Journal of Ambient Intelligence and Humanized Computing (2023) 14:6517–6529
https://doi.org/10.1007/s12652-021-03525-x

ORIGINAL RESEARCH

Framework for biometric iris recognition in video, by deep learning and quality assessment of the iris-pupil region

Eduardo Garea-Llano[1] · **Annette Morales-Gonzalez**[2]

Received: 15 October 2020 / Accepted: 23 September 2021 / Published online: 5 October 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
In the current world scenario the influence of the COVID19 pandemic has reached universal proportions affecting almost all countries. In this sense, the need has arisen to wear gloves or to reduce direct contact with objects (such as sensors for capturing fingerprints or palm prints) as a sanitary measure to protect against the virus. In this new reality, it is necessary to have a biometric identification method that allows safe and rapid recognition of people at borders, or in quarantine controls, or in access to places of high biological risk, among others. In this scenario, iris biometric recognition has reached increasing relevance. This biometric modality avoids all the aforementioned inconveniences with proven high efficiency. However, there are still problems associated with the iris capturing and segmentation in real time that could affect the effectiveness of a System of this nature and that it is necessary to take into account. This work presents a framework for real time iris detection and segmentation in video as part of a biometric recognition system. Our proposal focuses on the stages of image capture, iris detection and segmentation in RGB video frames under controlled conditions (conditions of border and access controls, where people collaborate in the recognition process). The proposed framework is based on the direct detection of the iris-pupil region using the YOLO network, the evaluation of its quality and the semantic segmentation of iris by a Fully Convolutional Network. (FCN). The proposal of an evaluation step of the quality of the iris-pupil region reduce the passage to the system of images with problems of out of focus, blurring, occlusions, light changing and pose of the subject. For the evaluation of image quality, we propose a measure that combines parameters defined in ISO/IEC 19794-6 2005 and others derived from the systematization of the knowledge of the specialized literature. The experiments carried out in four different reference databases and an own video data set demonstrates the feasibility of its application under controlled conditions of border and access controls. The achieved results exceed or equal state-of-the-art methods under these working conditions.

Keywords Iris detection · Iris segmentation · Video · Image quality

1 Introduction

The COVID19 pandemic has negatively influenced the flow of passengers in ports and airports worldwide, affecting most countries. Current border control systems are based on biometric face recognition, in the same way a good part of access control systems are based on fingerprint biometrics. Due to the need for the mandatory use of sanitary masks established by many countries, the need to reduce direct contact with sensors in order to avoid or reduce the risk of infection, it is necessary to have biometric recognition system that allows the safe recognition of people at the borders, quarantine controls, access to places of high biological risk, among others. In this circumstance, iris biometric recognition has become vitally important, because precisely this modality avoids all the aforementioned inconveniences with a proven high efficiency.

Since its inception in the 1990s, the process of capturing iris images for biometric recognition has been characterized by the use of near infrared (NIR) sensors. The use of this type of sensors has shown to be effective for iris images with low contrast between the iris and the pupil. This problem occurs in people with dark iris colors due to the high content of melanin, especially those of African or

✉ Eduardo Garea-Llano
eduardo.garea@cneuro.edu.cu

Annette Morales-Gonzalez
amorales@cenatav.co.cu

[1] Cuban Neuroscience Center, Havana, Cuba

[2] Advanced Technologies Application Center, Havana, Cuba

mixed origin (Proenc¸a 2016). However, this fact, which in certain circumstances may be an advantage, in others becomes a limitation due to the relative high cost of the hardware necessary to implement a practical solution. For example, for the person identification at the borders, biometric face recognition applications are generally used. For this task, cameras in the visible spectrum (VS) are used for image capturing (web cameras or IP video cameras), so under current conditions, a change from facial biometric identification method to one based on iris using NIR cameras would be very expensive. On the other hand, in recent years there has been a substantial increase in biometric applications based on the use of VS cameras integrated into mobile devices. Iris biometrics has not been left out of this development (Raja et al. 2016). In this context, video in VS as a way to capture the iris in real-time applications is a topic that acquires an increasing importance (Hollingsworth et al. 2009; Garea-Llano et al. 2018). For these reasons, it is necessary to direct research towards this type of iris image capture due to the problems that characterize it.

In this same sense, iris recognition in VS represents one of the greatest challenges. Some recent works (Proenc¸a 2016) confirm that if the iris images in the VS are captured appropriately, they can exhibit a level of detail in their texture similar to those captured in the NIR, while generally the amount of information they have is less. Nevertheless, different noises, such as specular reflections and shadows, may appear within the images in the VS, leading to increased intra-class variations.

In Raja et al. (2016) the use of white LED light is proposed as a way to attenuate the effect of low contrast between dark-colored irises and pupils in images captured by mobile devices in the VS. This work shows that this proposal increases recognition rates to a similar level obtained with images in the NIR.

In this work we explore the idea of the combination of fast and accurate object detection method, the evaluation of the image quality of detected objects and their semantic segmentation for video in real time as steps prior to recognition in an iris biometric system. The final objective would be the application of this system under controlled scenarios. Then, this work presents a framework for iris detection and segmentation in video taken in the VS under controlled conditions (conditions of border controls and access controls, where people collaborate in the recognition process) with the purpose of biometric recognition.

The main contributions of this work are:

- The proposal of an iris-pupil region detector based on the YOLO network. The proposed detector increases the accuracy and speed of the detection and allows its use in real time video capture.
- The proposal of a stage and measure for image quality evaluation on the detected iris pupil region which prevent the entry into the recognition system of poor quality images (with problems of out of focus, blurring, occlusions, lighting and changes of pose of the subject). This measure is capable to perform the evaluation of the detected images in real time during the video capture.
- The proposal of an iris segmentation method from the iris-pupil region based on its semantic classification by a convolutional network

The work is structured as follows: In Sect. 2 related works are discussed, Sect. 3 presents the proposed framework and Sect. 4 presents the experimental design, its results and discussion. Finally the conclusions of this research are presented.

2 Related works

The main characteristic of the human iris is its complex structure in which there is a large amount of information that can be used for iris biometry. Today, due to recent technological advances, iris recognition has evolved towards the use of less sophisticated technologies compared to its first applications (Jayavadivel and Prabaharan 2021). In general, iris recognition includes different stages such as the image acquisition, pre-processing, segmentation, normalization, feature extraction, and matching. A good part of the most recent works focus on the three first stages of the process because they guarantee the quality of the information that is going to be introduced to the recognition system.

Iris recognition has proven to be less accurate when performed in poorly controlled environments, such as outdoors, at a distance greater than 1 m, and with little or no cooperation from the user. To achieve a correct identification, the image of the iris must be of high quality. An image of the iris with a very large and off-centered pupil may be rejected in the acquisition phase. Consequently, many attempts are made to acquire the iris which delay recognition and irritate the user. This results in poor iris identification. The rate of failed registrations or the rate of rejection of poor quality images would tend to increase in these cases (Bansal 2020).

The problem of object detection is a key issue in computer vision. This topic has two fundamental research lines and challenges to be faced, the object detection (position) and their class estimation (classification).

Computer vision-based object detection methods have been used in a wide variety of applications, including facial recognition tasks, vehicle and people tracking in video, autonomous driving, and medical diagnostics. The first methods described in the literature for object detection are mainly based on the use of key points, edges or templates,

which leads to low detection precision and limited their practical applications to highly controlled environments at laboratory levels. A number of methods that have obtained better results in the detection of objects use the extraction of characteristics such as Haar characteristics, HOG (Histogram of oriented gradient) and LBP (Local binary patterns) (Agarwal et al. 2019).

In Verma et al. (2018). The eye detection is done by Viola–Jones algorithm and iris detection is done by circular Hough transform technique for checking the eyes state as previous step for a real time Driver Assistance System. They achieve an accuracy of 99% during daytime and an accuracy of 96% during nighttime and 91% accuracy for noisy frames.

Naz et al. (2019) proposed a scheme for Driver Fatigue Detection, to detect the eyes in video, they used Viola-Jones algorithm. Once the eyes have been localized, they are classified as open or closed using three different techniques. The experiments showed high performance in detecting driver fatigue, but the authors did not present results in terms of the accuracy and speed of ocular detection.

Garea et al. (2019) proposed a detector based on the classic Viola and Jones (2004) for the direct detection of the iris-pupil region in video scenes. This detector takes advantage of the integral image representation to achieve a quick calculation of the features used in the process. The Adaboost-based learning algorithm allows the selection of a small number of features from the initial set and a cascade of simple classifiers to detect the iris-pupil region.

The main problem with these approaches is their relatively low precision in the process of detecting the edges of the iris-pupil region and the speed of detection. These problems lead to the loss of video frames in which the detector cannot find the presence of the iris-pupil regions. Another drawback is associated with the nature of the detector, which is highly dependent on the nature of the positive and negative training samples used in the training process. This problem leads to low extrapolation ability, so that when capture conditions change, detection rates drop significantly.

In early 2010s, as a result of the development achieved by graphic processors and the creation of large image databases, convolutional neural networks (CNN) were proposed with a powerful performance of image expression and understanding.

In Minhaz et al. (2021), a video framework for detecting abnormal behavior based on frequent iris movements was presented. They proposed a Multi-task Cascaded Convolutional Networks classifier to detect eyes and observe irises from eye representations using Circular Hough Transformation (CHT). In the experiments on their own video dataset they reached an accuracy of 94% on detecting a frequent iris movement in real time.

According to Pavlov and Galeeva (2019), CNN methods for object detection can be divided into two categories: single-state methods, which provide an estimate the position and the class in a single step and two-state methods, which first detect image regions where the object may be located and then apply a classifier to these regions. The first type of these methods is the one with the highest speed and at the same time they have an accuracy similar to those of the second type. Among them, methods that offer the best results are the Single-Shot Detector (SSD) (Liu et al. 2016) and the You Only Look Once (YOLO) (Redmon et al. 2016). Both methods provide the probability of each possible class as outputs, not for the entire image, but for a series of regularly spaced positions for different scales of rectangles called boxes.

In Redmon et al. (2016) the first regression-based method was proposed, YOLO (You Only Look Once), this method is capable of simultaneously predict the bounding box coordinates and classify the objects in an end-to-end neural network. The introduction of YOLO allowed to detect objects in real time, although it was limited to large objects and the accuracy in the bounding box detection was also low. In this way. Liu et al. (2016) proposed the SSD which introduced reference boxes and detected objects on multi-scale feature maps to improve detection accuracy. Redmon and Farhadi (2017) proposed the YOLO-V2 method, which achieves higher precision and speed compared to the YOLO method. YOLO-V2 removed the fully-connected layers and introduced anchor boxes to predict bounding boxes, making the YOLO detector faster and more robust than YOLO. YOLO-V3 (Redmon and Farhadi 2018) uses the residual structure to further deepen the network layer and achieves a breakthrough in accuracy. Tiny-YOLO-V3 (Redmon and Farhadi, 2018) is an improvement of YOLO with a relatively small model size for constrained environments.

Estimating the quality of images is one of the topics that has taken on great importance in biometrics. Quality measures are used to decide when the image should be discarded or processed by the recognition systems. In iris biometry, some works have been presented (Schmid et al. 2016; Daugman and Downing 2016).

In Schmid et al. (2016) the quality of iris images is determined by many factors that depend on the environmental conditions, the characteristics of the sensor and person to be recognized. Among some of the quality measures reported in the literature, Daugman and Downing (2016) focus on evaluating the quality after the iris segmentation process, which results in saturation of the system with a significant number of low-quality images. Another problem with the proposed quality measures for the iris is that they only evaluate a single parameter or a small set of them (Garea-Llano et al. 2018), such as blurring, motion blur and some types of occlusions.

Other authors (Zuo and Schmid 2008; Daugman and Downing 2016) use a large set of factors that include, in

addition to those already mentioned, specular reflections, lighting problems and off-angle gaze. These authors consider that if at least one of the estimated parameters falls below the established thresholds, the measure that integrates them automatically takes a zero value ("veto power"). This can be a drawback for a video-based system, since the video capture conditions will not always be optimal and a quality measure of this nature can limit the passage of images that may have an identifying value.

In Garea-Llano et al. (2018 and 2019), quality measures were proposed to estimate the quality within the complete eye image captured by a video VS camera. These measures consider parameters established in the ISO/IEC 19794-6: 2005 standard and combine them with methods to estimate image sharpness. However, the quality estimation is developed in the whole eye image and not in the specific iris-pupil region (see Fig. 1). Additionally, the level of sharpness and diversity of the iris texture is not considered.

In this work we present a quality measure that considers the parameters established in the standard combined with the estimation of image sharpness and level of diversity of the iris texture through the image entropy. Unlike the previous proposals, quality estimation is made not on the whole eye image, but on the iris-pupil image previously detected from the proposal presented in Sect. 3.2.

Iris segmentation involves detection of iris outer and inner borders in an eye image (Fig. 1a). This task may also include the detection of the eyebrows, the exclusion of the eyelashes as well as edges of the contact lenses and specular reflections (Daugman 2004). The accurate segmentation of the iris region in images taken in the VS is still a complex task due to the large number of factors influencing the image quality (Proenc¸a and Alexandre 2010).

In the first part of the Noisy Iris Challenge Evaluation (NICE.I) (Proenc¸a and Alexandre 2012), the optimized versions of the integrodifferential operator proposed by Daugman (Tan et al. 2010; Sankowski et al. 2010) were among the best positioned methods. More recent work demonstrates that image classification into two classes (iris and non-iris) based on machine learning concepts further improve segmentation accuracy (Liu et al. 2016).

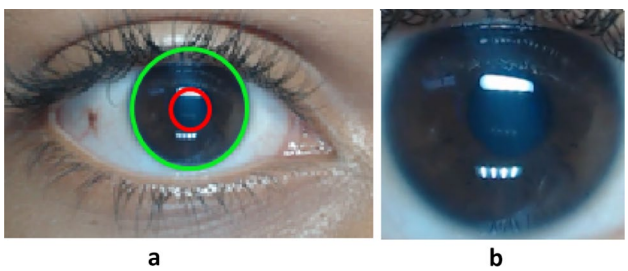

Fig. 1 Eye region (**a**); iris-pupil region (**b**)

In recent years, studies based on the use of FCNs have achieved the best results (Liu et al. 2016). Jalilian and Uhl (2017) explored the performance of three types of FCNs in the task of iris segmentation, the studied networks presented difficulties segmenting images with problems of vision angle and taken from various distances with different types of occlusions (including glasses). The main limitation of these methods is that they only take into account the existence of two classes (iris, non-iris).

In Liu et al. (2016), the obtained results were poor for a number images of dark-skinned people or images that did not contain eyes. As expressed by the authors, this is due to the limited number of training examples in the NICE I database.

To face these problems, in Osorio-Roig et al. (2017) a semantic segmentation algorithm is presented to extract iris regions. The algorithm based on HMRF-PyrSeg, was trained with 53 manually annotated eye images taken from the training set of the NICE I database. Although the results did not improve the state of the art of segmentation methods, the authors showed that the idea of multi-class segmentation was a promising path for iris segmentation in VS images.

In Osorio-Roig et al. (2018), a method of multiclass iris segmentation was presented. This method is based on the image classification by means of an FCN into different semantic classes (sclera, iris, pupil, eyebrows and eyelashes, skin, hair, specular reflections and background regions). The experimental results showed that for iris segmentation, the use of information of different semantic classes of an eye image can help in the classification of iris and non-iris regions.

However, the application of this method directly in a video-based system is not feasible because the image classification is developed on a previously detected eye image, which leads to a considerable calculation time to classify the entire eye image, this undermines the performance of a system in real time.

3 Proposed framework

Our proposal is based on the combination of deep learning methods for the effective and efficient development of the stages of real-time detection of iris-pupil region, the evaluation of its quality and its semantic classification as a form of iris segmentation on video. Figure 2 shows the general outline of the proposed framework. The detection of the iris-pupil region is carried out directly on the video frames as well as the quality evaluation to subsequently classify the region into five classes (iris, pupil, sclera, eyelashes and specular reflections). The application of this procedure in a video-based iris recognition system can ensure its real-time performance while achieving accuracy similar to

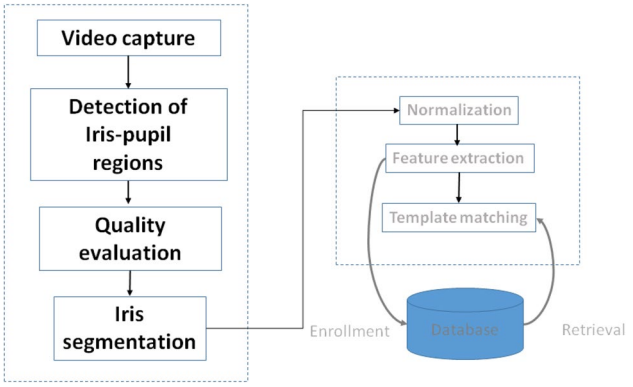

Fig. 2 General diagram of the proposed framework (left) and its interaction with the rest of the stages of a biometric recognition system (right)

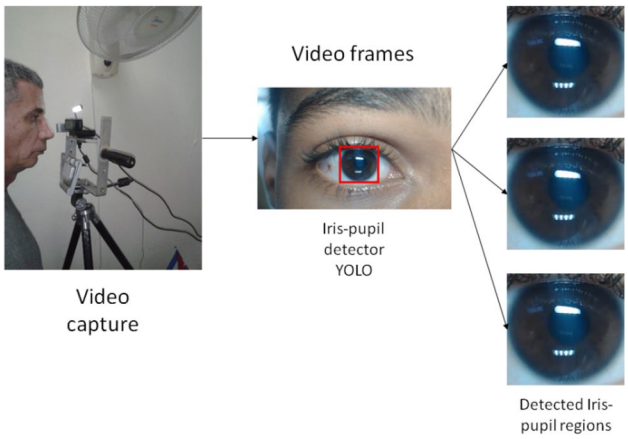

Fig.3 Diagram of the proposal for the capture and detection of the iris-pupil region in real time

image-based systems. Classification applied only to the iris-pupil image and not over the entire eye image will reduce computational time.

3.1 Real time capture and detection of iris-pupil region in video

In this section we present the design of the video capture device following the precepts established in Raja et al. (2015), we also present the algorithm for real-time detection of the iris-pupil region in video. Figure 3 shows the scheme of the proposal for capturing and detecting the iris-pupil region in video frames.

The design is based on the use of low cost materials, which allows, for example, the adaptation of face recognition systems to this biometry with minimal additional costs.

A review of the literature (Pavlov and Galeeva 2019), shows that the methods of the YOLO family are one of the most effective at the moment for object detection. Considering that our proposal is based on the use of video and its performance is required in real time, we thought that the use of the YOLO method in its tiny-Yolo version (Redmon and Farhadi 2017) could give us better results than the classic Viola and Jones (2004) algorithm.

For the direct detection of the iris-pupil region in the video frames, a detector was obtained based on the YOLO network in its tiny-Yolo version (Redmon and Farhadi, 2018). This method allows a fast estimation of the regions of interest.

In Bobrovsky et al. (2019) it is shown that this network can be applied to images of different resolution without the need to change its parameters. The training set consisted of 3300 images of the iris-pupil region taken from the MobBio (Monteiro et al. 2013) (800), UTIRIS (Hosseini et al. 2010) (800) databases and 1700 taken from the video frames of our own database (see 4.2). The training set was manually prepared by selecting the rectangular regions that enclose the iris-pupil region (Fig. 1b).

3.2 Real time image quality evaluation of the iris-pupil region in video

Focal length is one of the most important parameters considered in the standard, this parameter indicates the optimal distance between the subject and the sensor for a standard pixel density. Pixel density is defined in the standard as the quantitative sum of pixels that lie on the diagonal length of the iris image. The density regulated in the standard must be at least 200 pixels and must be composed of at least 2 lines of pixels per millimeter (2 lppmm). From the medical literature, it can be assumed that the iris represents 60% of iris-pupil region under a well-illuminated environment. Then the pixel density (*Ird*) can be calculated using the classical Pythagorean Theorem by Eq. 1 where w and h are the width and length (in pixels) of the iris-pupil images.

$$Ird = 0.6(\sqrt{w^2 + h^2}) \tag{1}$$

Of the variety of factors that negatively influence the quality of eye images, the degree of sharpness is one of the most important because, depending on whether the image is blurred or out of focus, it will lose the details of the iris texture, i.e., the iris internal structures which have identifying value.

In Garea-Llano et al. (2018) an analysis of the use of the Kang and Park filter was performed using eye images taken in the NIR. The filter consisting in a 5×5 convolutional kernel, was used to filter the high frequencies. It is composed of three functions of 5×5 and amplitude -1, one of 3×3 and amplitude $+5$ and four of 1×1 and amplitude -5. The kernel is capable of estimating the high frequencies

within the iris texture in the NIR better than other state-of-the-art operators, in addition to having a low computational time due to the reduced size of the kernel. In the case of the images in the VS, the effect of the kernel could be similar, so this procedure can be used in our proposal.

The entropy of an iris image has proven to be a good indicator of the amount of information contained in it (Waleed et al. 2018). Entropy only depends on the number of gray levels and their frequency.

Taking these elements into account and considering that the pixel density of the iris image is another of the fundamental elements for image quality, we propose its combination with the Kang and Park filter and the estimation of entropy to obtain a quality measure of the iris-pupil image (*Qiris*). The proposed *Qiris* measure is obtained by Eq. 2

$$Qiris = \frac{Ird * kpk * ent}{tird * tkpk * tent} \tag{2}$$

where *kpk,* is the average pixel value obtained as result of the convolution of the input iris-pupil image with the Kang and Park kernel.

tird is the threshold established by the standard for the minimum *Ird* to obtain a quality image.

tkpk is the estimated threshold of *kpk* to obtain a quality image. In Garea-Llano et al. (2018) the authors, based on their experimental results, recommend a threshold = 15.

ent is the value of the image entropy.

tent is the estimated threshold of *ent* with which it is possible to obtain a quality image.

The entropy of an iris image is calculated by Eq. 3, where the value of *Pi* is the probability of occurrence of a given pixel value within the iris region and *n* is the number of pixels in the image.

$$ent = -\sum_{i=1}^{n} p_i * log_2 p_i \tag{3}$$

For the definition of the entropy threshold (*tent*) we carried out an experiment that allowed us to establish that the iris-pupil images taken in the VS with a quality according to the international standard have an entropy ≥ 4.

For the experiment, we created a set of 300 images of the iris-pupil region manually cut from eye images and standardized to the size of 260×260. The images were selected from the MobBio (Monteiro et al. 2013) (150 images) and UBIRIS (Hosseini et al. 2010) (150 images) databases.

To determine the minimum entropy value for a quality image, each of them was evaluated using the pixel density parameters (*Ird*) and their response to the Kang and Park filter. The experimental results are shown in Fig. 4, they established that images with values greater than the corresponding thresholds (*tird* = 200; *tkpk* = 15) have an entropy greater than or equal to 4, so we can assume this value as the value of *tent.*

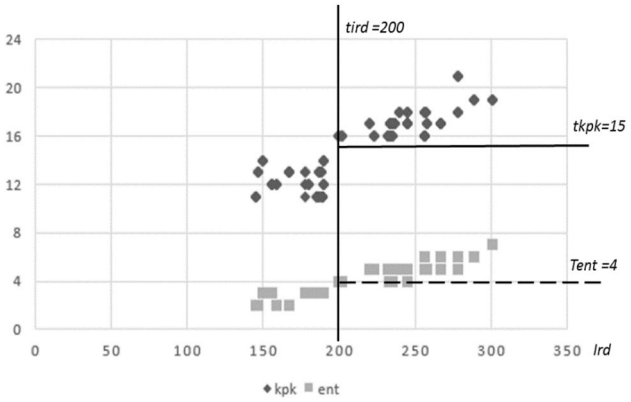

Fig. 4 Experimental results to determine the *tent*

The *Qiris* measure can reach values that depend on the thresholds selected for *Ird, kpk* and *ent*. Thus, considering the threshold values *tird* = 200 established by the standard, *tkpk* = 15 obtained experimentally in Garea-Llano et al. (2018) and *tent* = 4 experimentally obtained by us, the minimum *Qiris* value to obtain a quality iris-pupil image will be = 1, higher values denote higher quality images and values less than 1 denote lower quality images than that established by the standard.

3.3 Iris segmentation by semantic classification of the iris-pupil region

Figure 5 shows the general scheme of the proposed segmentation method. The scheme consists of two main parts. The semantic classification of the iris-pupil region and the iris segmentation.

3.3.1 Semantic classification of the iris-pupil region

The semantic classification of ocular images has shown a better result in iris segmentation avoiding or reducing the erroneous pixel classifications (Osorio et al. 2018). As result, to each region of the ocular image is assigned a label that denotes their anatomical class or the artifact (specular reflection, glasses).

In this work, we propose the use of a training scheme where a trained FCN was adapted for the semantic classification of the iris-pupil region through an inductive transfer learning process (Torrey and Shavlik 2009). We obtained a FCN fcn8s-at-once model, which was fine-tuned of the pre-trained model VGG-16 (Simonyan and Zisserman 2014). The fine tuning process allows us to select the most promising hypothesis space to adjust the objective knowledge. The implementation of the fcn8s-atonce architecture was carried out in the Caffe framework (Jia et al. 2014).

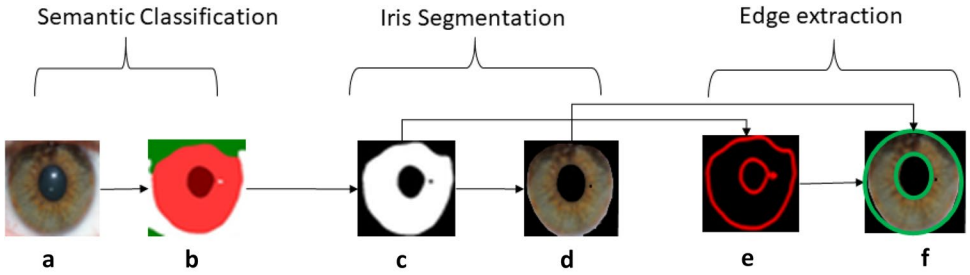

Fig. 5 General scheme of the proposed segmentation method

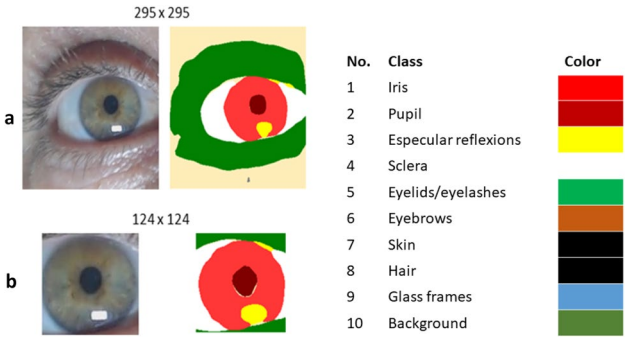

Fig. 6 Semantic classification. (**a**) Example taken from public annotations (Osorio et al. 22) (**b**) Sample for training

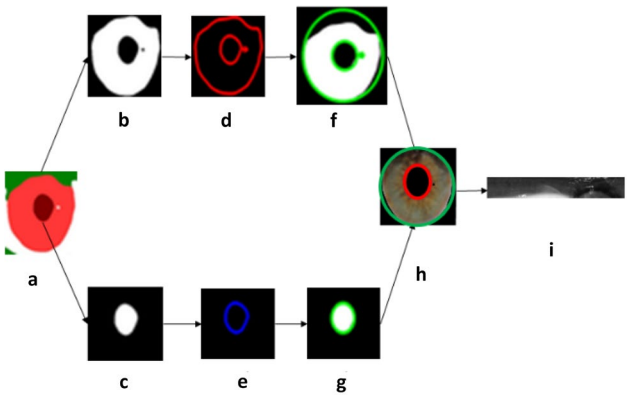

Fig. 7 Process for extraction of iris inner and outer edges

The intermediate upsampling layers were initialized by bilinear interpolation. The model output number was set according to the 5 possible classes that describe the iris-pupil region (Fig. 6a). Unlike the proposal of Osorio et al. (2018), the classification made on the image in our proposal is made only on the iris-pupil region (Fig. 6b) and not on the entire ocular image (Fig. 6a). We believe that this proposal significantly reduces the computation time compared to the proposal of Osorio et al. (2018) due to the smaller image size to be classified and the fewer number of classes to be considered.

For the training process we used a set consisting of 500 ocular ground truth images of semantic segmentation manually created and taken from the NICE I database (Proenc¸a et al. 2010). These public annotations (Osorio et al. 2018) were constructed at the pixel level where each pixel is labeled with its corresponding class. For the network training, the 500 images of this training set were manually cut, leaving us with only the iris-pupil region of each image as shown in Fig. 6b.

3.3.2 Iris segmentation

Once the classified image is obtained (Fig. 5b) the segmentation mask is calculated (Fig. 5c), which is obtained through a binarization process of the classified image, transforming the values of the pixels corresponding to the iris class to 1 and the rest of the classes (eyelids/eyelashes, specular reflections, pupil) to 0. To obtain the segmented image (Fig. 5d), a logical bitwise operation is performed between the image (Fig. 5a) and the mask (Fig. 5c), resulting in the segmented iris image. In this way, the segmented iris image is obtained free of the false characteristics of texture that can cause the occlusion of part of the iris by the eyelashes or specular reflections that remain after the quality evaluation process.

3.3.3 Edge extraction

Due to the need to normalize the iris texture using the Rubbersheet model (Daugman 2004), it is necessary to find the inner and outer edges of the iris in the segmented image. For this task we propose to combine the output of the semantic classification with an edge-based approach that allows detection of the center and radius coordinates of the iris and pupil regions.

Given the image of the iris-pupil region $S\,(x, y)$, S_k the obtained pixel classification using 5 classes such that $S_k\,(x, y) \in \{1, 2, 3, 4, 5\}$ are the k labels associated with the iris, pupil, eyelids/eyelashes and specular reflections respectively.

1. Obtain *Ik* and *Pk*: iris and pupil masks respectively for S*k*. (Figs. 7b, c)

$$I_k = \begin{cases} 1, if S_k(x,y) = iris \\ 0, in\ another\ case \end{cases} \quad (4)$$

$$P_k = \begin{cases} 1, if S_k(x,y) = pupil \\ 0, in\ another\ case \end{cases} \quad (5)$$

2. Obtain n candidate to iris (ICn) and pupil (PCn) contours applying canny filter. (Fig. 7d, e)
3. From the set of iris candidate contours ICn, obtain candidate circles using the circular Hough transform (Fig. 7f).
4. From the candidate circles found in the previous step, designate as iris the largest radius (Ci) (Fig. 7h).
5. From the set of pupil candidate contours PCn, obtain candidate circles using the circular Hough transform, as was done in step 3 (Fig. 7g).
6. From the candidate circles found in the previous step, discard those that are not contained in Ci, designate as pupil (Cp) the circle with largest radius (Fig. 7h).
7. Normalize the iris region from the coordinates of the Ci and Cp circles using the Rubbersheet model (Fig. 7i).

4 Experimental design

To validate our proposal, the experimental design was aimed at verifying the influence of the proposed framework in an iris recognition system by evaluating it using four referenced image databases and one own video database.

Three basic functionalities make up the experimental pipeline implemented in the c++language. The first two: Video Acquisition and Segmentation contain the proposed methods described in the previous sections. In the third, Feature Extraction and Comparison, for the purpose of the experiments in this work, we implemented the combination of two state-of-the-art feature extraction methods in order to verify the robustness of the proposed method regarding the use of different features for recognition, Scale-Invariant Feature Transform (SIFT) (Lowe 2004), and Uniform Local Binary Patterns (LBP) (Liao et al. 2007).

For comparison we use the similarity measures of each of these methods. The first estimates the similarity between two sets of SIFT key points (Lowe 2004). The second uses the Chi-square distance (Liao et al. 2007) which is a non-parametric test to measure the similarity between two images in a class. In both cases the minimum distance found will give the measure of maximum similarity between them.

4.1 Databases

MobBio (Monteiro et al. 2013) is a multi-biometric database that includes face, iris and voice of 105 volunteers. The iris subset contains 16 ocular images in the VS of each individual with a resolution of 250 X 200. The images were taken under two different lighting conditions, with different degrees of occlusion and iris orientation. These were captured with an Asus Transformer Pad TF300T Tablet at a distance of between 10 and 50 cm.

UTIRIS (Hosseini et al. 2010), is a hybrid database proposed by members of Teheran University. It has 1540 NIR and VS images of 79 people, separated into 154 classes. In the case of images under a VS, the capture was made with a Canon EOS 10D camera, the images were captured with 3MP and their resolution is 2048×1360.

NICE.I database is a subset of the UBIRIS v2 database (Proenc¸a. et al. 2010), whose images were acquired under non-ideal and non-cooperative conditions including images at-a-distance and on-the-move. It consists of 500 eye images for training and 445 eye images for testing. NICE.I images present problems of occlusions, specular reflections, off-angle and out-of-focus blur.

MICHE-I (De Marsico et al. 2015) is a challenging dataset captured with mobile devices in non-ideal difficult situations. This database consists of 3,191 images captured from 92 subjects collected using mobile devices: iPhone 5, Galaxy Samsung IV and Galaxy Tablet II (1,262, 1,297 and 632 images, respectively), with many different resolutions. For our experiments we used a subset of 344 images from the Galaxy Tablet II with resolution 640×480.

Our database (DatIris) is composed by 82 videos of 41 people taken in two 10-s sessions under the conditions and with the video acquisition device presented in Sect. 3. The camera used was a Logitech C920 HD Pro webcam, at a resolution of 920×1080. The videos were taken indoors. The database contains videos of light-skinned people of Caucasian origin, dark skinned people of African origin and mestizos, it contains videos of 26 men and 15 women, in an age range of 10 to 65 years (see examples in Fig. 8).

5 Experimental results and discussion

5.1 Evaluation of the proposed iris-pupil detection stage

The effectiveness of the proposed detection method was measured by the total degree of pixel-to-pixel coincidence ($Cdet$) between the images of the iris-pupil zone automatically detected in a 10-s video and their equivalents manually obtained (ground truth) and it is calculated by expression (6).

$$Cdet = \frac{\sum_{i=1}^{k} \frac{Idt(i)}{Igt(i)}}{k} \quad (6)$$

Fig. 8 Examples of the used databases

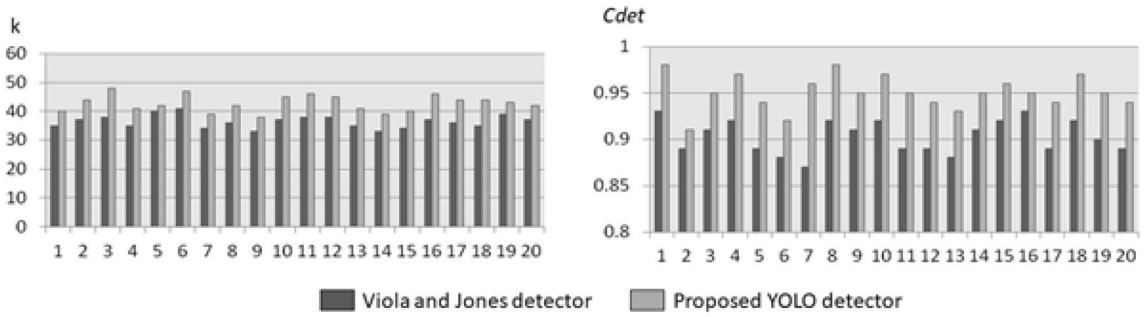

Fig. 9 Results of the evaluation of the YOLO-based detector

where, *Idt (i)* is the *i* automatically detected iris-pupil image in the video with $Qiris >= 1$; *Igt (i)* is the *i* ground truth image manually obtained from the corresponding video frame from where the image was automatically detected; *k* is the total number of true detections with $Qiris >= 1$ occurred in the video.

The performance of the proposed detector was compared with the performance of the detector proposed in Garea-Llano et al. (2019) based on algorithm of Viola and Jones (2004). The detector was tested on a set of 20 videos corresponding to 20 different people from our DatIris database (see 4.1).

Another parameter that was taken into account to assess the performance of the proposed detector was the estimation of the precision in the detection, it was computed by expression (7) commonly used to verify the effectiveness of machine learning systems.

$$Precision = \frac{TP}{(TP + FP)} \tag{7}$$

where, *TP* is the number of true (positive) detections and *FP* is the number of false (negative) detections made by the detector.

Figure 9 shows the results of the first part of the experiment. The results show that the proposed detector is capable of detecting a greater quantity of quality images by video compared to the detector based on Viola and Jones, similarly our proposal get a better accuracy than the Viola and Jones based detection.

Table 1 Results of the computed *Precision*

Detector	TP	FP	*Precision*
Viola and Jones detector (Garea et al. 2019)	728	25	0.967
Proposed YOLO detector	856	5	0.994

Table 1 shows the results of the computed *Precision* obtained by the proposed detector and its comparison with the detector based on Viola and Jones (Garea et al. 2019). These results show that, our detector achieves better *Precision* with fewer false detections.

The results of this experiment show the ability of the proposed YOLO-based detector to detect the iris-pupil region in real time by matching the times obtained by the detector based on the Viola and Jones algorithm. However, the proposed detector achieves a higher number of detections in the 10-s duration of the experienced videos. Additionally, these detections are of higher quality in terms of accuracy and precision with a low number of false detections.

5.2 Evaluation of the proposed iris segmentation method

In order to evaluate the performance of the proposed segmentation method two experiments were developed. The

Table 2 Experimental results of the segmentation stage on MobBio database

Osorio-Roig et al. (2018)		Garea-Llano et al. (2019)		Proposed method	
E^1%	t (s)	E^1%	t (s)	E^1%	t (s)
2.5	4	2.4	0.3	**1.7**	**0.3**

Values in bold highlight the results obtained by our proposal in comparison with those of the other two authors, which show that our proposal has a lower error and also is faster

Table 3 Experimental results of the segmentation stage on NICE I database

Segmentation method	E^1%
Liu et al. (2016)	0.9
Zhao and Kumar (2015)	1.21
Tan and Sun (2010)	1.31
Tan and Kumar (2013)	1.9
Osorio et al. (2017)	2.38
Osorio et al. (2018)	1.13
Chai et al. (2020)	0.7
Proposed method	**1.8**

The value in bold highlights the value of the error obtained by our proposal that, although it does not improve the state of the art, is comparable with this

first was oriented to compare de performance of our proposal with two semantic segmentation methods (Osorio et al. 2018; Garea et al. 2019) in terms of accuracy and efficiency in a set of quality images captured under controlled conditions. The second experiment was aimed to compare our results against others state of-the-art segmentation methods in terms of accuracy in a set of images acquired under non-ideal and non-cooperative conditions.

For the first experiment, the quality of the MobBio database was evaluated by the proposed quality index and 200 images with a $Qiris > 1$ were selected in order to guarantee the image quality simulating the projected work conditions of the proposed framework. This database provides manual annotations of the pupil and iris contours (ground truth).

For the second experiment, we used 500 images from the NICEI dataset without previous quality evaluation. To achieve comparability of segmentation accuracies, the comparisons were performed upon the bi-class (iris, non-iris) ground truth iris segmentation masks available for the NICE.I database.

As a metric for the evaluation we consider the E^1 measure proposed by the NICE I protocol (http://nice1.di.ubi.pt/). This metric estimates the proportion of the corresponding mismatched pixels.

To estimate the efficiency of the proposed method, the time (in seconds) to obtain a segmented iris image was calculated. It was computed using a PC with a 3.2 GHz Core i5-3470 processor and 8 GB of RAM.

Table 2 lists the average values of E^1 by Osorio et al. (2018), Garea-Llano et al. (2019) and our proposal. Table 3 shows the results of comparing the performance of our proposal with 7 state-of-the-art algorithms on the NICEI database.

The results of first experiment show a better performance of the proposed method in terms of accuracy and a calculation time for the segmentation process similar to Garea-Llano et al. (2019) and less than Osorio et al. (2018). These results show that reducing the size of the image to be processed (image of the iris-pupil region and not the complete eye image), increases the efficacy of the segmentation process when the images have a $Qiris > 1$. This is due in part to the better accuracy in detection achieved by the detector based on YOLO and to the fewer classes that the CNN should process. Then, the proposed segmentation process reduces the computation time with respect to Osorio et al. (2018) in more than 3 s, maintaining a time similar to that achieved by Garea et al. (2019), but keeping the same efficacy. These results demonstrate that it is possible to use the proposed segmentation method in the working conditions of the proposed framework for real-time video.

The results of the second experiment show that despite the conditions of the NICE I database, the efficacy of the proposed method is comparable to the state-of-the-art methods under these conditions, which indicates that it could be used in systems whose working conditions are less controlled. This topic would be part of the future research to be carried out.

5.3 Evaluation of the proposed image quality measure and its effect on recognition process

The experiment was developed to check the effect on the verification task of the proposed image quality method on the subsequent steps involved in the process of recognizing people in a biometric system.

The accuracy of the verification was measured using false rejection rate (FRR) at a false acceptance rate (FAR) $<= 0.001\%$. Comparisons were made in a form of all against all to obtain the genuine and impostor distributions. The experiment was developed on four databases (MobBio, UTIRIS, MICHEI and DatIris).

For the DatIris database, the 41 videos from Session 1 were processed by taking two images of each iris and comparing them against a set consisting of two images of each iris taken from the frames of the videos in Session 2. Table 4 shows the comparison of the FRR error at FAR $<= 0.001\%$ obtained by Osorio et al. (2018), Garea-Llano et al. (2019)

Table 4 Comparison of the FRR obtained in the verification process

Datasets	*Qiris*	% of images *	Osorio et al. (2018)		Garea-Llano et al. (2019)		Our proposal	
			SIFT	LBP	SIFT	LBP	SIFT	LBP
MobBio	–	100	0.36	0.28	0.29	0.27	0.28	0.26
	< 1.0	73.7	0.28	0.26	0.27	0.25	0.27	0.25
	≥ 1.0	26.3	0.24	0.23	0.25	0.22	0.23	0.21
UTIRIS	–	100	0.05	0.06	0.08	0.05	0.06	0.04
	< 1.0	20.5	0.06	0.05	0.07	0.06	0.07	0.05
	≥ 1.0	79.5	0.07	0.04	0.06	0.04	0.05	0.03
MICHEI	–	100	0.38	0.37	0.38	0.34	0.37	0.32
	< 1.0	66	0.37	0.33	0.35	0.30	0.35	0.30
	≥ 1.0	34	0.27	0.25	0.26	0.24	0.25	0.23
DatIris	–	100	0.04	0.03	0.07	0.04	0.04	0.03
	< 1.0	10	0.03	0.02	0.04	0.03	0.03	0.03
	≥ 1.0	90	0.02	0.01	0.02	0.01	0.02	**0.009**

The value in bold highlights the best result obtained by our proposal over the rest of the results

*Percentage of images processed with quality values specified in the *Qiris* column

Table 5 Comparison of the quality proposed measure (*Qiris*)

Datasets	% imag proc. **		*FRR Qindex*	*FRR Qiris*
	Qindex	Qiris		
MobBio	19.5	26.3	0.23	**0.21**
UTIRIS	71.8	79.5	0.04	**0.03**
MICHEI	73.6	66	0.25	**0.23**
DatIris	88	90	0.02	**0.009**

Values in bold highlight the fact that all the results obtained using the proposed quality index (Qiris) are better than the results obtained with the quality index with which it was compared (Qindex)

**Percentage of images processed with quality values > = 1

and our proposal on the tested databases, without quality evaluation and taking *Qiris* > *1.0* and *Qiris* < *1.0* intervals to accept or reject the iris-pupil images to be processed.

The results show that by increasing the *Qiris* value, the system supports higher quality images and rejects low quality images. This increase in quality results in a decrease in the FRR, with a significant result in UTIRIS where a FRR = 0.03 is achieved with 79% of the database and a *Qiris* > 1 using LBP features.

However in the MobBio and MICHE I databases, increasing the *Qiris* threshold results in a significant decrease in the number of images to be compared. This is because these databases were taken in less restrictive and more challenging environments in order to test and validate new algorithms developed for these conditions.

The results obtained in the DatIris database show a high performance of the verification process obtaining a FRR = 0.009, which corroborates the relevance of the quality index proposed for the projected work conditions.

It is also observed that the FRR levels become very close for the three evaluated methods, which shows that the decrease of the classification area from the whole eye image in Osorio-Roig et al. (2018) to the iris-pupil region in Garea-Llano et al. (2019) and in our proposal does not affect the accuracy of the recognition process, and a slight increase was observed in the case of our proposal, which may be related to the increase in accuracy experimented in the proposed detection process based on YOLO.

Table 5 shows the results of the comparison of the quality measure (*Qindex*) proposed in Garea-Llano et al. (2018) with our proposal (*Qiris*), at values > = 1 using the LBP features and the proposed segmentation method. It can be seen that starting from the inclusion of the entropy of the iris image; the system considers a higher percentage of images to be processed while maintaining a similar level of recognition accuracy.

The results corroborate that the proposed quality index is capable of decanting those images not suitable for recognition under the projected conditions of border and access controls. On the other hand, the simplicity of the calculations involved in the *Qiris* computation allows the evaluation of the images in real time, which is essential for video work. The conditions in which the experiment was developed corroborate that the proposal has a similar behavior under different capture conditions (sensors and environmental conditions).

6 Conclusions

In this work we presented a framework for real time iris detection and segmentation in RGB video frames under controlled conditions (conditions of border and access controls, where people collaborate in the recognition process) as part of a biometric recognition system.

The proposed framework integrates three main contributions: An iris-pupil region detector based on the YOLO network that increases the accuracy and speed of the detection and allows its use in video capture. The experimental results shown the ability of the proposed YOLO-based detector to detect the iris-pupil region in real time reaching a high number of quality detections in terms of accuracy and precision with a low number of false positives.

The proposal of a stage and measure (*Qiris*) for image quality evaluation reduce the input to the recognition system of low quality images which leads to an increase in the recognition effectiveness. In addition, the simplicity of the calculations involved in the *Qiris* computation allows the evaluation of the images in real time, which is essential for video work. The experimental results on four different datasets corroborate that the proposal has a similar behavior under different capture conditions.

The proposal of an iris segmentation method based on semantic classification of the iris-pupil region by a CNN, produces segmented iris images free of the false texture characteristics that remain after the quality evaluation process. The experimental results show a better performance of the proposed method in terms of accuracy for the segmentation process when images have a $Qiris >= 1$. The proposed segmentation process reduces the computation time with respect to others methods, demonstrating that it is possible to use in the working conditions of the proposed framework. On the other hand the proposed method is comparable to the state-of-the-art methods, which indicates that it could be used in systems whose working conditions are less controlled. This topic would be part of our future works.

Acknowledgements This work is an extension and continuity of the results presented by us at the 24th Iberoamerican Congress on Pattern Recognition (CIARP 2019): Garea-Llano E., Morales-González A., Osorio-Roig Video Iris Recognition Based on Iris Image Quality Evaluation and Semantic Classification, Proceedings. Lecture Notes in Computer Science 11896, Springer 2019, ISBN 978-3-030-33903-6. We want to thank the third author of this cited work, Daile Osorio-Roig for her contribution to the conception of the idea that gave rise to the development of the results presented here.

References

Agarwal S, Du Terrail JO, Jurie F (2019) Recent advances in object detection in the age of deep convolutional neural networks. ffhal-01869779v2f

Bansal A (2020) Iris recognition system: a review. Int Res J Eng Technol (IRJET) 07:05

Bobrovsky A, Galeeva M, Morozov A, Pavlov V, Tsytsulin A (2019) Automatic detection of objects on star sky images by using the convolutional neural network. J Phys: Conf Ser 1236:1066

Chai TY, Goi BM, Hong YY (2020) End-to-end automated iris segmentation framework using U-net convolutional neural network. Lecture Notes in Electrical Engineering, vol 621. Springer, Singapore

Daugman J (2004) How iris recognition works. IEEE Trans Circ Syst Video Technol 14(1):21–30

Daugman J, Downing C (2016) Iris image quality metrics with veto power and nonlinear importance tailoring. In: Rathgeb C, Busch C (eds) Iris and periocular biometric recognition. IET Publisher, pp 83–100

De Marsico M, Nappi M, Riccio D, Wechsler H (2015) Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols. Pattern Recogn Lett 57:17–23

Garea-Llano E, Morales-González A, Osorio-Roig D (2019) Video iris recognition based on iris image quality evaluation and semantic classification.CIARP2019. LNCS 11896:198–208

Garea-Llano E, Osorio-Roig D, Hernández O (2018) Image quality evaluation for video iris recognition in the visible spectrum. Biosensors and bioelectronics open access (ISSN: 2577–2260)

Hollingsworth K, Peters T, Bowyer K (2009) Iris recognition using signal-level fusion of frames from video. IEEE Trans Inf Forens Secur 2009(4):837–848

Hosseini MS, Araabi BN, Soltanian-Zadeh H, Pigment M (2010) Pattern for iris recognition. IEEE Trans Instr Meas 2010(59):792–804

Jalilian E, Uhl A (2017) Iris segmentation using fully convolutional encoder–decoder networks. Deep learning for biometrics. Springer, Berlin, pp 133–155

Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: Convolutional architecture for fast feature embedding. In Proc. of the 22nd ACMMM2014, pp. 675–678

Liao S, Zhu X, Lei Z, Zhang L, Li SZ (2007) Learning multi-scale block local binary patterns for face recognition. In Proc. ICB'07, pp 828–837

Liu W et al (2016) SSD: single shot multibox detector. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision—ECCV 2016, LNCS, vol 9905. Springer, Berlin

Liu N, Li H, Zhang M, Liu J, Sun Z, Tan T (2016) Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In Proc. Int'l Conf. on Biometrics (ICB'16). IEEE, 2016, pp. 1–8.

Lowe DG (2004) Distinctive image features from scale-invariant key points. Int J Comput Vision 2004(60):91–110

Minhaz M, Rahman U, Hasan M, Taief AM (2021) A new framework for video-based frequent iris movement analysis towards anomaly observer detection. Int J Image Graph Signal Process 13(1):13–27

Monteiro C, Oliveira HP, Rebelo A, Sequeira AF, Mobbio A (2013) 1st biometric recognition with portable devices competition. Available in: https://paginas.fe.up.pt/~mobbio2013/.

Naz S, Ziauddin S, Shahid A (2019) Driver fatigue detection using mean intensity, SVM, and SIFT. Int J Interact Multimed Artif Intell 5(4):86–93

Osorio-Roig D, Morales-Gonzalez A, Garea-Llano E (2017) Semantic segmentation of color eye images for improving iris segmentation. In: Proc.CIARP'17. LNCS. Springer, pp. 466–474.

Osorio-Roig D, Rathgeb C, Gomez-Barrero M, Morales-González Quevedo A, Garea-Llano E, Busch C(2018) Visible wavelength iris segmentation: a multi-class approach using fully convolutional neuronal networks. BIOSIG 2018, IEEE.

Pavlov VA, Galeeva MA (2019) Detection and recognition of objects on aerial photographs using convolutional neural networks. J Phys Conf Ser 1326:012038

Proenca H (2016) Unconstrained iris recognition in visible wavelengths. In: Bowyer WK, Burge JM (eds) Handbook of iris recognition, 2nd edn. Springer, London, pp 321–358

Proenca H, Alexandre LA (2010) Iris recognition: analysis of the error rates regarding the accuracy of the segmentation stage. Image vis Comput 28(1):202–206

Proenca H, Alexandre L (2012) Toward covert iris biometric recognition: experimental results from the nice contests. IEEE Trans Inf Forens Secur 7(2):798–808

Proenca H, Filipe S, Santos R, Oliveira J, Alexandre LA (2010) The ubiris.v2: A database of visible wavelength iris images captured on-themove and at-a-distance. IEEE Trans Pattern Anal Mach Intell 32(8):1529–1535

Raja KB, Raghavendra R, Vemuri VK, Busch C (2015) Smartphone based visible iris recognition using deep sparse filtering. Pattern Recogn Lett 2015(57):33–42

Raja KB, Raghavendra R, Busch C (2015) Iris imaging in visible spectrum using white LED. In proc. BTAS 2015, IEEE.

Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 7263–7271.

Redmon J, Farhadi A (2018) YOLOV3: An incremental improvement,'' 2018, arXiv: 1804.02767. [Online]. Available: https://arxiv.org/abs/1804.02767

Redmon J, Divvala S, Girshick R, Farhad A (2016) You only look once: unified, real-time object detection. In 2016 IEEE Conference on computer vision and pattern recognition (CVPR), Las Vegas, pp. 779–788

Sankowski W, Grabowski K, Napieralska M, Zubert M, Napieralski A (2010) Reliable algorithm for iris segmentation in eye image. Image vis Comput 28(2):231–237

Schmid N, Zuo J, Nicolo F, Wechsler H (2016) Iris quality metrics for adaptive authentication. In: Bowyer KW, Burge MJ (eds) Handbook of iris recognition, 2nd edn. Springer, London, pp 101–118

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv: 1409.1556

Tan C-W, Kumar A (2013) (2013), Towards online iris and periocular recognition under relaxed imaging constraints. IEEE Trans Image Processing 22(10):3751–3765

Tan T, He Z, Sun Z (2010) (2010), Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. Image vis Comput 28(2):223–230

Torrey L, Shavlik J (2009) Transfer learning. Handbook of research on machine learning applications. Springer, Berlin

Verma S, Girdhar A, Jha RRK (2018) Real-time eye detection method for driver assistance system. Advances in intelligent systems and computing, vol 696. Springer, Singapore

Viola P, Jones M (2004) Rapid object detection using a boosted cascade of simple features. Mitsubishi Electric Research Laboratories Inc, Cambridge

Waleed SA, Fathy A, Ali HS (2018) Entropy with local binary patterns for efficient iris liveness detection. Wirel Person Commun 102(3):2331–2344

Zhao Z, Kumar A (2015) An accurate iris segmentation framework under relaxed imaging constraints using total variation model. In Proc. IEEE Int'l Conf. on Computer Vision (ICCV'15), 2015, pp. 3828–3836.

Zuo J, Schmid NA (2008) An automatic algorithm for evaluating the precision of iris segmentation. In: BTAS'08, Washington, DC, USA

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.