-scale, open, independent technology evaluation. The goal of the evaluation was to promote the development and advancement of iris recognition technology, and therefore it contained challenging images, mainly due to poor focus and occlusions by eyelashes. The second edition of the evaluation, ICE 2006 [102], contained substantially more data, 59,558 images from 240 subjects (480 classes). Currently, the databases from both evaluations are only available as subsets of the larger ND-IRIS-0405 database [100,101] which contains 64,980 iris images (712 classes).

[9] http://iris2017.livdet.org

[10] See the supplementary materials for more details

Table 11
Iris databases for liveness detection.

Database name	Genuine irises	Number of images (classes) Textured contact lenses	Fake printouts
LivDet-Iris-2013 [90]			
Notre Dame subset	2,800 (N/A)	1,400 (N/A)	–
Clarkson subset	516 (64)	840 (6)	–
Warsaw subset	852 (284)	–	815 (276)
LivDet-Iris-2015 [91]			
Clarkson LG subset	828 (45)	1,152 (7)	1,746 (45)
Clarkson Dalsa subset	1,078 (N/A)	1,431 (N/A)	1,746 (N/A)
Warsaw subset	2,854 (100)	–	4,705 (100)
LivDet-Iris-2017 [79]			
Notre Dame subset	1,500 (N/A)	1,500 (N/A)	–
IIITD-WVU subset	2,952 (N/A)	1,701 (N/A)	5,806 (N/A)
Clarkson subset	3954 (50)	1887 (12)	2,254 (49)
Warsaw subset	5,168 (470)	–	6,845 (470)
ATVS-FIr [92,93]	800 (100)	–	800 (100)
Sclera Liveness Dataset [95,96]	500 (50)	–	500 (50)
ETPADv1 [97,98]	200 (100)	–	2 × 400 recordings (100)
ETPADv2 [97,99]	400 (200)	–	2 × 400 recordings (200)

The Jilin University Iris Biometric and Information Security Lab (JLUBRIRIS V1-V6) Database [107–109] contains 729 video recordings (360 classes) captured under the NIR spectrum using a self-developed camera. The database contains five-second video recordings that are stored as individual images.

The Multi-Angle Sclera Dataset (MASD) v. 1 [95,104] contains 2,624 images (164 classes) of eyes captured in the visible spectrum for the purpose of processing the sclera region of the eye; however, images also have sufficient resolution for iris recognition.

Another database containing challenging, off-angle, iris images is the Off Axis/Angle Iris Dataset, Release 1 [76]. This database contains images captured in the visible spectrum and in night vision (NIR-enabled) mode, as discussed in Section 4.4.4.

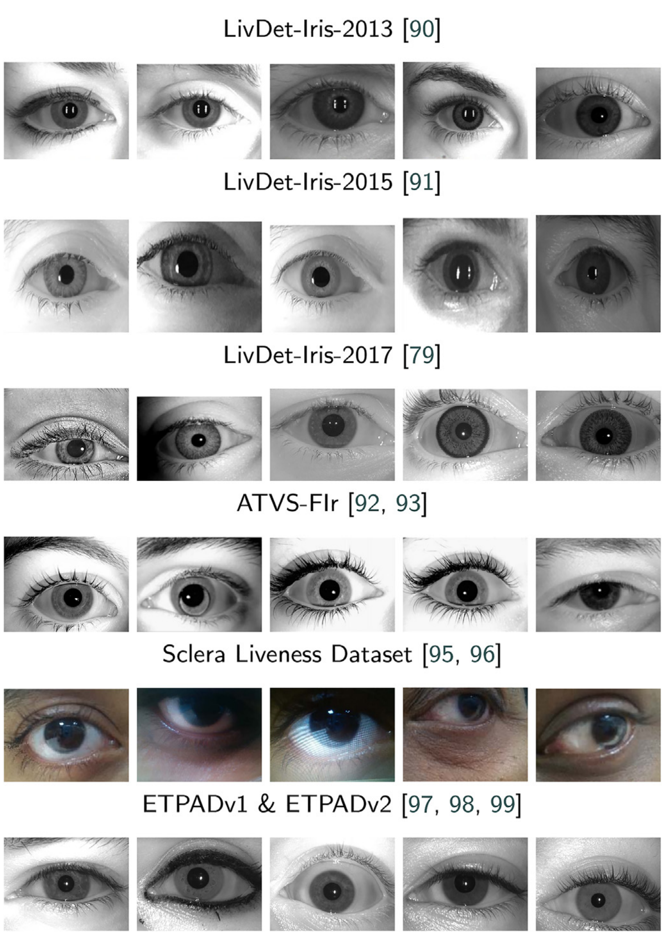

Fig. 15. Iris images from databases created for spoofing and liveness detection.

4.5.4. Periocular databases

Several datasets have been created for periocular recognition research that might find applications in iris recognition research (see Table 13 and Fig. 17). The two most popular are the Multiple Biometrics Grand Challenge (MBGC) [111,112] and the Face and Ocular Challenge Series (FOCS) [113,114] datasets.

The databases developed for MBGC (two versions v. 1 and v. 2) were designed to investigate the potential of the fusion of face and iris biometrics as well as off-angle iris images from video. The database contains 14 images and 3 video sequences for 140 different subjects (280 classes) captured in the NIR spectrum.

The database developed for FOCS [113,114] consists of 9,588 still images (136 subjects, 272 classes) of a single iris and eye region. The still images were captured in the NIR spectrum. The eye regions were extracted from NIR video sequences collected from the Iris on the Move system used in MBGC v. 2. Therefore, there is an overlap between the FOCS and MBGC databases.

The IIITD Cataract Mobile Periocular Database [117,118] contains 2,380 ocular images (145 classes), captured before and after cataract surgery. Despite the large resolution of the images (4608 × 3456 pixels), the quality of the images is rather low, as they were captured by a regular smartphone camera (MicroMax A350 Canvas Knight). Therefore, its suitability for iris recognition research is questionable.

The UBI Periocular Dataset [115,116] is a periocular image database captured under non-controlled acquisition conditions, containing more variations in pose, illumination, and scale than MBGC and FRGC. The database contains 10,950 images (522 classes) collected at various distances from 4 to 8 m.

Another periocular dataset originating from UBI is the CSIP database [58,59] described in Section 4.4.2 captured by state-of-the-art smartphones (2014). The CSIP database contains 2004 images of 50 people (100 classes) captured by four different smartphones. The images vary in resolution and are also provided with iris segmentation masks and information on the acquisition conditions.

4.5.5. Genetic relationship

It has been demonstrated that the genetic relationship between two irises has no influence on the random pattern extracted as proposed by

Table 12
Iris databases containing different types of noise.

Database name	Spectrum	Number of images (classes)
ND-IRIS-0405 [100, 101]	NIR	64,980 (712)
ICE 2006 [102]	NIR	59,558 (480)
ICE 2005 [103]	NIR	2,953 (264)
MASD v.1 [104, 95]	VIS	2,624 (164)
MMU v.2 [105, 106]	NIR	995 (200)
WVU Off Axis/Angle [76]	NIR, VIS	865 (38 ∪ 146)
JLUBRIRIS V1-V6 [107–109]	NIR	729 videos (360)
MMU v.1 [105, 110]	NIR	450 (90)

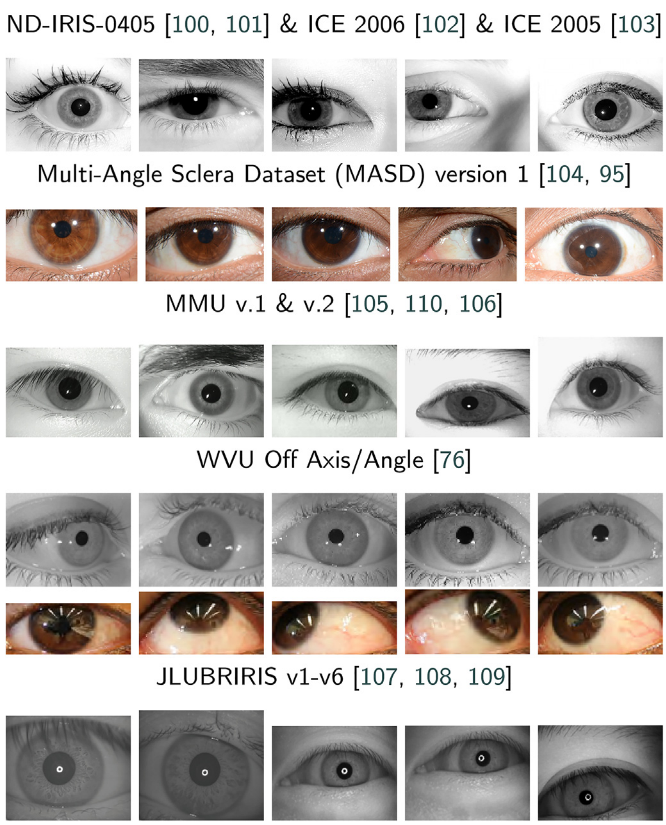

Fig. 16. Iris images captured with different types of noise.

J. Daugman [45]; however, other features might exhibit correlation [119], allowing the extraction of additional information from iris images (e.g., gender and ethnicity).

Analyzing the relationship between two genetically equal irises is an easy task, as we can study the similarity between the left and right eyes of the same person. In most existing databases, both eyes are labeled. To study genetic relationships, iris images of identical twins were collected.

Table 13
Periocular databases.

Database name	Spectrum	Number of images (classes)
UBI Periocular [115,116]	VIS	10,950 (522)
FOCS [113,114]	NIR	9,588 (272)
IIITD Cataract Mobile [117,118]	VIS	2,380 (145)
CSIP [58,59]	VIS	2,004 (100)
MBGC v.1 & v.2 [111, 112]	NIR	1,960 (280) +420 *videos*

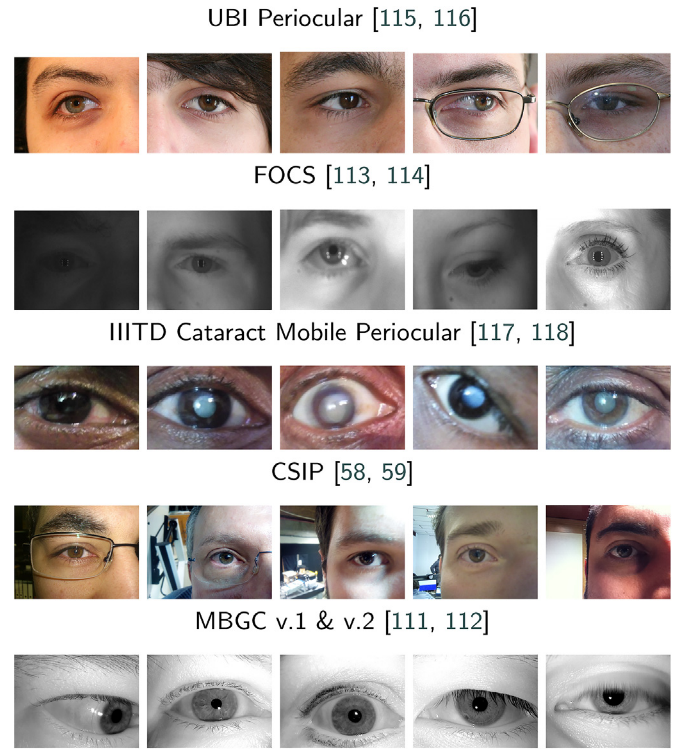

Fig. 17. Iris image databases used for periocular recognition research.

Table 14
Iris image databases of identical twins.

Database name	Number of images (classes)
CASIA Iris v.4 Twins [27]	3,183 (400)
WVU Twins Day Dataset [18–20]	N/A (152+)

We identified two publicly available twin-iris databases (see Table 14 and Fig. 18).

The first publicly available iris database of twins was CASIA Iris v.3 Twins [27] (now available in v. 4). It contains 100 pairs of twins captured using OKI's IRISPASS-h camera. In total, it contains 3,183 images in 400 classes from 200 people. Most of the subjects were children, and the images were captured in outdoor conditions.

The WVU Twins Day Dataset 2010–1015[11] [18–20] is a multimodal database containing fingerprints, visible-spectrum facial images, and NIR-spectrum images of 152 irises. The images were captured in five sessions in semi-outdoor conditions (in an outdoor tent). The WVU Twins Day Dataset 2010–1015 was captured during the same event as the ND-TWINS-2009-2010 database [120]. However, although the authors declare that irises were captured during data collection [19], the ND-TWINS-2009-2010 contains only still face images.

4.5.6. Aging

The structure of the human iris has been presented as one of the most stable biometric characteristics that does not change significantly over a lifetime [24]. Due to a lack of data collected over larger time spans, it is difficult to verify this assumption. Some researchers observed that for longer time spans, the distribution of genuine iris features shifts closer to the distribution of impostor iris features; thus, the recognition

[11] The WVU Twins Day Dataset 2010–1015 is not available to non-US nationals or researchers outside the USA.

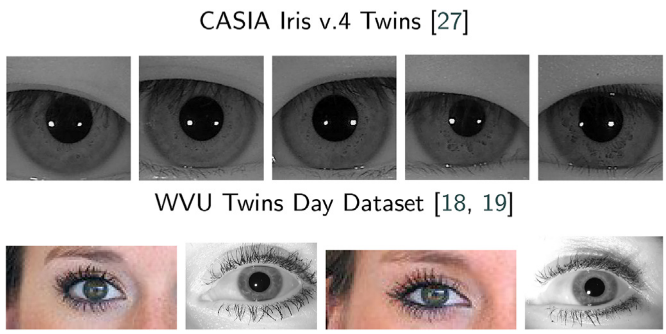

Fig. 18. Iris images of twins.

performance decreases [121]. To quantify this shift, several databases have been introduced (see Table 15 and Fig. 19).

The ND-IrisTemplate-Aging-2008-2010 [25] dataset contains 11,776 images of fewer than 200 people[12] collected in three sessions over two years.

The CASIA-IrisV1-Aging [21,22] database contains 36,240 images of 50 or fewer subjects[13] collected in 5 sessions within 4 years.

Table 15
Iris image databases collected to study the effects of aging.

Database name	Number of images (classes)
CASIA-IrisV1-Aging [21,22]	36,240 (100)
ND-IrisTemplate-Aging-2008-2010 [25]	11,776 (362)

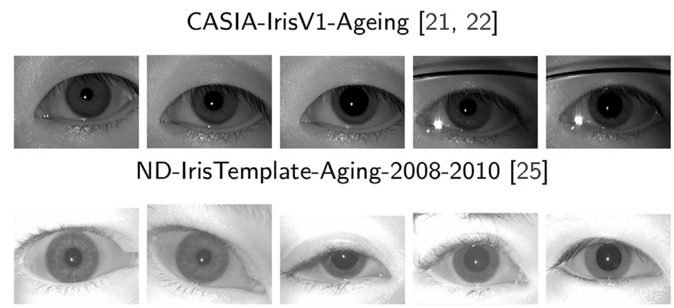

Fig. 19. Iris images from databases compiled to study the impact of aging.

4.5.7. Other properties

Another interesting database is the Children Multimodal Biometric Database [122], which contains 2,660 images of 3 modalities (iris, fingerprints, and face) belonging to 100 children aged from 18 months to 4 years.

ND-Gender-From-Iris-Dataset (ND-GFI) [123] is a database combined with the *gender information* about the subjects. The core part of the database contains 3,000 images, one image per class for 750 males and 750 females. An additional part of the database contains three images per class.

4.6. Uncommon research databases

In view of the wide and challenging field of iris recognition, it is understandable that research has been spread in various and uncommon directions.

The influence of different chemical substances on the iris (e.g., pupil dilation) is well known. However, the lack of available data led experts to different conflicting conclusions [124,125]. The IIITD Iris under Alcohol Influence (IUAI) database[14] [126] was created to help solve this problem. It contains 440 NIR iris images of 55 subjects captured in a semi-controlled environment before and after alcohol consumption.

Similarly, the Iris Degradations Data Set (IDDS) [127] contains 2,183 iris images of drug-induced pupil constriction and dilation to investigate their impact on bit errors. In both cases, volunteers applied pilocarpine and tropicamide solutions in the form of eye drops, which caused miosis and mydriasis, respectively. This resulted in an increase in the bit error in both cases compared to the control state, as the original information was distorted either by the addition of previously unavailable information (miosis) or by the removal of previously existent information (mydriasis). All images were captured in the NIR domain, with the database containing images of both eyes.

The Eye Cancer Foundation Dataset[15] [128,129] (see Fig. 20) contains images of eyes with various forms of severe diseases (e.g., iris tumors, choroidal tumors, conjunctival tumors, and infiltrative intraocular tumors). The database contains both malignant and benign patient samples captured in an unconstrained environment. Although the database covers a wide range of diseases, it contains only one image per patient, which is a considerable drawback.

Fig. 20. Images from the Eye Cancer Foundation Dataset.

The IIITD Cataract Mobile Periocular database [118] contains pre- and post-cataract surgery images. It is captured in the visible spectrum using a high-resolution mobile camera, and contains 56 common patients across both sessions (pre - and post-surgery).

Another database capturing various eye diseases is the Warsaw-BioBase-Disease-Iris v1.0 - v2.1[16] [130,131]. The database contains NIR and partial color images of healthy eyes as well as eyes with various pathologies (including cataract, acute glaucoma, posterior and anterior synechiae, retinal detachment, rubeosis iridis, corneal vascularization, corneal grafting, iris damage, atrophy, corneal ulcers, haze, and opacities). As the patient usually suffers from multiple diseases, the authors decided to divide images according to the type of physical impact that affects the eyes or other structures as an external symptom of the specific disease. The authors' results show that diseases affecting iris geometry or tissue structure or producing other visible manifestations significantly affect recognition accuracy, but also a disease that does not produce distinctive changes can increase the likelihood of an error.

Another uncommon area of research is iris recognition for animal identification in general [132]. The SEU Bovine Iris Database [133] is the only available database of cow iris images that we identified. It was captured in the NIR domain. Although the results suggest that the methodology for identifying cattle based on their irises is appropriate, the authors point out that further research is needed in this area. Another animal iris database is the Warsaw-BioBase-HorseIris v1.0 [134], which contains near-infrared images of horse eyes captured by a specialized veterinary camera. It includes images captured during two sessions at 60 frames per second to compensate for animal movement

[12] The number s of subjects in the individual sessions were as follows: 88, 157, and 181.
[13] The number s of subjects in the individual sessions were as follows: 48, 49, 50, 49, and 49.
[14] Despite claims in the paper [126], the database is not yet available on the declared website: http://research.iiitd.edu.in/groups/iab/resources.html
[15] Images in this publication were provided by The Eye Cancer Foundation, Inc. http://eyecancer.com
[16] Despite multiple attempts, we did not succeed in obtaining any database from http://zbum.ia.pw.edu.pl/EN/node/46 for legal reasons. However, if available, datasets from the Warsaw University of Technology would be a valuable contribution to the research community. For instance, post-mortem datasets and pupil dynamics datasets.

during the shooting. In total, it contains data from 28 horses (14 mares, 10 stallions, and 4 geldings).

5. Common properties of popular databases

After identifying the iris databases with the highest impact, we identified four common properties of the most cited databases[17]:

1. Longevity - Our findings confirm that the most cited databases have been available for the longest time. A good example is the CASIA-Iris-V1 database, which is the most cited database and has been available continuously since its introduction in 2003. Despite a rather small scale and recommendations not to use it, it is still being cited and often serves as a benchmark dataset. Because there are already many published research papers, this dataset has become a tool for comparing new results with a wide range of past results. In contrast to CASIA-Iris-V1, the extensive BATH database [135] reached only limited popularity mainly due to its relatively short lifespan.
2. Limited access restrictions - The datasets that are publicly available without requiring researchers to sign a license are more popular (e.g., the UPOL iris dataset [65,66]). In addition, licenses for many datasets require a signature of the institutional legal representative (e.g., datasets from Notre Dame University [136]. These datasets are less popular than datasets where the signature of an individual researcher is sufficient. (e.g., the CASIA Iris dataset). A license is legally binding a document and typically protects the subjects as well as the institution providing the database. In addition because iris images are a type of personal data, this protection is in many countries regulated and enforced by law (e.g., the EU General Data Protection Regulation [137]).
3. Scale - The numbers of subjects and images in a dataset also influence the popularity. A sufficient number of samples in a dataset is a significant requirement for performing statistically relevant research. Datasets with more samples (i.e., with higher statistical relevance) can often serve as objective benchmarks. While the size of the dataset is a strong requirement, it is often not sufficient, and other properties, such as a good protocol, must be met as well.
4. Timing - We found that the most popular datasets introduced novel aspects and enabled research that was not possible with the public datasets available prior to their publication. Good examples are *(i)* the CASIA-Iris-V1 dataset–the first publicly available iris dataset, *(ii)* the UBIRIS v.1 dataset–the first database to introduce unconstrained imaging, and *(iii)* the UPOL [65] database–the first that contains high-resolution images taken in the visible spectrum.

6. Discussion

Existing iris datasets have matured over the last two decades, but there are still limitations that need to be addressed. The major limitation is their availability, meaning that many of the datasets have a rather short lifespan during which they are available. A good example is the BATH iris dataset [135], which has been used by many studies; however, it is no longer available. In some cases, we also observed selective availability, where the authors decided to share the dataset depending on the requesting institution (institutions with a lower profile might typically have more problems obtaining a dataset). This undermines the reproducibility of the research by independent researchers and prevents new researchers from publishing results obtained using such a database.

The effect of aging has been studied on public datasets created in a time span of at most eight years [138]. The limits of long - term iris recognition are defined by the difficulties in following up on a large group of people over a long period [139].

A relatively new problem arises from the new data privacy regulations that protect the privacy of subjects. For instance, in Europe, the GDPR regulation contains the right to erasure (a.k.a. *the right to be forgotten*) [140] that guarantees to the subjects the possibility to retract their agreement for the use of their data and removal of the subject-associated information from the database (if possible). Due to the nature of the biometric data, it is possible to uniquely identify the subject (in the case of iris data, even with a high precision) [141]. Thus, possible changes in the datasets might undermine the purpose and consistency of the reported data over years. Due to recent issues related to the leaking of personal data, similar regulations are being discussed globally.

Other substantial limitations of the available databases are imperfections in the described capture setup and protocol. Many of the parameters are considered irrelevant by the authors for their own research, but that information might be vital for others and can therefore broaden the potential application areas of the database. The properties of optical systems are rarely described. Many databases also lack descriptions of at least one of the following: sensor type and model, spectral range of captured images, capture distance, and ecological validity. Because many datasets provide only cropped regions of the eye, information about aperture, shutter speed, and sensitivity are missing, although most of the capture systems include this information in the images by default. In the case of capturing with mobile devices (e.g., smartphones), data from the inertial measurement unit (IMU) (i.e., an accelerometer and a gyroscope) might provide valuable information to remove the negative effects of the rolling shutter and identify motion blur [142]. In addition, many datasets provide images in a compressed format only, thus reducing the information captured by the sensor.

A detailed description of the protocol and the capture setup is important because of differences in imaging requirements for different directions of research. For instance, iris capture in motion shares the core problems that are recognized in iris capture by mobile handheld devices - the relative movement of subjects with respect to the sensor and a substantially smaller sensor compared to the distance between the sensor and the eye. Despite similarities in research problems, research on the iris in motion focuses on using novel sensors and optical systems [2,143], while mobile iris capture focuses on using additional sensor information available in mobile devices (IMU or multiple imaging sensors) [144,145] and computational methods to process the captured images [146,147].

Many research papers point out the absence of databases suitable for testing a particular parameter (i.e., a constrained environment with variability of only a single parameter), which makes research conclusions and underlying reasons rather unclear, typically stating that more research is needed. In these cases, having a detailed description of the protocol could solve the problem. Some of these issues could be avoided by keeping the EXIF/metadata (their removal is a standard practice for reasons of privacy) that includes camera parameters. Databases created using custom-built cameras typically do not provide this information and lack any description of the protocol. While publicly available cameras include such specifications in the image file by default, in the case of custom hardware, many aspects remain hidden to other users of the database.

We also observed an inconsistency among datasets that were captured under visible light. In some cases, the authors used a monochromatic sensor with a band-pass filter that captures the entire visible band of light; others use mass market cameras that capture the visible light in three separate spectral bands (separately for the colors red, green, and blue). The spectral sensitivity of the visible light filter differs from that of the individual color filters (even when the color bands are combined) and therefore should not be treated as analogous. In addition, most consumer color cameras contain a Bayer filter that limits the resolution of individual bands to one quarter for red and blue spectra and one half in the case of green; thus, 2/3 of the color information is actually computed and not measured.

From our review, we also conclude that datasets with synthetic images have not gained popularity in iris recognition research. Despite their containing large numbers of samples (larger than traditional datasets), researchers opt for real-world images. We suspect that these datasets lack the realism of studied effects that appear in less constrained environments.

[17] Sorted according to importance, where the first is the most important

Much of the research on iris capture at a distance has focused on building a traditional optical system with mirrors for the capture, but only a limited amount is focused on computational iris capture, e.g., using super - resolution [148].

6.1. Related research

We have presented a review of existing databases that were created for the evaluation of biometric methods, more specifically, iris and ocular recognition. There are other research areas that use images of the human eye and iris that could benefit from the databases reviewed for this study.

Gaze estimation and eye tracking have been used in many applications from human–computer interaction [149] to healthcare applications as a diagnostic tool. These applications typically have other well-established datasets (e.g., [150]), but share mostly initial processing of iris images, i.e., iris localization and segmentation. Hence, iris recognition databases could be of use as an additional data source. In addition, pupil tracking (capturing changes in the size of the pupil over a short time span, typically a few seconds), has been shown to be an important medical biomarker reflecting levels of neurotransmitter and neuronal activity [151]. Another related research area is eye blink detection. It has been used as a countermeasure against spoofing in face [152] and iris recognition methods [153] and in medical applications, e.g., to diagnose computer vision syndrome [154].

Databases traditionally used for related research have limited use in biometrics, because they typically do not contain information about identity. Owing to insufficient metadata, the potential in iris recognition lies mainly in unsupervised machine-learning methods and unsupervised pre-training of the models with a large number of parameters.

6.2. Challenges and competitions

Objective and independent evaluation helps in comparing scientific methodology, and ultimately stimulates the progress within particular domains. The iris recognition method introduced by J. Daugman [45], has been the baseline for years; this is in part due to its simplicity and near perfect recognition rate when combined with appropriate imaging. Numerous improvements have been proposed, however, even with this well-defined baseline the results have varied tremendously; a result of the implementation methods, datasets, and evaluation protocols used [6]. These differences prevent objective comparison between methods, a problem which is amplified by the frequent unavailability of the dataset itself, the effect of which is highlighted in section 3.

Many of these problems can be prevented through the creation of benchmark datasets and independent evaluation. This guarantees the objective comparison between methods through both unified protocols and conditions. Such evaluations are typically organized in the form of competitions and/or challenges. In addition to the creation of publicly available datasets with uniform metrics, this solution encourages competition among researchers.

Details of existing competitions and challenges are beyond the scope of this review, hence in Table 16 only the most popular challenges are presented; grouped in terms of the number of participating research teams, and sorted chronologically. Further details can be found in the following review articles [5,155–157].

7. Summary

Our review shows that there is a substantial heterogeneity in available databases in terms of size (from 18 to > 1,300 classes), sensors used, image quality, etc. This variability means that, for many research questions, there is a database available, but it is not always easy for researchers to identify the best option. Our review provides guidance for identifying the proper database, but also provides recommendations for creating new ones. Because there are many attributes that might be of interest to researchers, it is a challenging task to present a global overview in the format of a research article. For this reason, we added a large table as supplementary material summarizing the properties, as well as additional information (e.g., download links where the datasets can be obtained). We encourage the reader to review the supplementary table. From our bibliometric study, we found that databases from CASIA are the most cited [26,27] (although we discourage using the CASIA Iris v.1 Database), followed by the databases collected at the University of Beira Interior [46,62]. We therefore recommend these datasets as benchmarks when it is a priority to compare their methods with published state of the art. In addition, to obtain these datasets, a license signed by a researcher is sufficient, in contrast to the signature of the institutional legal representative, typically required by others. For comparison studies, it is advisable to use datasets created for certain challenges or competitions, as they come with a standardized protocol for evaluation. For the evaluation of iris images captured by smartphone cameras, those are MICHE [55] and VISOB [50]. To study spoofing and liveness detection, there is LivDet 2017 [161]. To study combinations of different modalities, there are two MBGC datasets [112]. In addition to CASIA and UBI, the Computer Vision Research Lab at the University of Notre Dame [136] has invested a substantial effort in the development of publicly available biometric datasets (including human iris data). Their website hosts 14 different high-quality iris datasets (and more from other modalities), which is the most concentrated web resource that we identified [136]. Although our bibliometric review did not show as high popularity for these as for the datasets at CASIA or UBI, we would like to encourage researchers to explore these datasets in more detail.

8. Recommendation for creating a good iris database

Important aspects of collecting and sharing research data have been discussed by various scientific communities [174,175].

- Plan availability for years to come - The adoption of a new benchmark in the area or iris recognition tends to be rather slow. To keep a database available, it is important to allocate resources necessary for the database distribution for several years into the future. The main resources needed are *(i)* technical - a stable url for the promoting website as well as infrastructure that keeps the website online, and *(ii)* personal – an appointed person responsible for license management as well as for solving problems that interested users might experience.
- Make access simple - we observed that the databases with licenses that can be signed by the individual researchers tend to be more popular. Requiring the signature of the legal institutional representative, especially within a university environment (typically being the rector), is a major hurdle for young researchers. In many cases, they opt to create their own database instead. If the signature from the institutional representative is necessary, we recommend publishing the full license agreement as well as sample of the images from the database on the project website. This helps in deciding whether the database is useful for particular research before starting the administrative procedures to obtain necessary approvals.
- Include a statistically relevant number of samples - Acquiring and handling test subjects is one of the most challenging tasks when creating a biometric database. The number of subjects included should be as large as possible [176]; however, there is always a minimum size for obtaining statistically relevant results. Although this minimum is difficult to quantify for the general case, the statistical significance of 100 samples obtained from the same subjects is not the same as 1000 samples obtained from 100 different subjects.
- Make the database unique - Many authors that adopt certain da tabases, continue using them in subsequent publications. As we have shown in previous sections, a database is typically introduced to explore certain properties or problems in a systematic

Table 16
Competitions and challenges related to iris recognition.

Challenge	Focus	Year	Participants submitted (registered)	Used dataset
ICE 2005 [103]	all stages of iris recognition	2005	9	ND-IRIS-0405
ICE 2006 [102]	all stages of iris recognition on large scale	2006	14	ND-IRIS-0405
NICE-I [158]	iris segmentation in an unconstrained environment	2008	(97)	UBIRIS-V2
MBGC v. 1 [111, 112]	still & portal (walk through) face and iris recognition	2008	14 (68)	MBGC
MBGC v. 2 [111, 112]	still & portal face and iris recognition	2009	13 (78)	MBGC (extended)
NICE-II [159]	iris feature extraction and matching in an unconstrained environment	2010	(67)	UBIRIS-V2
ICB ICIR 2013 [160]	iris recognition in real-world applications	2013	8	CASIA-Iris-Thousand, IR-TestV1
LivDet 2013 [90,161]	iris liveness detection	2013	3	LivDet-Iris-2013 combined dataset
MobBIO 2013 [162]	iris recognition using portable devices	2013	2	MobBIO Multimodal
MICHE-I [56]	mobile iris recognition and segmentation using consumer devices	2014	6	MICHE DB
MobILive 2014 [163]	mobile iris liveness detection	2014	6	MobBIOfake
CCBR CCIR 2014 [164]	iris recognition in real-world applications	2014	–	CASIA-IrisV3-Lamp, IR-TestV1
ICB CSIR 2015 [81]	cross-sensor iris recognition	2015	–	ICB CSIR
LivDet 2015 [91,161]	iris liveness detection	2015	4	LivDet-Iris-2015 combined dataset
SSBC 2015 [165]	sclera segmentation	2015	4 (10)	MASD
MICHE-II [60]	mobile iris feature extraction	2016	7	MICHE DB
Cross-Eyed 2016 (IJCB) [71]	iris and periocular recognition across spectra	2016	3	CROSS-EYED
BTAS MIR 2016 [54]	mobile iris recogniton	2016	3	BTAS MIR
ICIP 2016 CMOBR [49]	mobile ocular recognition	2016	4	VISOB 1
SSRBC 2016 [166]	sclera segmentation and recognition	2016	3 (12)	MASD
Cross-Eyed 2017 (IJCB) [167]	iris and periocular recognition across spectra	2017	5	CROSS-EYED
SSERBC 2017 [168]	sclera segmentation and eye recognition	2017	5 (16)	MASD
LivDet 2017 [79,161]	iris liveness detection	2017	3 (12)	LivDet-Iris-2017 combined dataset
SSBC 2018 [169]	sclera segmentation in cross sensor environment	2017	4	MASD, MSD
SSBC 2019 [170]	sclera segmentation in cross resolution environment	2019	4	MASD, MSD
LivDet 2020 [161,171]	iris liveness detection	2020	3 (3)	LivDet-Iris-2020 combined test dataset
WCCI/IJCNN2020 [172]	long-term mobile ocular recognition	2020	3	VISOB 2
SSBC 2020 [173]	sclera segmentation in mobile environment	2020	13 (26)	MASD, MSD, MOBIUS

manner. A successful database should help users make novel findings and research conclusions. Hence, the database should fulfill the needs of novel research areas where the benchmarks are not yet established. The authors of this review aspire to assist in this task with this review and make the needs more obvious to database creators.

- Extensive protocol and setup description - Although most of the available datasets were created for testing a particular hypothesis or for a particular research purpose, authors frequently argue that the dataset can be useful beyond a single research topic. To increase the potential of the dataset, it is important to provide an extensive description of the protocol and setup. We observed that important information is frequently missing, limiting the use of the datasets, for instance, the wavelength of the NIR illumination, distance at which the images were collected, and descriptions of the sensor or the optical system used.

9. Conclusions

The aim of biometric datasets is to enable the testing of biometric systems or methods. The publicly available datasets, serving as benchmarks, enable objective and reproducible research. Despite the claims of the authors, many of the datasets declared publicly available are not available, for a variety of reasons.

In this paper, we reviewed the existing iris image datasets. We focused mainly on the availability and popularity of the datasets. We raised six different RQs. We identified 158 different datasets, of which only 81 were actually available (answering RQ 1). The full list is provided in the supplementary materials. We identified that databases created by CASIA are the most cited (answering RQ 2). We provided descriptions of the available databases in a structured form (answering RQ 3; see Section 4). Subsequently, we identified common properties of popular databases in Section 5 (answering RQ 4). In Section 6, we discussed the limitations of the databases and areas lacking in available databases. Lastly, based on our review, we formulated appropriate recommendations for creating an iris database in Section 8.

A limitation of this study is that we searched only the Web of Science online library when performing the bibliometric research. Another limitation is that we relied on e-mail and phone contacts when obtaining the dataset.

We aspire to bring clarity in the availability of databases to support reproducible research. However, new databases are continuously being created, while others become unavailable. This requires a continuous effort to keep track of the state-of-the-art and popular databases. We plan to keep this list updated and available on our website.[18]

[18] The full list of available datasets will be published at: https://irisdata.etrovub.be and https://irisdata.fei.stuba.sk

CRediT authorship contribution statement

Lubos Omelina: Conceptualization, Methodology, Writing - original draft, Investigation. **Jozef Goga**: Data curation, Writing - original draft, Investigation. **Jarmila Pavlovičová**: Writing - review & editing, Validation. **Miloš Oravec**: Writing - review & editing, **Funding acquisition**. Bart Jansen: Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Acknowledgments

The research described in the paper was done within the project No. 1/0867/17 of the Slovak Grant Agency VEGA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imavis.2021.104109.

References

[1] J. Daugman, The importance of being random: statistical principles of iris recognition, Pattern Recogn. 36 (2) (2003) 279–291.
[2] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, A. Ross, Long range iris recognition: a survey, Pattern Recogn. 72 (2017) 123–143.
[3] F. Alonso-Fernandez, J. Bigun, A survey on periocular biometrics research, Pattern Recogn. Lett. 82 (2016) 92–105 An insight on eye biometrics.
[4] F. Jan, Segmentation and localization schemes for non-ideal iris biometric systems, Signal Process. 133 (2017) 192–212.
[5] M.D. Marsico, A. Petrosino, S. Ricciardi, Iris recognition through machine learning techniques: A survey, Pattern Recogn. Lett. 82 (2016) 106–115 An insight on eye biometrics.
[6] K.W. Bowyer, K. Hollingsworth, P.J. Flynn, Image understanding for iris biometrics: a survey, Comput. Vis. Image Underst. 110 (2) (2008) 281–307.
[7] J. Neves, F. Narducci, S. Barra, H. Proença, Biometric recognition in surveillance scenarios: a survey, Artif. Intell. Rev. 46 (Dec 2016) 515–541.
[8] A. Rattani, R. Derakhshani, Ocular biometrics in the visible spectrum: a survey, Image Vis. Comput. 59 (2017) 1–16.
[9] K.B. Raja, R. Raghavendra, V.K. Vemuri, C. Busch, Smartphone based visible iris recognition using deep sparse filtering, Pattern Recogn. Lett. 57 (2015) 33–42 Mobile Iris CHallenge Evaluation part I (MICHE I).
[10] Y. Alvarez-Betancourt, M. Garcia-Silvente, An overview of iris recognition: a bibliometric analysis of the period 2000–2012, Scientometrics 101 (Dec 2014) 2003–2033.
[11] R.R. Jillela, A. Ross, Segmenting iris images in the visible spectrum with applications in mobile biometrics, Pattern Recogn. Lett. 57 (2015) 4–16 Mobile Iris CHallenge Evaluation part I (MICHE I).
[12] M. Trokielewicz, Iris recognition with a database of iris images obtained in visible light using smartphone camera, 2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA) Feb 2016, pp. 1–6.
[13] D. Yadav, N. Kohli, M. Vatsa, R. Singh, A. Noore, Unconstrained visible spectrum iris with textured contact lens variations: Database and benchmarking, 2017 IEEE International Joint Conference on Biometrics (IJCB) Oct 2017, pp. 574–580.
[14] P.J. Phillips, K.W. Bowyer, P.J. Flynn, Comments on the casia version 1.0 iris data set, IEEE Trans. Pattern Anal. Mach. Intell. 29 (Oct. 2007) 1869–1870.
[15] L. Ma, Y. Wang, T. Tan, Iris recognition based on multichannel gabor filtering, Proceedings of the International Conference on Asian Conference on Computer Vision 2002, pp. 279–283.
[16] S.S. Arora, M. Vatsa, R. Singh, A. Jain, On iris camera interoperability, 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS) Sept 2012, pp. 346–352.
[17] L. Xiao, Z. Sun, R. He, T. Tan, Coupled feature selection for cross-sensor iris recognition, 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS) Sept 2013, pp. 1–6.
[18] Twins Day Dataset 2010–-1015, https://biic.wvu.edu/data-sets/twins-day-dataset-2010-1015 Accessed: 2019-05-14.
[19] P.J. Phillips, P.J. Flynn, K.W. Bowyer, R.W.V. Bruegge, P.J. Grother, G.W. Quinn, M. Pruitt, Distinguishing identical twins by face recognition, *Face and Gesture 2011*, Pp. 185–192, March 2011.
[20] K. Hollingsworth, K.W. Bowyer, P.J. Flynn, Similarity of iris texture between identical twins, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops 2010, pp. 22–29.
[21] CASIA Iris Ageing database v. 1, http://biometrics.idealtest.org/dbDetailForUser.do?id=14 Accessed: 2018-11-01.
[22] P. Wild, J. Ferryman, A. Uhl, Impact of (segmentation) quality on long vs. short-timespan assessments in iris recognition performance, *IET,* Biometrics 4 (4) (2015) 227–235.
[23] ND-TimeLapseIris-2012 database, https://sites.google.com/a/nd.edu/public-cvrl/data-sets Accessed: 2018-11-02.
[24] S.E. Baker, K.W. Bowyer, P.J. Flynn, P.J. Phillips, Template aging in Iris biometrics, vols. 205–218, Springer London, London, 2013.
[25] S.P. Fenker, K.W. Bowyer, Analysis of template aging in iris biometrics, 2012 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* June 2012, pp. 45–51.
[26] CASIA Iris Image Database Version 1.0, http://biometrics.idealtest.org/dbDetailForUser.do?id=1 Accessed: 2019-12-12.
[27] CASIA Iris Database V4, http://biometrics.idealtest.org/dbDetailForUser.do?id=14 Accessed: 2018-11-05.
[28] IIT Delhi Iris Database (Version 1.0), http://www4.comp.polyu.edu.hk/csajaykr/IITD/Database_Iris.htm Accessed: 2018-10-24.
[29] A. Kumar, A. Passi, Comparison and combination of iris matchers for reliable personal authentication, Pattern Recogn. 43 (3) (2010) 1016–1026.
[30] C.-N. Chun, R. Chung, Iris recognition for palm-top application, in: D. Zhang, A.K. Jain (Eds.), Biometric Authentication, Springer Berlin Heidelberg, Berlin, Heidelberg 2004, pp. 426–433.
[31] CUHK Iris Image Dataset, http://www.mae.cuhk.edu.hk/~cvl/main_database.htm Accessed: 2020-05-20.
[32] Z. Wei, T. Tan, Z. Sun, Synthesis of large realistic iris databases using patch-based sampling, 2008 19th International Conference on Pattern Recognition Dec 2008, pp. 1–4.
[33] J. Zuo, N.A. Schmid, X. Chen, On generation and analysis of synthetic iris images, IEEE Transactions on Information Forensics and Security 2 (March 2007) 77–90.
[34] WVU Synthetic Iris Model Based, https://citer.clarkson.edu/biometric-dataset-collections/synthetic-iris-model-based Accessed: 2018-11-07.
[35] S. Shah, A. Ross, Generating synthetic irises by feature agglomeration, 2006 International Conference on Image Processing Oct 2006, pp. 317–320.
[36] WVU Synthetic Iris Textured Based, https://citer.clarkson.edu/biometric-dataset-collections/synthetic-iris-textured-based Accessed: 2018-11-07.
[37] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, *IEEE Transactions on Circuits and Systems for Video* Technology, 14, Jan 2004, pp. 4–20.
[38] S. Crihalmeanu, A. Ross, S. Schuckers, L. Hornak, A protocol for multibiometric data acquisition, storage and dissemination, tech. Rep., WVU, lane department of computer Science and Electr. Eng, 2007.
[39] P.A. Johnson, P. Lopez-Meyer, N. Sazonova, F. Hua, S. Schuckers, Quality in face and iris research ensemble q-fire, 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS) 2010, pp. 1–6.
[40] Quality-Face/Iris Research Ensemble (Q-FIRE), https://citer.clarkson.edu/research-resources/biometric-dataset-collections-2/quality-faceiris-research-ensemble-q-fire/ Accessed: 2020-05-20.
[41] N. Kihal, S. Chitroub, A. Polette, I. Brunette, J. Meunier, Efficient multimodal ocular biometric system for person authentication based on iris texture and corneal shape, IET Biometrics 6 (6) (2017) 379–386.
[42] Biometric Iris-Cornea Database, http://www-labs.iro.umontreal.ca/labimage/IrisCorneaDataset/ Accessed:2019-05-03.
[43] Y. Yin, L. Liu, X. Sun, Sdumla-hmt: A multimodal biometric database, in: Z. Sun, J. Lai, X. Chen, T. Tan (Eds.), Biometric Recognition, Springer Berlin Heidelberg, Berlin, Heidelberg 2011, pp. 260–268.
[44] Shandong University, Machine Learning and Applications Group (SDUMLA) - the Homologous Multi-modal Traits Database, http://mla.sdu.edu.cn/info/1006/1195.htm Accessed: 2020-05-25.
[45] J. Daugman, How Iris recognition works, IEEE Transact. Circ. Syst. Video Technology 14 (Jan. 2004) 21–30.
[46] H. Proenca, S. Filipe, R. Santos, J. Oliveira, L.A. Alexandre, The ubiris.v2: a database of visible wavelength iris images captured on-the-move and at-a-distance, IEEE Trans. Pattern Anal. Mach. Intell. 32 (Aug 2010) 1529–1535.
[47] C.N. Padole, H. Proenca, Periocular recognition: Analysis of performance degradation factors, 2012 5th IAPR International Conference on Biometrics (ICB) March 2012, pp. 439–445.
[48] W. Dong, Z. Sun, T. Tan, A design of iris recognition system at a distance, 2009 Chinese Conference on Pattern Recognition Nov 2009, pp. 1–5.
[49] A. Rattani, R. Derakhshani, S.K. Saripalle, V. Gottemukkula, Icip 2016 competition on mobile ocular biometric recognition, 2016 IEEE International Conference on Image Processing (ICIP) Sept 2016, pp. 320–324.
[50] Visible light mobile Ocular Biometric (VISOB) Dataset ICIP2016 Challenge Version, http://sce2.umkc.edu/cibit/dataset.html Accessed: 2019-04-14.
[51] CASIA-Iris-Mobile-V1.0 - Casia mobile database (datasets S1, S2 and S3), http://biometrics.idealtest.org/dbDetailForUser.do?id=13 Accessed: 2018-10-24.
[52] Q. Zhang, H. Li, Z. Sun, T. Tan, Deep feature fusion for iris and periocular biometrics on mobile devices, IEEE Transactions on Information Forensics and Security 13 (11) (2018) 2897–2912.
[53] Q. Zhang, H. Li, M. Zhang, Z. He, Z. Sun, T. Tan, Fusion of face and iris biometrics on mobile devices using near-infrared images, *Biometric Recognition* (J. Yang, J. Yang,

Z. Sun, S. Shan, W. Zheng, and J. Feng, Eds.), Springer International Publishing, Cham 2015, pp. 569–578.
[54] M. Zhang, Q. Zhang, Z. Sun, S. Zhou, N.U. Ahmed, The btas competition on mobile iris recognition, 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS) Sep. 2016, pp. 1–7.
[55] MICHE - Mobile Iris CHallenge Evaluation, http://biplab.unisa.it/MICHE/ Accessed: 2020-05-27.
[56] M.D. Marsico, M. Nappi, D. Riccio, H. Wechsler, Mobile iris challenge evaluation (miche)-i, biometric iris dataset and protocols, Pattern Recogn. Lett. 57 (2015) 17–23 Mobile Iris CHallenge Evaluation part I (MICHE I).
[57] M. De Marsico, M. Nappi, F. Narducci, H. Proena, Insights into the results of miche i - mobile iris challenge evaluation, Pattern Recogn. 74 (Feb. 2018) 286–304.
[58] G. Santos, E. Grancho, M.V. Bernardo, P.T. Fiadeiro, Fusing iris and periocular information for cross-sensor recognition, Pattern Recogn. Lett. 57 (2015) 52–59 Mobile Iris CHallenge Evaluation part I (MICHE I).
[59] Cross Sensor Iris and Periocular Database, http://csip.di.ubi.pt Accessed: 2019-04-18.
[60] M. De Marsico, M. Nappi, H. Proença, Results from miche ii – mobile iris challenge evaluation ii, Pattern Recogn. Lett. 91 (2017) 3–10 Mobile Iris CHallenge Evaluation (MICHE-II).
[61] Q. Zhang, H. Li, Z. Sun, Z. He, T. Tan, Exploring complementary features for iris recognition on mobile devices, 2016 International Conference on Biometrics (ICB) June 2016, pp. 1–8.
[62] H. Proença, L.A. Alexandre, Ubiris: A noisy iris image database, *Proceedings of the 13th International Conference on Image Analysis and Processing*, ICIAP'05, Springer-Verlag, Berlin, Heidelberg 2005, pp. 970–977.
[63] M. Edwards, A. Gozdzik, K. Ross, J. Miles, E. Parra, Quantitative measures of iris color using high resolution photographs, Am. J. Phys. Anthropol. 147 (01) (2012) 141–149.
[64] MILES Iris Dataset, https://drive.google.com/drive/folders/0B5OBp4zckpLnU3YxMnozSGhGelE Accessed: 2020-05-20.
[65] M. Dobeš, L. Machala, P. Tichavský, J. Pospíšil, Human eye iris recognition using the mutual information, Optik - Int. J. Light Electron Optics 115 (9) (2004) 399–404.
[66] Palackỳ University Olomouc (UPOL) Iris Image Dataset, http://phoenix.inf.upol.cz/iris/ Accessed: 2020-10-10.
[67] M. Dehnavi, M. Eshghi, Design and implementation of a real time and train less eye state recognition system, EURASIP J. Adv. Signal Process. 2012 (2012) 02.
[68] Eye SBU database, http://facultymembers.sbu.ac.ir/eshghi/index.html Accessed: 2020-05-20.
[69] The Hong Kong Polytechnic University Cross-Spectral Iris Images Database, http://www4.comp.polyu.edu.hk/~csajaykr/polyuiris.htm Accessed: 2018-10-24.
[70] P.R. Nalla, A. Kumar, Toward more accurate iris recognition using cross-spectral matching, Trans. Img. Proc. 26 (Jan. 2017) 208–221.
[71] A. Sequeira, L. Chen, P. Wild, J. Ferryman, F. Alonso-Fernandez, K.B. Raja, R. Raghavendra, C. Busch, J. Bigun, Cross-eyed - cross-spectral iris/periocular recognition database and competition, 2016 International Conference of the Biometrics Special Interest Group (BIOSIG) 2016, pp. 1–5.
[72] Cross-Spectrum Iris/Periocular Recognition Competition Database, https://sites.google.com/site/crossspectrumcompetition/cross-eyed-2016 Accessed: 2020-05-21.
[73] A. Sharma, S. Verma, M. Vatsa, R. Singh, On cross spectral periocular recognition, 2014 IEEE International Conference on Image Processing (ICIP) Oct 2014, pp. 5007–5011.
[74] M.S. Hosseini, B.N. Araabi, H. Soltanian-Zadeh, Pigment melanin: pattern for iris recognition, IEEE Trans. Instrum. Meas. 59 (April 2010) 792–804.
[75] University of Tehran IRIS (UTIRIS) database, https://utiris.wordpress.com/ Accessed: 2018-12-13.
[76] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic, "Image quality assessment for iris biometric," in *SPIE Proceedings Vol. 6202: Biometric Technology for Human Identification III* , Vol. 6202, Pp. 6202–11, SPIE, 2006.
[77] IIITD Multi-spectral Periocular Database, http://iab-rubric.org/resources/impdatabase.html Accessed: 2018-10-30.
[78] IIITD-WVU Mobile Iris Spoofing Dataset, http://iab-rubric.org/resources.html Accessed: 2018-10-30.
[79] D. Yambay, B. Becker, N. Kohli, D. Yadav, A. Czajka, K.W. Bowyer, S. Schuckers, R. Singh, M. Vatsa, A. Noore, D. Gragnaniello, C. Sansone, L. Verdoliva, L. He, Y. Ru, H. Li, N. Liu, Z. Sun, T. Tan, Livdet iris 2017 - iris liveness detection competition 2017, 2017 IEEE International Joint Conference on Biometrics (IJCB) Oct 2017, pp. 733–741.
[80] ND-CrossSensor-Iris-2013 database, https://sites.google.com/a/nd.edu/public-cvrl/data-sets Accessed: 2018-11-02.
[81] Dataset provided within the ICB Competition on Cross-sensor Iris Recognition, http://biometrics.idealtest.org/2015/csir2015.jsp Accessed: 2018-11-05.
[82] S.E. Baker, A. Hentz, K.W. Bowyer, P.J. Flynn, Degradation of iris recognition performance due to non-cosmetic prescription contact lenses, Comput. Vis. Image Underst. 114 (9) (2010) 1030–1044.
[83] J.S. Doyle, K.W. Bowyer, Robust detection of textured contact lenses in iris recognition using bsif, IEEE Access 3 (2015) 1672–1683.
[84] D. Yadav, N. Kohli, J.S. Doyle, R. Singh, M. Vatsa, K.W. Bowyer, Unraveling the effect of textured contact lenses on iris recognition, IEEE Transactions on Information Forensics and Security 9 (May 2014) 851–862.
[85] J.S. Doyle, K.W. Bowyer, P.J. Flynn, Variation in accuracy of textured contact lens detection based on sensor and lens pattern, 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS) Sept 2013, pp. 1–7.
[86] N. Kohli, D. Yadav, M. Vatsa, R. Singh, A. Noore, Detecting medley of iris spoofing attacks using desist, 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS) Sept 2016, pp. 1–6.
[87] Iris Combined Spoofing Database, http://iab-rubric.org/resources.html Accessed: 2018-10-31.
[88] P. Gupta, S. Behera, M. Vatsa, R. Singh, On iris spoofing using print attack, 2014 22nd International Conference on Pattern Recognition Aug 2014, pp. 1681–1686.
[89] Iris Combined Spoofing Database, http://iab-rubric.org/resources.html Accessed: 2018-10-30.
[90] D. Yambay, J.S. Doyle, K.W. Bowyer, A. Czajka, S. Schuckers, Livdet-iris 2013 - iris liveness detection competition 2013, *IEEE International Joint Conference on Biometrics*, Pp. 1–8, Sept 2014.
[91] D. Yambay, B. Walczak, S. Schuckers, A. Czajka, Livdet-iris 2015 - iris liveness detection competition 2015, 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA) Feb 2017, pp. 1–6.
[92] ATVS-FIr iris database, https://atvs.ii.uam.es/atvs/fir_db.html Accessed: 2019-05-04.
[93] J. Fierrez, J. Ortega-Garcia, D.T. Toledano, J. Gonzalez-Rodriguez, Biosec baseline corpus: A multimodal biometric database, Pattern Recogn. 40 (4) (2007) 1389–1392.
[94] V. Ruiz-Albacete, P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez, J. Ortega-Garcia, Direct attacks using fake images in iris verification, in: B. Schouten, N.C. Juul, A. Drygajlo, M. Tistarelli (Eds.), Biometrics and Identity Management, Springer Berlin Heidelberg, Berlin, Heidelberg 2008, pp. 181–190.
[95] Multi-Angle Sclera Dataset (MASD) version 1, https://sites.google.com/site/dasabhijit2048/datatsets Accessed: 2020-05-27.
[96] A. Das, U. Pal, M.A. Ferrer, M. Blumenstein, A framework for liveness detection for direct attacks in the visible spectrum for multimodal ocular biometrics, Pattern Recogn. Lett. 82 (2016) 232–241 An insight on eye biometrics.
[97] I. Rigas, O.V. Komogortsev, Gaze estimation as a framework for iris liveness detection, IEEE International Joint Conference on Biometrics Sep. 2014, pp. 1–8.
[98] Eye Tracker Print-Attack Database (ETPAD) v1, https://userweb.cs.txstate.edu/~ok11/etpad_v1.html Accessed: 2019-05-16.
[99] Eye Tracker Print-Attack Database (ETPAD) v2, https://userweb.cs.txstate.edu/~ok11/etpad_v2.html Accessed: 2019-05-16.
[100] K.W. Bowyer, P.J. Flynn, The ND-IRIS-0405 iris image dataset, Technical Report, University of Notre Dame, CoRR (2009).
[101] "ND-Iris-0405 Data Set." https://sites.google.com/a/nd.edu/public-cvrl/data-sets. Accessed: 2018-11-02.
[102] P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, M. Sharpe, Frvt 2006 and ice 2006 large-scale experimental results, IEEE Trans. Pattern Anal. Mach. Intell. 32 (May 2010) 831–846.
[103] P. J, K.W. Bowyer, P.J. Flynn, X. Liu, W.T. Scruggs, The iris challenge evaluation 2005, 2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems Sept 2008, pp. 1–8.
[104] A. Das, U. Pal, M.A.F. Ballester, M. Blumenstein, Multi-angle based lively sclera biometrics at a distance, 2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM) 2014, pp. 22–29.
[105] Multimedia University Iris Database (MMU) - V1 and V2, https://mmuexpert.mmu.edu.my/ccteo Accessed: 2020-10-01.
[106] C.C. Teo, H.F. Neo, G.K.O. Michael, C. Tee, K.S. Sim, A robust iris segmentation with fuzzy supports, in: K.W. Wong, B.S.U. Mendis, A. Bouzerdoum (Eds.), Neural Information Processing. Theory and Algorithm, Springer Berlin Heidelberg, Berlin, Heidelberg 2010, pp. 532–539.
[107] Y. Chen, Y. Liu, X. Zhu, H. Chen, F. He, Y. Pang, Novel approaches to improve iris recognition system performance based on local quality evaluation and feature fusion, TheScientificWorldJournal 670934 (2014) 02.
[108] G. Huo, Y. Liu, X. Zhu, H. Dong, Secondary iris recognition method based on local energy-orientation feature, Journal of Electronic Imaging 24 (1) (2015) 1–13.
[109] Jilin University Iris Biometric and Information Security Lab (JLUBRIRIS v1-v6) Database, http://biis.jlu.edu.cn/ Accessed: 2019-05-01.
[110] C.C. Teo, H.T. Ewe, An efficient one-dimensional fractal analysis for iris recognition, Proceedings of WSCG 2005, 2005.
[111] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A., Scallan, I., and S. Weimer, "Overview of the multiple biometrics grand challenge," in *Proceedings of the Third International Conference on Advances in Biometrics*, ICB '09, (Berlin, Heidelberg), pp. 705–714, Springer-Verlag, 2009.
[112] Multiple Biometric Grand Challenge (MBGC), https://www.nist.gov/programs-projects/multiple-biometric-grand-challenge-mbgc Accessed: 2019-04-14.
[113] R. Jillela, A.A. Ross, V.N. Boddeti, B.V.K.V. Kumar, X. Hu, R. Plemmons, P. Pauca, Iris segmentation for challenging periocular images, vols. 281–308, Springer London, London, 2013.
[114] Face and Ocular Challenge Series (FOCS), https://www.nist.gov/programs-projects/face-and-ocular-challenge-series-focs Accessed: 2019-04-14.
[115] C.N. Padole, H. Proença, Compensating for pose and illumination in unconstrained periocular biometrics, IJBM 5 (2013) 336–359.
[116] "UBI Periocular Dataset." http://socia-lab.di.ubi.pt/~ubipr/index.html. Accessed: 2019-04-08.
[117] Cataract Mobile Periocular Database (CMPD), http://www.iab-rubric.org/resources/cmpd.html Accessed: 2018-10-29.
[118] R. Keshari, S. Ghosh, A. Agarwal, R. Singh, M. Vatsa, Mobile periocular matching with pre-post cataract surgery, 2016 IEEE International Conference on Image Processing (ICIP) Sept 2016, pp. 3116–3120.
[119] M. Singh, S. Nagpal, M. Vatsa, R. Singh, A. Noore, A. Majumdar, Gender and ethnicity classification of iris images using deep class-encoder, In: IEEE Int. Jt. Conf. Biometrics, IJCB 2017, 2018 (2018)https://doi.org/10.1109/BTAS.2017.8272755.
[120] ND-TWINS-2009-–2010 Still Face database, https://cvrl.nd.edu/projects/data/#nd-twins-2009-2010 Accessed: 2019-05-14.

[121] H. Mehrotra, M. Vatsa, R. Singh, B. Majhi, Does iris change over time? PLoS One e78333 (2013) 11.
[122] P. Basak, S. De, M. Agarwal, A. Malhotra, M. Vatsa, R. Singh, Multimodal biometric recognition for toddlers and pre-school children, 2017 IEEE International Joint Conference on Biometrics (IJCB) Oct 2017, pp. 627–633.
[123] J.E. Tapia, C.A. Perez, K.W. Bowyer, Gender classification from the same iris code used for recognition, IEEE Transactions on Information Forensics and Security 11 (Aug 2016) 1760–1770.
[124] B. Brown, A.J. Adams, G. Haegerstrom-Portnoy, R.T. Jones, M.C. Flom, Pupil size after use of marijuana and alcohol, Am J. Ophthalmol. 83 (3) (1977) 350–354.
[125] J.E. Richman, K.G. McAndrew, D. Decker, S.C. Mullaney, An evaluation of pupil size standards used by police officers for detecting drug impairment, Optometry - J. Am. Optomet. Associat. 75 (3) (2004) 175–182.
[126] S.S. Arora, M. Vatsa, R. Singh, A. Jain, Iris recognition under alcohol influence: A preliminary study, 2012 5th IAPR International Conference on Biometrics (ICB) March 2012, pp. 336–341.
[127] I. Tomeo-Reyes, V. Chandran, Part based bit error analysis of iris codes, Pattern Recogn. 60 (2016) 306–317.
[128] P. Jain, P. Finger, Iris varix: 10-year experience with 28 eyes, Indian J. Ophthalmol. 67 (03) (2019) 350.
[129] The Eye Cancer Foundation Dataset, http://www.eyecancercure.com/research/image-gallery https://eyecancer.com/eye-cancer/image-galleries/iris-tumors Accessed: 2019-05-16.
[130] M. Trokielewicz, A. Czajka, P. Maciejewicz, Database of iris images acquired in the presence of ocular pathologies and assessment of iris recognition reliability for disease-affected eyes, 2015 IEEE 2nd International Conference on Cybernetics (CYBCONF) June 2015, pp. 495–500.
[131] M. Trokielewicz, A. Czajka, P. Maciejewicz, Assessment of iris recognition reliability for eyes affected by ocular pathologies, 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS) Sep. 2015, pp. 1–6.
[132] Z. Luo, J. Dai, Y. Jia, J. He, An improved bovine iris segmentation method, MATEC Web of Conferences, 267, 2019, p. 03002 , 01.
[133] L.Z. Menglu Zhang, An iris localization algorithm based on geometrical features of cow eyes, *Proc SPIE*, vol. 7495, 2009.
[134] M. Trokielewicz, M. Szadkowski, Iris and periocular recognition in arabian race horses using deep convolutional neural networks, 2017 IEEE International Joint Conference on Biometrics (IJCB) Oct 2017, pp. 510–516.
[135] D.M. Monro, S. Rakshit, D. Zhang, Dct-based iris recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (apr 2007) 586–595.
[136] The Notre Dame Computer Vision Research Laboratory (CVRL), Datasets, https://cvrl.nd.edu/projects/data/ Accessed: 2019-12-15.
[137] Data protection in EU, https://ec.europa.eu/info/law/law-topic/data-protection_en Accessed: 2019-12-15.
[138] A. Czajka, Template ageing in iris recognition, BIOSIGNALS 2013 - Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing, 04, 2013, pp. 1–8.
[139] K. Browning and N. M. Orlans, “Biometric Aging - Effects of Aging on Iris Recognition.” https://www.mitre.org/publications/technical-papers/biometric-aging-effects-of-aging-on-iris-recognition, 2014. MITRE Corp.,Techniacl paper, Accessed: 2020-04-28.
[140] Council of European Union, Regulation (eu) 2016/679 of the european parliament and of the council, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679 2016.
[141] E. Simperl, K. O'Hara, R. Gomer, Analytical Report 3: Open Data and Privacy, tech. rep., European Data Portal, 06, 2016 https://www.europeandataportal.eu/sites/default/files/open_data_and_privacy_v1_final_clean.pdf Accessed: 2020-04-28.
[142] A. Karpenko, D. Jacobs, J. Baek, M. Levoy, Digital video stabilization and rolling shutter correction using gyroscopes, tech. rep, Computer Science Tech Report, Stanford University, Sept. 2011.
[143] J.R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D.J. LoIacono, S. Mangru, M. Tinker, T.M. Zappia, W.Y. Zhao, Iris on the move: acquisition of images for iris recognition in less constrained environments, Proc. IEEE 94 (11) (2006) 1936–1947.
[144] D. Crouse, H. Han, D. Chandra, B. Barbello, A.K. Jain, Continuous authentication of mobile user: Fusion of face image and inertial measurement unit data, 2015 International Conference on Biometrics (ICB) 2015, pp. 135–142.
[145] A. Das, C. Galdi, H. Han, R. Ramachandra, J. Dugelay, A. Dantcheva, Recent advances in biometric technology for mobile devices, In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2018.
[146] Z. Ali, U. Park, J. Nang, J. Park, T. Hong, S. Park, Periocular recognition using umlbp and attribute features, KSII Transactions on Internet and Information Systems 11 (12) (2017) 6133–6151.
[147] S. Barra, A. Casanova, F. Narducci, S. Ricciardi, Ubiquitous iris recognition by means of mobile devices, Pattern Recogn. Lett. 57 (2015) 66–73 Mobile Iris CHallenge Evaluation part I (MICHE I).
[148] F. Alonso-Fernandez, R.A. Farrugia, J. Bigun, J. Fierrez, E. Gonzalez-Sosa, A survey of super-resolution in iris biometrics with evaluation of dictionary-learning, IEEE Access 7 (2019) 6519–6544.
[149] M. Bielikova, M. Konopka, J. Simko, R. Moro, J. Tvarozek, P. Hlaváč, E. Kuric, Eye-tracking en masse: group user studies, lab infrastructure, and practices, J. Eye Mov. Res. 11 (2018) 08.
[150] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, A. Torralba, Eye tracking for everyone, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) June 2016, pp. 2176–2184.
[151] S. Joshi, Y. Li, R. Kalwani, J. Gold, Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi and cingulate cortex, Neuron 89 (1) (2016) 221–234.
[152] M. Szwoch, P. Pieniążek, Eye blink based detection of liveness in biometric authentication systems using conditional random fields, Computer Vision and Graphics, Springer Berlin Heidelberg, Berlin, Heidelberg 2012, pp. 669–676.
[153] A.S. Adhau, D.K. Shedge, Iris recognition methods of a blinked eye in nonideal condition, 2015 International Conference on Information Processing (ICIP) 2015, pp. 75–79.
[154] A. Fogelton, W. Benesova, Eye blink completeness detection, Comput. Vis. Image Underst. 176-177 (2018) 10.
[155] A. Czajka, K.W. Bowyer, Presentation attack detection for iris recognition: an assessment of the state-of-the-art, ACM Comput. Surv. (July 2018) 51.
[156] D. Yambay, A. Czajka, K. Bowyer, M. Vatsa, R. Singh, A. Noore, N. Kohli, D. Yadav, S. Schuckers, Review of Iris Presentation Attack Detection Competitions, Springer International Publishing, Cham, 2019 169–183.
[157] L. A. Zanlorensi, R. Laroca, E. Luz, A. S. B. J. au2, L. S. Oliveira, And D. Menotti, “Ocular recognition databases and competitions: A survey,” 2019.
[158] H. Proenca, L.A. Alexandre, The nice.I: Noisy iris challenge evaluation - part i, 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems 2007, pp. 1–4.
[159] K.W. Bowyer, The results of the nice.ii iris biometrics competition, Pattern Recognition Letters, 33, 2012, pp. 965–969 , Noisy Iris Challenge Evaluation II - Recognition of Visible Wavelength Iris Images Captured At-a-distance and On-the-move.
[160] Man Zhang, Jing Liu, Z. Sun, T. Tan, Su Wu, F. Alonso-Fernandez, V. Némesin, N. Othman, K. Noda, P. Li, E. Hoyle, A. Joshi, The first icb* competition on iris recognition, *IEEE International Joint Conference on Biometrics*, Pp. 1–6, 2014.
[161] LivDet - Liveness Detection Competitions, http://livdet.org/ Accessed: 2020-05-27.
[162] A. Sequeira, J. Monteiro, A. Rebelo, H. Oliveira, Mobbio: A multimodal database captured with a portable handheld device, *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, Vol. 3, 01, 2014.
[163] A. Sequeira, H. Oliveira, J. Monteiro, J. Monteiro, J. Cardoso, Mobilive 2014 -mobile iris liveness detection competition, *IJCB 2014–2014 IEEE/IAPR International Joint Conference on Biometrics*, 09, 2014.
[164] The First CCBR Competition on Iris Recognition, http://biometrics.idealtest.org/2014/CCIR2014.jsp Accessed: 2020-10-10.
[165] A. Das, U. Pal, M. Ferrerc, M. Blumensteina, Ssbc 2015: Sclera segmentation benchmarking competition, 2015 *IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 09, 2015, pp. 1–6.
[166] A. Das, U. Pal, M. Ferrer, M. Blumenstein, Ssrbc 2016: Sclera segmentation and recognition benchmarking competition, 2016 *International Conference on Biometrics (ICB)*, 06, 2016, pp. 1–6.
[167] A.F. Sequeira, L. Chen, J. Ferryman, P. Wild, F. Alonso-Fernandez, J. Bigun, K.B. Raja, R. Raghavendra, C. Busch, T. de Freitas Pereira, S. Marcel, S.S. Behera, M. Gour, V. Kanhangad, Cross-eyed 2017: Cross-spectral iris/periocular recognition competition, 2017 IEEE International Joint Conference on Biometrics (IJCB) 2017, pp. 725–732.
[168] A. Das, U. Pal, M. Ferrer, M. Blumenstein, D. Stepec, P. Rot, Z. Emersic, P. Peer, V. Struc, S.V. Aruna Kumar, B.S. Harish, Sserbc 2017: Sclera segmentation and eye recognition benchmarking competition, 2017 *IEEE International Joint Conference on Biometrics (IJCB)*, 10, 2017, pp. 742–747.
[169] A. Das, U. Pal, M. Ferrer, M. Blumenstein, D. Stepec, P. Rot, Z. Emersic, P. Peer, V. Struc, Ssbc 2018: Sclera segmentation benchmarking competition, 2018 International Conference on Biometrics (ICB), 02, 2018, pp. 303–308.
[170] A. Das, U. Pal, M. Blumenstein, z. sun, Sclera segmentation benchmarking competition in cross-resolution environment, *ICB 2019*, 06, 2019.
[171] P. Das, J. McGrath, Z. Fang, A. Boyd, G. Jang, A. Mohammadi, S. Purnapatra, D. Yambay, S. Marcel, M. Trokielewicz, P. Maciejewicz, K. Bowyer, A. Czajka, S. Schuckers, J. Tapia, S. Gonzalez, M. Fang, N. Damer, F. Boutros, A. Kuijper, R. Sharma, C. Chen, A. Ross, Iris liveness detection competition (livdet-iris) – the 2020 edition, 2020.
[172] VISible light mobile Ocular Biometric (VISOB) 2.0 dataset (WCCI/IJCNN2020 Challenge Version), https://sce.umkc.edu/research-sites/cibit/dataset.html Accessed: 2020-10-10.
[173] M. Vitek, A. Das, Y. Pourcenoux, A. Missler, C. Paumier, S. Das, I. Ghosh, D.R. Lucio, L. Zanlorensi, D. Menotti, F. Boutros, N. Damer, J. Grebe, A. Kuijper, J. Hu, Y. He, C. Wang, H. Liu, Y. Wang, R. Vyas, Ssbc 2020: Sclera segmentation benchmarking competition in the mobile environment, *IJCB 2020*, 10, 2020.
[174] A. de Waard, H. Cousijn, and I. J. Aalbersberg, “10 aspects of highly effective research data.” https://www.elsevier.com/connect/10-aspects-of-highly-effective-research-data, 2015. https://www.elsevier.com/connect/10-aspects-of-highly-effective-research-data, Accessed: 2019-12-12.
[175] H. Aguinis, N.S. Hill, J. Bailey, Best practices in data collection and preparation: recommendations for reviewers, editors, and authors, Organ. Res. Methods 109442811983648 (2019) 03.
[176] ISO IEC 19795–1:2006 - Information technology – Biometric performance testing and reporting – Part 1: Principles and framework, tech. rep, 2006.