

# **Método de Reconocimiento Biométrico por el Iris en Video, basado en Aprendizaje Profundo y la Evaluación de la Calidad de la región Iris-pupila**

Eduardo Garea-Llano<sup>1</sup>, Annette Morales-Gonzalez<sup>2</sup>

<sup>1</sup> Neuroinformatics Department, Cuban Neuroscience Center, Havana, Cuba, <sup>2</sup> Advanced Technologies Application Center, Havana, Cuba

eduardo.garea@cneuro.edu.cu, amorales@cenatav.co.cu

## **Resumen:**

En el escenario mundial actual donde la influencia de la pandemia de COVID19 ha alcanzado proporciones universales afectando a la casi totalidad de los países se hace necesario contar con un método de identificación biométrica que permita el reconocimiento seguro de las personas en frontera, los controles de cuarentena, el acceso a locales de alto riesgo biológico, entre otros. Las características de estos escenarios exigen que las personas deban estar con el rostro cubierto por mascarillas sanitarias (nasobuco), las manos protegidas por guantes o que el contacto directo con objetos sea el menor posible como puede ser el caso de los sensores para la captación de huellas dactilares o palmares. En esta circunstancia el reconocimiento biométrico por el iris ha cobrado vital importancia debido a que precisamente esta modalidad evita todos los inconvenientes antes mencionados con una demostrada alta eficacia. El gran inconveniente de esta modalidad biométrica ha sido la necesidad del uso de cámaras en el espectro infrarrojo y de otro tipo de hardware complementario que para aplicaciones de amplia distribución resulta caro. En este sentido el uso de video tomado por sensores en el espectro visible para aplicaciones biométricas ha alcanzado un gran auge en los últimos 5 años, la biometría del iris no ha estado exenta de ello por lo que el uso de cámaras de video y sensores acoplados a dispositivos móviles se ha incrementado y han conllevado a un amplio espectro de aplicaciones que utilizan esta modalidad biométrica.

En este trabajo se presenta un método para el reconocimiento biométrico del iris en las condiciones mencionadas. Nuestra propuesta se centra en las etapas de captación y segmentación del iris y se basa en la combinación de métodos de aprendizaje profundo para la detección directa de la región del iris-pupila utilizando la red YOLO, la evaluación de su calidad y la clasificación semántica de las regiones que la componen por una Red Convolutiva Completa (Fully Convolutional Network) para realizar la segmentación del iris. Para la evaluación de la calidad se propone el uso de una medida que combina parámetros definidos en el ISO / IEC 19794-6 2005 y otros derivados de la sistematización del conocimiento presente en la literatura especializada. Los experimentos desarrollados sobre bases de datos de referencia y una base de datos propia demuestran la efectividad de la propuesta y la pertinencia de la misma en las actuales circunstancias tanto en eficacia como en eficiencia.

## **Introducción**

Desde sus comienzos en la década de los noventa del pasado siglo XX, el proceso de captura de las imágenes de iris para el reconocimiento biométrico ha estado caracterizado por el uso de sensores infrarrojos (NIR). El uso de este tipo de sensores ha demostrado su efectividad ante imágenes de iris con bajo contraste entre el iris y la pupila como puede ser el caso de personas que tienen colores oscuros de iris por el alto contenido de melanina, sobre todo aquellas personas de origen africano o mestizo.

Sin embargo esta característica que en determinadas circunstancias puede ser una ventaja, en otras se convierte en una limitante por el relativo alto costo del hardware necesario para implementar una solución práctica. Por ejemplo para la identificación de personas en frontera generalmente se utilizan aplicaciones de reconocimiento biométrico por el rostro, las que para la captación de las imágenes utilizan cámaras que captan imágenes o video en el espectro visible (VS), generalmente se utilizan para ello cámaras web o cámaras de video IP, por lo que en las actuales circunstancias un cambio de método de identificación biométrica facial hacia uno basado en iris en IR resultaría muy costoso. Por otro lado, en años recientes se ha

producido un incremento sustancial de aplicaciones biométricas basadas en el uso de este tipo de sensores (VS) integrados a dispositivos móviles. La biometría del iris no ha estado al margen de este desarrollo con algunas aplicaciones de este tipo [1]. En este contexto el video en el VS como modo de captación del iris en aplicaciones en tiempo real es un aspecto que adquiere gran importancia [2] [3]. Por todo ello se hace necesario dirigir las investigaciones hacia este tipo de captación de imágenes de iris debido a los problemas que la caracterizan.

En este mismo sentido el reconocimiento del iris en el VS representa uno de los mayores retos. Algunos trabajos recientes [4] confirman que si las imágenes de iris en el VS son capturadas de forma apropiada estas pueden exhibir un nivel de detalle en su textura similar a aquellas capturadas en el NIR, mientras que generalmente el volumen de información que estas tienen es menor. En su lugar, dentro de las imágenes en el VS pueden aparecer diferentes ruidos, como reflexiones especulares y sombras lo que conlleva a un incremento de las variaciones intra-clase.

En este trabajo se presenta un método para el reconocimiento biométrico del iris en video tomado en el VS. Nuestra propuesta se centra en las etapas de captación y segmentación del iris y se basa en la combinación de métodos de aprendizaje profundo para la detección directa de la región de iris-pupila (YOLO), la evaluación de su calidad y la clasificación semántica de las regiones que la componen por una Red Convolutiva Completa (FCN, por sus siglas en inglés) para realizar la segmentación del iris. Para la evaluación de la calidad se propone el uso de una medida de calidad que combina parámetros definidos en el ISO / IEC 19794-6 2005 [16] y otros derivados de la sistematización del conocimiento presente en la literatura especializada. Esta propuesta asegura que las imágenes de iris obtenidas tengan un bajo porcentaje de artefactos inherentes a la calidad de la imagen que influyan de forma negativa en el posterior proceso de reconocimiento biométrico y evita otros problemas relacionados con la captación en tiempo real y la no cooperación de las personas a ser identificadas como pueden ser los ojos cerrados y la mirada fuera de ángulo. El trabajo se estructura de la siguiente forma: En la sección 2 se discuten los trabajos relacionados, la sección 3 presenta la propuesta del método y la sección 4 presenta el diseño experimental, sus resultados y discusión. Finalmente se presentan las conclusiones de esta investigación.

## **2. Trabajos Relacionados**

### **2.1. Captación y detección del iris dentro en la escena**

En [5] se propone el uso de luz blanca LED como una forma de atenuar el efecto del bajo contraste entre los iris de color oscuro y las pupilas en imágenes capturadas por dispositivos móviles en el VS. En este trabajo queda demostrado que esta propuesta incrementa las tasas de reconocimiento hasta un nivel similar obtenido con imágenes en el NIR.

En [6] se presenta la detección directa de la región iris-pupila en escenas de video mediante un detector basado en el algoritmo clásico de Viola and Jones [7]. Este detector aprovecha las características de la representación integral de la imagen para lograr un cálculo rápido de los rasgos utilizados por el detector. El algoritmo de aprendizaje basado en Adaboost, permite la selección de un pequeño número de rasgos del conjunto inicial y obtener una cascada de clasificadores simples para detectar la región iris-pupila dentro de una escena donde predomina la imagen de ojo.

El principal problema de este enfoque es la relativamente baja exactitud en el proceso de detección de las fronteras exteriores de la región del iris-pupila y la velocidad de detección en video que propicia que se pierdan cuadros del video en los que el detector no puede encontrar presencia de las regiones de iris-pupila. Otro problema está asociado con la propia naturaleza del detector, el cual depende en gran medida de la naturaleza de las muestras de entrenamiento positivas y negativas utilizadas en el proceso de entrenamiento, lo que conlleva a una baja capacidad de extrapolación, de forma que cuando se cambian las condiciones de captura de las imágenes las tasas de detección disminuyen significativamente.

Por otro lado, el problema de la detección de objetos es un tema clave en la visión por computadoras. Este tema tiene dos líneas fundamentales de investigación y de retos a enfrentar [8], la detección de los objetos en la imagen y su posición y la estimación de la clase a la cual pertenece (clasificación). En los últimos años han sido propuestos algunos métodos basados en Redes Neuronales Convolucionales (CNN) para atacar este problema que han tenido un éxito significativo. Según [8] estos métodos pueden ser divididos en dos categorías:

métodos de un solo estado, los que proporcionan una estimación de la posición y la clase a la que pertenece el objeto en un solo paso y los métodos de dos estados, los que primero detectan regiones de la imagen donde puede estar presente el objeto y luego aplican estas regiones a un clasificador.

Para el caso de la detección de la región iris-pupila en video los métodos que más se adaptan a la naturaleza de esta tarea son los de la primera categoría. El primer tipo de estos métodos es el de mayor velocidad y a su vez tienen una exactitud similar a los del segundo tipo [9]

Entre ellos los que mejores resultados ofrecen son el Single-Shot Detector (SSD) [10] y el You Only Look Once (YOLO) [11]. Ambos métodos proporcionan como salidas la probabilidad de cada posible clase, no para toda la imagen, sino para una serie de posiciones regularmente espaciadas para diferentes escalas de rectángulos llamados cajas.

Una revisión de la literatura [9] demuestra que los métodos de la familia YOLO son los más efectivos en estos momentos. Considerando que nuestra aplicación se basa en el uso del video para la captación de las imágenes de iris por lo que se requiere de su desempeño en tiempo real pensamos que el uso del método YOLO [11] pudiese darnos mejores resultados que el algoritmo clásico de Viola y Jones [7].

## 2.2. Evaluación de la calidad de la imagen.

La estimación de la calidad de las imágenes es uno de los temas que ha tomado gran importancia en la biometría. Las medidas de calidad son utilizadas para decidir cuando la imagen debe ser desechada o procesada por los sistemas de reconocimiento. En el caso del iris algunos trabajos han sido presentados en este sentido [12][13]. En [12] la calidad de las imágenes de iris se determina por muchos factores que dependen de las condiciones ambientales, las características del sensor y de la persona a ser reconocida. Algunas de las medidas de calidad reportadas en la literatura [13] se centran en la evaluación de la calidad después del proceso de segmentación del iris lo que trae como consecuencia la saturación del sistema con un número significativo de imágenes de baja calidad. Otro de los problemas de las medidas de calidad propuesta para el iris es que solo evalúan un solo parámetro o un conjunto pequeño de ellos [3], como la borrosidad por estar las imágenes fuera de foco, borrosidad por movimiento y algunos tipos de oclusiones.

Otros autores [13][14] utilizan un conjunto mayor de factores que incluyen además de las ya mencionadas, a las reflexiones especulares, los problemas de iluminación y la mirada fuera de ángulo. Estos autores consideran que si se produce al menos la degradación de uno de los parámetros estimados por debajo de los umbrales establecidos la medida que los integra automáticamente toma valor cero ("veto power"). Esto para un sistema basado en video puede ser un inconveniente debido a que no siempre las condiciones de captura del video serán las óptimas y una medida de calidad de esta índole puede limitar el paso de imágenes que si pueden tener valor identificativo.

En [6] y [15] se propusieron medidas de calidad que estiman la calidad de la imagen dentro de la imagen completa del ojo capturada por una cámara de video en el VS. Estas medidas consideran parámetros establecidos en el estándar ISO / IEC 19794-6: 2005 [16] y los combinan con métodos de estimación de la nitidez de la imagen. Sin embargo, la estimación de la calidad se desarrolla en toda la imagen del ojo y no en la región específica de iris-pupila (ver figura 1). Adicionalmente, el nivel de nitidez y diversidad de la textura del iris no es considerado.

En este trabajo presentamos una medida de calidad que considera los parámetros establecidos en el estándar [16] combinada con la estimación de la nitidez de la imagen y la estimación del nivel de diversidad de la textura del iris a través de la entropía de la imagen. En esta propuesta a diferencia de las propuestas anteriores la estimación de la calidad se realiza no sobre la imagen de ojo, sino sobre la imagen de iris-pupila previamente detectada a partir de la propuesta que presentamos en la sección 3.2.



Figura.1. Región del ojo (a); región de iris-pupila (b)

## 2.3. Segmentación del iris

La llamada segmentación del iris involucra la detección de las fronteras exteriores e interiores del iris en una imagen de ojo (fig.1a). Esta tarea también puede incluir la detección de las cejas, la exclusión de las pestañas así como los bordes de los lentes de contacto y las reflexiones especulares [17]. La segmentación exacta de la región del iris en imágenes tomadas en el VS es aún una tarea compleja debido a la gran cantidad de factores que influyen en la calidad de las imágenes [18].

En la primera parte del Noisy Iris Challenge Evaluation (NICE.I) [19] las versiones optimizadas del operador integrodiferencial propuesto por Daugman [20], [21], estuvieron entre los métodos mejor posicionados. Trabajos más recientes demuestran que la clasificación de la imagen en dos clases iris y no-iris basada en conceptos de máquinas de aprendizaje mejoran aún más la exactitud en la segmentación [22].

En los últimos años los trabajos basados en el uso Redes Convolucionales Completas (FCNs) han alcanzado los mejores resultados [22], [23]. Jalilian y Uhl [23] exploraron el desempeño de tres tipos de FCNs en la tarea de segmentación del iris, las redes estudiadas presentaron dificultades ante imágenes que presentaban problemas de ángulo de visión y que fueron tomadas desde varias distancias con diferentes tipos de oclusiones (incluyendo gafas). La principal limitante de estos métodos está en que ellos solo toman en cuenta la existencia de dos clases (iris-no iris).

En [22] Liu et al. Los resultados fueron pobres para una serie de ejemplos de imágenes de personas de piel oscura o imágenes que no contenían ojos. De acuerdo con lo expresado por los autores esto se debe a la limitada cantidad de ejemplos de entrenamiento en la base de datos NICE I.

Para enfrentar estos problemas en [24] se presenta un algoritmo de segmentación semántica para extraer la región del iris. El algoritmo basado en HMRF-PyrSeg [24], fue entrenado con 53 imágenes de ojo anotadas manualmente y tomadas del set de entrenamiento de la base de datos NICE I. Aunque los resultados experimentales no mejoraron a los métodos de segmentación del estado del arte, los autores demostraron que la idea de la segmentación multi-clase era un camino prometedor para la segmentación del iris en imágenes VS.

En [25] se presenta un método de segmentación multiclase del iris. Este método se basa en la clasificación por medio de una FCN de la imagen en diferentes clases semánticas presentes en una imagen de ojo (esclera, iris, pupila, cejas y pestañas, piel, cabello, reflexiones especulares y regiones del fondo). Los resultados experimentales mostraron que para la segmentación del iris, el uso de la información de las diferentes clases semánticas presentes en una imagen de ojo es mejor que su clasificación en iris-no iris.

Sin embargo la aplicación de este método de forma directa en un sistema basado en video no es factible debido a que la clasificación de la imagen se desarrolla sobre una imagen de ojo previamente detectada, lo que conlleva a un considerable tiempo de cálculo para clasificar toda la imagen, esto atenta contra el desempeño de un sistema en tiempo real.

### **3. Método propuesto**

La figura 2 muestra el esquema general del método propuesto. La propuesta se basa en la combinación de métodos de aprendizaje profundo para el desarrollo eficaz y eficiente de las etapas de detección en tiempo real de la región iris-pupila, la evaluación de la calidad de la misma y su clasificación semántica como forma de segmentación del iris en video. La detección de la región iris-pupila se realiza directamente sobre los cuadros del video al igual que la evaluación de la calidad para seguidamente realizar la clasificación de la región en cinco clases (iris, pupila, esclera, pestañas y reflexiones especulares). La aplicación de este procedimiento en un sistema de reconocimiento de iris basado en video puede asegurar su trabajo en tiempo real a la par de alcanzar exactitudes similares a la de los sistemas basados en imágenes. La clasificación aplicada solamente a la imagen de la región iris-pupila y no sobre toda la imagen de ojo reducirá el tiempo computacional.

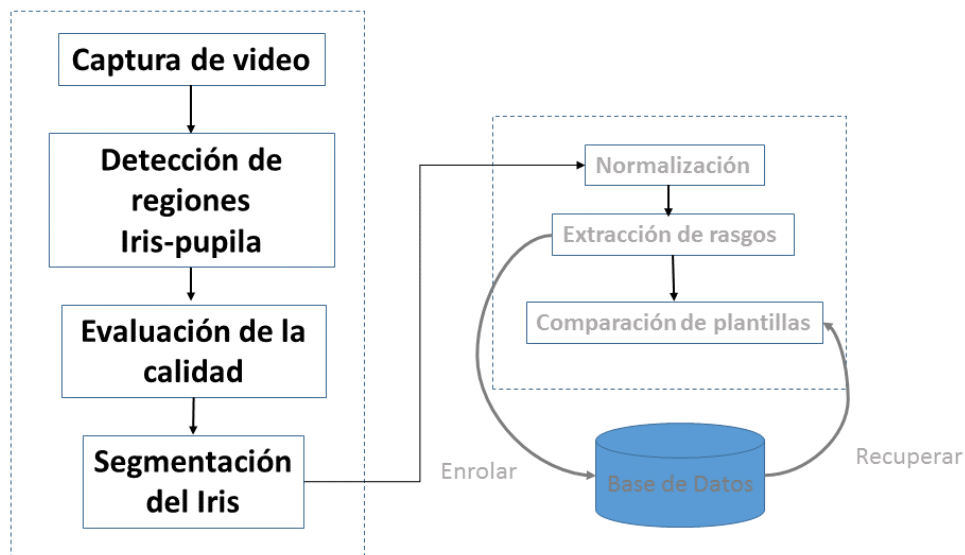


Figura.2. Esquema general del método propuesto

### 3.1. Captación y detección de la región iris-pupila en tiempo real en el video

En esta sección presentamos el diseño del dispositivo de captura de video siguiendo los preceptos establecidos en [5], presentamos además el algoritmo para la detección en tiempo real de la región iris-pupila en video. La figura 3 muestra el esquema de la propuesta para la captación y detección de la región iris-pupila en tiempo real a partir de los cuadros del video.

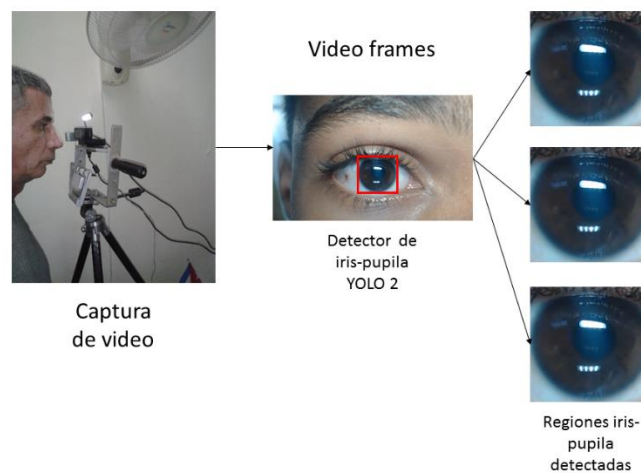


Figura.3. Esquema de la propuesta para la captación y detección de la región iris-pupila en tiempo real

#### ***Diseño de dispositivo de captura del video***

Tomando en cuenta la experiencia descrita en [5] realizamos la propuesta de un diseño de dispositivo de captura. El diseño se basa en el uso de materiales de fácil adquisición y muy bajo costo lo que permite, por ejemplo, la adaptación de sistemas de reconocimiento de rostros a esta biometría con un mínimo de gastos adicionales. La figura 4 muestra las componentes del dispositivo de captura experimental obtenido y como estas se integran.

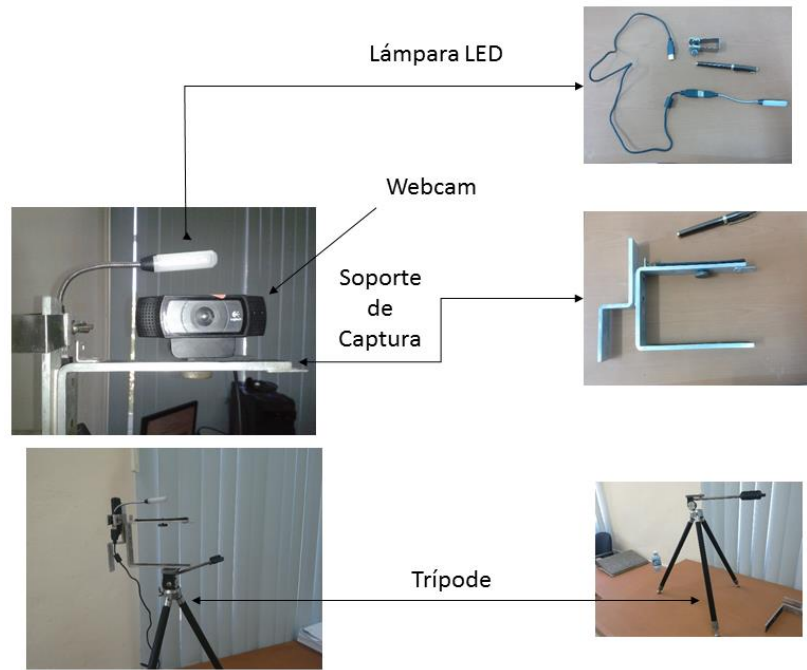


Figura 4. Componentes e integración del dispositivo de captura

### ***Detección de la región iris-pupila en tiempo real en el video***

Para la detección directa de la región iris-pupila en los frames de video se obtuvo un detector basado en técnicas de aprendizaje profundo sobre la base de la red YOLO [11]. Este método permite la estimación rápida de las regiones de interés.

Para la extracción de los rasgos de las imágenes en YOLO, se utiliza la red convolucional Darknet-19 [11]. Esta red está compuesta por 19 capas convolucionales, 5 capas de max pooling, una capa pooling promedio y una softmax. La red Darknet-19 usa fundamentalmente filtros de convolución de 3x3. Darknet-19, en lugar de capas completamente conectadas se aplica una capa promedio global de pooling en lugar de la capa softmax. La red neuronal convolucional que contiene esta capa en la estructura es de convolución completa [26].

En [27] se demuestra que esta red puede ser aplicada a imágenes de diferente resolución sin necesidad de cambiar los parámetros de la misma.

El set de entrenamiento consistió de 3300 imágenes de región iris-pupila tomadas de las bases de datos MobBio [28] (800), UTIRIS [29] (800) y 1700 tomadas de los cuadros de video de una base de datos propia (ver 4.2). El conjunto de entrenamiento fue preparado manualmente mediante la selección de las regiones rectangulares que encierran a la región iris-pupila (figura 1.b).

### **3.2 Medida para la evaluación de la Calidad de la región Iris-pupila en tiempo real en el video**

La distancia focal es uno de los parámetros más importantes considerados en el estándar [16], este parámetro indica la distancia óptima entre el sujeto y el sensor para una densidad de píxeles normada. La densidad de píxeles está definida en el estándar como la suma cuantitativa de los píxeles que se encuentran en la longitud diagonal de la imagen del iris. La densidad normada en el estándar debe ser al menos de 200 píxeles y debe estar compuesta de al menos 2 líneas de píxeles por milímetro (2 lppmm). De la literatura médica, es posible asumir que el iris representa el 60% de una imagen de la región iris-pupila bajo un ambiente bien iluminado. Entonces la densidad de píxeles ( $Ird$ ) puede ser calculada utilizando el teorema clásico de Pitágoras por la ecuación 1 donde  $w$  y  $h$  son el ancho y el largo en píxeles de las imágenes de iris-pupila.

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

De la variedad de factores que influyen de forma negativa en la calidad de las imágenes de ojo, el grado de nitidez es uno de los más importantes debido a que dependiendo de si la imagen se encuentra borrosa o fuera de foco la misma perderá los detalles de la textura del iris, o sea de las estructuras internas del mismo que son las que poseen el valor identificativo.

En [3] se realizó un análisis del uso del filtro de Kang and Park utilizando imágenes de ojos tomadas en el NIR. El filtro fue utilizado para filtrar las altas frecuencias. El filtro utilizado se compone de un kernel convolucional de 5x5. El mismo está compuesto de tres funciones de 5x5 y amplitud -1, una de 3x3 y amplitud +5 y cuatro de 1x1y amplitud -5 (ver figura 5). El kernel es capaz de estimar de las altas frecuencias dentro de la textura del iris en el NIR mejor que otros operadores del estado del arte, además de poseer un bajo tiempo computacional debido al tamaño reducido del kernel. En el caso de las imágenes en el VS el efecto del kernel pudiese ser similar por lo que este procedimiento puede ser utilizado en nuestra propuesta.

$$\begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & 4 & 4 & 4 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix} + \begin{bmatrix} 5 & 5 & 5 \\ 5 & 5 & 5 \\ 5 & 5 & 5 \end{bmatrix} + \begin{bmatrix} -5 & -5 \\ -5 & -5 \end{bmatrix}$$

Figura 5. Esquema general del filtro de Kang y Park

La entropía de una imagen de iris ha probado ser un buen indicador del volumen de información contenida en ella [30]. La entropía solo depende del volumen de niveles de gris y la frecuencia de cada nivel.

Tomando en cuenta estos elementos y considerando que la densidad de pixeles de la imagen de iris es otro de los elementos fundamentales para la calidad de la imagen, proponemos su combinación con el filtro de Kang y Park y la estimación de la entropía para obtener una medida de la calidad de la imagen iris-pupila ( $Q_{iris}$ ). La medida  $Q_{iris}$  propuesta es obtenida por la ecuación 2

$$Q_{iris} = \frac{Ird * kpk * ent}{tird * tkpk * tent} \quad (2)$$

Donde  $kpk$  es el valor promedio de los pixeles de la imagen obtenida como resultado de la convolución de la imagen de iris-pupila de entrada con el kernel de Kang y Park.  $tird$  es el umbral establecido por el estándar [16] para el mínimo  $Ird$  para obtener una imagen de calidad.  $tkpk$  es el umbral estimado de  $kpk$  para obtener una imagen de calidad, en [3] los autores, a partir de sus resultados experimentales, recomiendan un umbral=15.  $ent$  es el valor de la entropía de la imagen.  $tent$  es el umbral estimado de  $ent$  con el cual es posible obtener una imagen de calidad.

La entropía de una imagen de iris se calcula por la ecuación 3, donde el valor de  $p_i$  es la probabilidad de ocurrencia de un valor de pixel dado dentro de la región del iris y  $n$  es el número de pixeles de la imagen.

$$ent = - \sum_{i=1}^n p_i \log_2 p_i \quad (3)$$

Para la definición del umbral de entropía ( $tent$ ) realizamos un experimento que nos permitió establecer que las imágenes de iris-pupila tomadas en el VS con una calidad acorde al estándar internacional poseen una entropía mayor que 4.

Para el experimento, creamos un conjunto de 300 imágenes de la región iris-pupila cortadas de forma manual y estandarizada al tamaño de 260x260 a partir de imágenes de ojo. Las imágenes originales fueron seleccionadas de las bases de datos MobBio [28] (150 imágenes) y UBIRIS [29] (150 imágenes).

Para determinar el valor mínimo de entropía para una imagen de calidad se realizó la evaluación de cada una de ellas por los parámetros de densidad de píxeles (*lrd*) y su respuesta al filtro de Kang y Park. Los resultados experimentales se muestran en la figura 6, estos establecen que aquellas imágenes con valores mayores que los umbrales correspondientes (*tird*=200; *tkpk*= 15) poseen una entropía mayor o igual que 4, entonces podemos asumir este valor como el valor de *tent*.

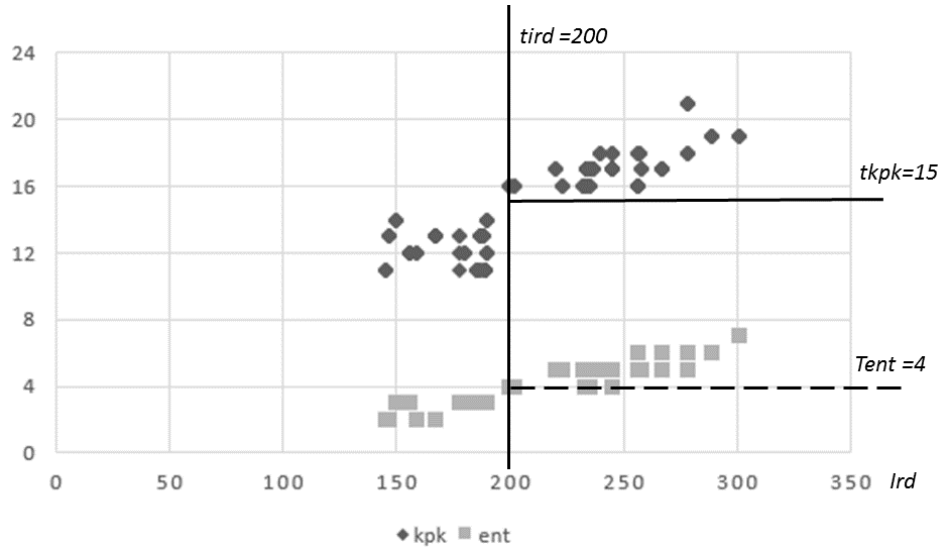


Figura 6. Resultados experimentales para determinar el *tent*

La medida *Qiris* puede alcanzar valores que dependen de los umbrales seleccionados para *lrd*, *kpk* y *ent*. De esta manera considerando los valores de los umbrales *tird*= 200 establecido por el estándar, *tkpk*=15 obtenido experimentalmente en [3] y *tent*=4 obtenido experimentalmente por nosotros, el mínimo valor de *Qiris* para obtener una imagen iris-pupila de calidad será de 1, valores mayores denotarán a imágenes de mayor calidad y valores menores que 1 denotarán imágenes de menor calidad que la establecida por el estándar.

Un tema a explorar en la etapa experimental de este trabajo es determinar bajo que valores mínimos de *Qiris* es posible obtener exactitudes aceptables en el reconocimiento para una configuración dada del sistema.

### 3.3. Clasificación semántica de la región iris-pupila para la segmentación del iris.

La clasificación semántica de las imágenes oculares ha demostrado un mejor resultado en la segmentación del iris evitando o disminuyendo las clasificaciones erróneas de píxeles que no pertenecen al iris [25]. Como resultado de esta clasificación a cada región de la imagen ocular se le asigna una etiqueta que denota la clase anatómica o al artefacto (reflexión especular, gafas) a que pertenece la región.

A partir de los resultados obtenidos en [25] en este trabajo proponemos el uso de un esquema de entrenamiento donde una FCN entrenada fue adaptada para el caso de la clasificación semántica de la región iris-pupila a través de un proceso de transferencia de aprendizaje inductivo [31]. Para ello seleccionamos el modelo FCN fcn8s-at-once el que fue afinado y tuneado a partir del modelo VGG-16 [32]. El proceso de tuneado fino nos permite seleccionar el espacio de hipótesis más prometedor para ajustar el conocimiento objetivo. La implementación de la arquitectura fcn8s-atonce se realizó en el framework Caffe [33]. Las capas intermedias de submuestreo fueron inicializadas por una interpolación bilineal. El número de salida del modelo fue fijado acorde a las 5 posibles clases que describen la región iris-pupila (fig.7a). A diferencia de la propuesta de [25], la clasificación hecha sobre la imagen en nuestra propuesta se realiza solo sobre la región iris-pupila (fig.7b) y no sobre toda la imagen ocular (fig. 7a). Pensamos que esta propuesta reduzca significativamente el tiempo de cómputo respecto a la propuesta de [25] debido al menor tamaño de la imagen a ser clasificada y el menor número de clases a ser consideradas.

Para el proceso de entrenamiento se tomó el conjunto conformado por 500 imágenes oculares ground truth de segmentación semántica creado de forma manual en [25] a partir de un conjunto de imágenes de la base de datos NICE I [34] Estas anotaciones públicas [35] fueron construidas al nivel de pixel en donde cada pixel fue etiquetado con su correspondiente clase. Para el entrenamiento de la red se tomaron las 500 imágenes de este set de entrenamiento y se recortaron de forma manual quedándonos solo con la región iris-pupila de cada imagen como se muestra en la figura 7b.



Figura. 7. Clasificación semántica. a) Ejemplo tomado de [35] b) ejemplo de muestra para el entrenamiento

### Segmentación del iris

El problema principal para incorporar FCN a un sistema de reconocimiento de iris, es que este clasifica la imagen pixel a pixel. Debido a la necesidad de normalizar la textura del iris a partir del modelo Rubbersheet [17], en este trabajo proponemos combinar el algoritmo basado en pixeles (FCN) con un enfoque basado en bordes que permita la detección de las coordenadas de centro y radio de las regiones iris y pupila.

Dada la imagen de la región iris-pupila  $I(x,y)$ ,  $S_k$  la clasificación de pixeles realizada usando 5 clases tal que  $s_k(x,y) \in \{1,2,3,4,5\}$  sean las etiquetas asociadas al iris y la pupila respectivamente:

1. Obtener  $I_k$  y  $P_k$ : mascaras binarias de iris y pupila respectivamente para  $S_k$ . (figuras 8c y 8d)

$$I_k = \begin{cases} 1, & \text{si } S_k(x,y) = i \\ 0, & \text{eoc} \end{cases} \quad (4)$$

$$P_k = \begin{cases} 1, & \text{si } S_k(x,y) = p \\ 0, & \text{eoc} \end{cases} \quad (5)$$

2. Obtener contornos de  $CI_k$ (iris) y  $CP_k$ (pupila) aplicando el filtro de Canny. (fig. 8e, 8f)
3. Para cada contorno  $CI_k$  obtener el menor círculo que lo contiene completamente, mediante métodos de aproximación circular (fig. 8g).
4. De todos los círculos candidatos a iris encontrados en el paso anterior, designar como iris ( $C_i$ ) al de mayor radio (fig.8i)
5. Para cada contorno de  $CP_k$  obtener el menor círculo que lo contiene completamente tal y como se hizo en el paso 3 (fig. 8h)
6. Para cada círculo candidato a pupila obtenido en el paso anterior, descartar aquellos que no estén contenidos en  $C_i$ , de los restantes, designar como pupila ( $C_p$ ) al de mayor radio (fig. 8i).
7. Normalizar la región del iris a partir de las coordenadas de los círculos  $C_i$  y  $C_p$  usando el modelo Rubbersheet (fig. 8j).

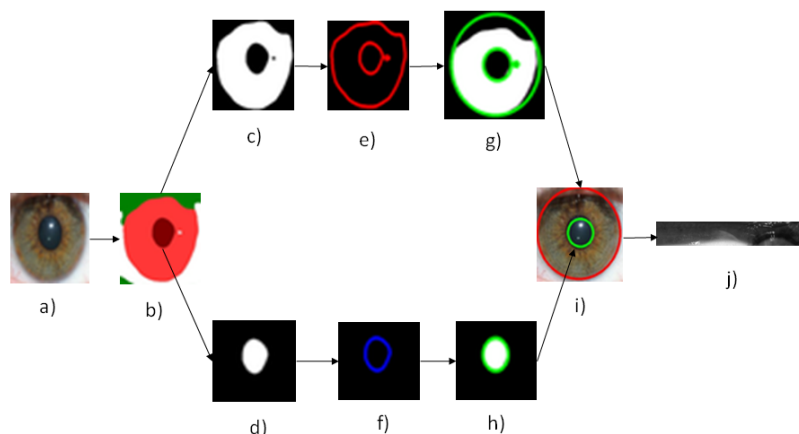


Figura 8. Procesos de obtención de centro y radio de pupila e iris a partir de la imagen de la región iris-pupila (a) clasificada por FCN (b).

## 4. Resultados Experimentales y Discusión

Para validar la propuesta, el diseño experimental fue dirigido a verificar la influencia del método propuesto en las tareas de segmentación y reconocimiento mediante su evaluación utilizando dos bases de datos de imágenes de referencia y una base de datos de video propia. Comparamos el desempeño de nuestra propuesta con los resultados obtenidos por dos métodos del estado del arte [6] y [25]

Tres funcionalidades básicas componen la plataforma experimental implementada en lenguaje c++. Las dos primeras: **Adquisición de video** y **Segmentación**, contienen los métodos propuestos descritos en las secciones anteriores.

En la tercera, **Extracción de rasgos y comparación**, para el propósito de los experimentos en este trabajo implementamos la combinación de dos métodos de extracción de rasgos del estado del arte con el objetivo de verificar la robustez del método propuesto respecto al uso de diferentes rasgos para el reconocimiento, Scale-Invariant Feature Transform (SIFT) [36], y Uniform Local Binary Patterns (LBP) [37]. Para la comparación utilizamos las medidas de similitud de cada uno de estos métodos. La primera estima la similitud entre dos conjuntos de puntos clave SIFT [36]. El segundo utiliza la distancia Chi-cuadrado [37] que es una prueba no paramétrica para medir la similitud entre dos imágenes de una clase. En ambos casos la distancia mínima encontrada dará la medida de máxima similitud entre ellas.

### 4.1. Bases de datos

MobBio [28] es una base de datos multi-biométrica que incluye rostro, iris y voz de 105 voluntarios. El subconjunto de iris contiene 16 imágenes oculares en el VS de cada individuo con una resolución de 250 X 200. Las imágenes fueron tomadas bajo dos condiciones de iluminación diferente, con diferentes grados de oclusión y orientación del iris. Estas fueron capturadas con un Tablet Asus Transformer Pad TF300T a una distancia de entre 10 y 50 centímetros.

UTIRIS [29], es una base de datos híbrida propuesta por miembros de la Universidad de Teheran. Esta posee imágenes bajo espectro de luz infrarroja y visible de 79 personas, para un total de 1540 imágenes separadas en 154 clases. En el caso de las imágenes bajo espectro de luz visible, la captura fue realizada con una cámara Canon EOS 10D, las imágenes fueron capturadas con 3MP y la resolución de las mismas es de 2048 x 1360.

Nuestra base de datos (DatIris) está compuesta por 82 videos de 41 personas tomados en dos sesiones de 10 segundos bajo las condiciones y con el dispositivo de adquisición de video presentado en la sección 3. La cámara utilizada fue una webcam Logitech C920 HD Pro, a una resolución de 920 x 1080. Los videos fueron tomados en condiciones interiores. La base de datos contiene videos de personas de piel clara de origen caucásico, piel oscura de origen africano y mestizos, esta compuesta por 26 hombres y 15 mujeres, en un rango de edades de 10 a 65 años (ver ejemplos en la figura 9).



Figura 9. Ejemplos de las bases de datos utilizadas. MobBio, UTIRIS y DatIris.

## 5. Resultados Experimentales

### 5.1. Evaluación de la etapa de detección de la región iris-pupila

La eficacia del método de detección propuesto fue medida mediante el grado total de coincidencia pixel a pixel ( $C_{det}$ ) entre las imágenes de la zona iris-pupila detectadas automáticamente en un video de 10 segundos por el detector y sus equivalentes obtenidas de forma manual (ground truth) y la misma se calcula por la expresión (6).

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

Donde,  $Idt(i)$  es la  $i$  imagen de iris-pupila detectada automáticamente en el video con  $Q_{iris} \geq 1$ ,  $Igt(i)$  es la  $i$  imagen ground truth obtenida manualmente del correspondiente cuadro del video donde fue detectada la imagen automáticamente;  $k$  es el numero total de detecciones con  $Q_{iris} \geq 1$  ocurridas en el video. El desempeño del detector propuesto fue comparado con el desempeño del detector propuesto en [6].

El detector fue probado sobre un set de 20 videos correspondientes a 20 personas diferentes de nuestra base de datos DatIris (ver 4.1).

La figura 10 muestra los resultados del experimento donde se evidencia que el detector propuesto es capaz de detectar una mayor cantidad de imágenes de calidad por video respecto al detector basado en Viola y Jones [6], de forma similar nuestra propuesta basada en YOLO logra una mayor exactitud que [6]. Por otro lado ambos detectores lograron una efectividad del 100% en la detección con calidad de ambos iris-pupila de cada individuo.

### 5.2. Evaluación de la etapa de segmentación propuesta

La exactitud del método de segmentación propuesto fue medida mediante la comparación por los resultados obtenidos en [25] y [6] con nuestra propuesta en la base de datos MobBio. Esta base de datos proporciona anotaciones manuales de los contornos pupilares y del iris (ground truth).

Para el experimento, se evaluó la calidad de la base de datos por el índice de calidad propuesto y fueron seleccionadas 200 imágenes con un  $Q_{iris} > 1$  con el objetivo de garantizar la calidad del experimento.

Como métrica para la evaluación consideramos la medida  $E^1$  propuesta por el protocolo NICE I (<http://nice1.di.ubi.pt/>). Esta métrica estima la proporción de los pixeles no coincidentes correspondientes.

Para la estimación de la eficiencia del método propuesto se calculó el tiempo (en segundos) que el método de segmentación toma para obtener una imagen segmentada del iris. El mismo se calculó utilizando una PC con procesador Core i5-3470 a 3.2 GHz y 8 GB de RAM. La tabla 1 lista los valores promedio de  $E^1$  por [25], [6] y nuestra propuesta. Los resultados muestran un mejor desempeño del método propuesto en cuanto a exactitud y un tiempo de cálculo para el proceso de segmentación similar a [6] e inferior a [25].

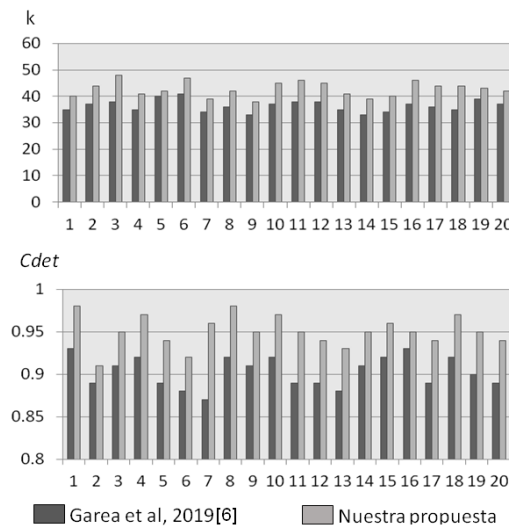


Figura 10. Resultados de la evaluación del detector basado en YOLO.

Tabla1. Resultados experimentales de la etapa de segmentación propuesta para la base de datos MobBio.

| Osorio-Roig et al [25] |         | Garea et al[6] |         | Nuestra propuesta |         |
|------------------------|---------|----------------|---------|-------------------|---------|
| $E^1$                  | t (sec) | $E^1$          | t (sec) | $E^1$             | t (sec) |
| 2.5                    | 4       | 2.4            | 0.3     | 1.7               | 0.3     |

Para determinar la influencia del método propuesto en los subsiguientes pasos que conlleva el proceso de reconocimiento de personas en un sistema biométrico, se desarrolló un experimento para comprobar su efecto en la tarea de verificación. La exactitud de esta tarea se midió mediante el error de falso rechazo (FRR por sus siglas en inglés) ante una tasa de falsa aceptación (FAR por sus siglas en inglés)  $\leq 0.001\%$ . Para ello se realizaron comparaciones en forma de todos contra todos para obtener la distribución de genuinos e impostores. El experimento se desarrolló sobre las tres bases de datos. Para la base de datos DatIris, los 41 videos de la sesión 1 fueron procesados tomando dos imágenes de cada iris y comparándolas contra un conjunto compuesto de dos imágenes de cada iris tomadas de los cuadros de los videos de la sesión 2.

La tabla 2 muestra la comparación del error FRR ante un FAR  $\leq 0.001\%$  obtenidos por [25], [6] y nuestra propuesta sobre las tres bases de datos experimentadas, tomando dos intervalos *Qiris* para aceptar o rechazar las imágenes iris-pupila a ser procesadas. Los resultados muestran que al aumentar el valor de *Qiris*, el sistema soporta imágenes de mayor calidad y rechaza imágenes de baja calidad. Este incremento en la calidad, resulta en un decrecimiento del FRR, con un resultado significativo en UTIRIS donde un FRR =0.03 es alcanzado con el 79% de la base de datos y un *Qiris*>1 utilizando los rasgos LBP. Sin embargo en la base de datos MobBio, el incremento del umbral de *Qiris* da como resultado un decrecimiento significativo en el número de imágenes a ser comparadas. Los resultados obtenidos en la base de datos DatIris muestran un alto desempeño del proceso de verificación obteniendo un FRR de 0.009, lo que corrobora la relevancia del índice de calidad propuesto para una aplicación real.

Se observa también que los niveles de FRR se tornan muy cercanos para los tres métodos evaluados lo que demuestra que la disminución del área de clasificación de la imagen ocular de toda la imagen de ojo en [25] a la región iris-pupila en [6] y nuestra propuesta no afecta la exactitud del proceso de reconocimiento observándose además un ligero aumento de la misma para el caso de nuestra propuesta, lo que puede estar relacionado con el aumento de la exactitud experimentado en el proceso de detección de la región iris-pupila propuesto basado en YOLO respecto al método basado en Viola y Jones [6]

La tabla 3 muestra los resultados de la comparación de la medida de calidad (*Qindex*) propuesta en [15] con nuestra propuesta (*Qiris*), en los valores  $\geq 1$  utilizando los rasgos LBP y el método de segmentación propuesto. Se puede apreciar que partir de la inclusión de la entropía de la imagen de iris, el sistema considera un porcentaje de imágenes mayor a ser procesado mientras mantiene un nivel similar de exactitud en el reconocimiento.

Tabla 2. Comparación del FRR obtenido en el proceso de verificación por [25], [6] y nuestra propuesta.

| Base de datos | Qiris | % de imágenes * | FCN[12] |      | CIARP 2019 |      | Nuestra propuesta |              |
|---------------|-------|-----------------|---------|------|------------|------|-------------------|--------------|
|               |       |                 | SIFT    | LBP  | SIFT       | LBP  | SIFT              | LBP          |
| MobBio        | <1.0  | 73.7            | 0.28    | 0.26 | 0.27       | 0.25 | 0.27              | <b>0.25</b>  |
|               | >=1.0 | 26.3            | 0.24    | 0.23 | 0.25       | 0.22 | <b>0.23</b>       | <b>0.21</b>  |
| UTIRIS        | <1.0  | 20.5            | 0.06    | 0.05 | 0.07       | 0.06 | 0.07              | <b>0.05</b>  |
|               | >=1.0 | 79.5            | 0.07    | 0.04 | 0.06       | 0.04 | <b>0.05</b>       | <b>0.03</b>  |
| DatIris       | <1.0  | 10              | 0.03    | 0.02 | 0.04       | 0.03 | <b>0.03</b>       | 0.03         |
|               | >=1.0 | 90              | 0.02    | 0.01 | 0.02       | 0.01 | 0.02              | <b>0.009</b> |

\* es el porciento de imágenes procesadas con valores de calidad especificados en la columna Qiris

Tabla 3 Comparación de la medida de calidad propuesta en [15] (Qindex) con nuestra propuesta (Qiris)

| Base de datos | % de imag proc. ** |       | FRR         |              |
|---------------|--------------------|-------|-------------|--------------|
|               | Qindex[15]         | Qiris | Qindex [15] | Qiris        |
| MobBio        | 19.5               | 26.3  | 0.23        | <b>0.21</b>  |
| UTIRIS        | 71.8               | 79.5  | 0.04        | <b>0.03</b>  |
| DatIris       | 88                 | 90    | 0.02        | <b>0.009</b> |

\*\* es el porciento de imágenes procesadas con valores de calidad >=1

## Conclusiones

En este trabajo se propuso un método para el reconocimiento biométrico del iris en video. Nuestra propuesta se centra en las etapas de captación y segmentación del iris y se basa en la combinación de métodos de aprendizaje profundo para la detección directa de la región del iris-pupila (YOLO), la evaluación de su calidad y la clasificación semántica de las regiones que la componen por una red FCN para realizar la segmentación del iris.

La obtención de un detector de la región iris-pupila basado en YOLO, permite detectar una mayor cantidad de imágenes de calidad en el video respecto al detector basado en Viola y Jones [6] a la par de lograr una mayor exactitud sin aumento de los tiempos de cómputo.

Se analizó la relevancia de la etapa de evaluación de la calidad de la imagen como paso fundamental para filtrar la información generada por el proceso de captura del video. Los resultados experimentales muestran que la inclusión de la evaluación de calidad propuesta como etapa previa al proceso de segmentación del iris limita el paso de imágenes de baja calidad al sistema, lo que conlleva a un incremento de las tasas de reconocimiento y una reducción en los tiempos de respuesta del sistema.

La aplicación de la clasificación semántica a la imagen de la región iris-pupila en el proceso de captura del video permite la reducción del tiempo computacional en el proceso de segmentación lo que en combinación con el método de detección propuesto.

## Referencias

1. Raja KB, Raghavendra R, Vemuri VK, Busch C. Smartphone based visible iris recognition using deep sparse filtering. Pattern Recognition Letters 2015, 57: 33-42.
2. Hollingsworth K, Peters T, Bowyer K. Iris recognition using signal-level fusion of frames from video, IEEE Trans. Inf. Forensics Security 2009, 4: 837-848.
3. Garea-Llano E, García-Vázquez M, Colores -Vargas JM, Zamudio-Fuentes LM, Ramírez-Acosta AA. Optimized robust multi-sensor scheme for simultaneous video and image iris recognition. Pattern Recognition Letters 2018, 101: 44-45.
4. H. Proenc,a, "Unconstrained iris recognition in visible wavelengths," in Handbook of Iris Recognition (2nd Edition), W. K. Bowyer and J. M. Burge, Eds. Springer London, 2016, pp. 321–358.
5. Raja KB, Raghavendra R, Busch C. Iris Imaging in Visible Spectrum using White LED. In proc. BTAS 2015, IEEE.
6. Eduardo Garea-Llano, Annette Morales-González, Dailé Osorio-Roig. Video Iris Recognition Based on Iris Image Quality Evaluation and Semantic Classification. LNCS. CIARP2019.
7. Viola P, Jones M Rapid Object Detection Using a Boosted Cascade of Simple Features. Mitsubishi Electric Research Laboratories, Inc., 2004 201 Broadway, Cambridge, Massachusetts 02139.

8. Zhihao Chen, Redouane Khemmar, Benoit Decoux, Amphani Atahouet, Jean-Yves Ertaud. Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility. 2019 Eighth International Conference on Emerging Security Technologies (EST), Jul 2019, Colchester, United Kingdom. 10.1109/EST.2019.8806222. hal-02343350.
9. Pavlov. V A and Galeeva. M A. Detection and recognition of objects on aerial photographs using convolutional neural networks. J. Phys.: Conf. Ser. 1326 012038, 2019
10. Liu. W. et al., SSD: Single Shot MultiBox Detector, ECCV 2016, pp 21-37.
11. Redmon J and Farhadi A 2017 YOLO9000: Better, Faster, Stronger IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 7263-7271
12. Schmid N, Zuo J, Nicolo F, Wechsler H. Iris Quality Metrics for Adaptive Authentication. In: Bowyer KW, Burge MJ. Handbook of Iris Recognition (2nd Edition), Springer-Verlag London 2016, 101-118.
13. Daugman J, Downing C. Iris Image Quality Metrics with Veto Power and Nonlinear Importance Tailoring. In: Rathgeb C, Busch C. Iris and Periocular Biometric Recognition. IET Pub. 2016,; 83-100.
14. Zuo J, Schmid NA. An automatic algorithm for evaluating the precision of iris segmentation. In: BTAS'08, Washington, DC, USA. 2008.
15. Garea-Llano E, Osorio-Roig D, Hernandez O. Image Quality Evaluation for Video Iris Recognition in the Visible Spectrum. Biosensors and Bioelectronics Open Access (ISSN:2577-2260), 2018.
16. ISO/IEC 19794-6:2005. Part 6: Iris image data, ISO.
17. Daugman. J., "How iris recognition works," IEEE Trans. on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 21–30, 2004.
18. Proença H. and Alexandre. L. A., "Iris recognition: Analysis of the error rates regarding the accuracy of the segmentation stage," Image and Vision Computing, vol. 28, no. 1, pp. 202–206, 2010.
19. Proença H. and Alexandre L. "Toward covert iris biometric recognition: Experimental results from the nice contests," IEEE Trans. on Information Forensics and Security, vol. 7, no. 2, pp. 798–808, 2012.
20. Tan T., He Z., and Sun Z., "Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition" Image and vision computing, vol. 28, no. 2, pp. 223–230, 2010.
21. W. Sankowski, K. Grabowski, M. Napieralska, M. Zubert, and A. Napieralski, "Reliable algorithm for iris segmentation in eye image," Image and Vision Computing, vol. 28, no. 2, pp. 231–237, 2010.
22. Liu N., Li H., Zhang M., Liu J., Sun Z., and Tan T., "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.
23. Jalilian E. and Uhl A., "Iris segmentation using fully convolutional encoder–decoder networks," in Deep Learning for Biometrics. Springer, 2017, pp. 133–155.
24. Osorio-Roig D., Morales-Gonzalez A., and Garea-Llano E., "Semantic segmentation of color eye images for improving iris segmentation," in Proc. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (CIARP'17). Springer, 2017, pp. 466–474.
25. Osorio-Roig D., Rathgeb C., Gomez-Barrero M., Morales-González Quevedo A., Garea-Llano E. and Busch C. Visible Wavelength Iris Segmentation: A Multi-Class Approach using Fully Convolutional Neuronal Networks. BIOSIG 2018, IEEE.
26. Lin M, Chen Q and Yan S 2013 Network In Network (Preprint arXiv: arXiv:1312.4400v3)
27. Bobrovsky A, Galeeva M, Morozov A, Pavlov V and Tsytsulin A 2019 Automatic detection of objects on star sky images by using the convolutional neural network Journal of Physics: Conference Series 1236 1066
28. Monteiro C, Oliveira HP, Rebelo A, Sequeira AF. Mobbio 2013: 1st biometric recognition with portable devices competition, available in: <https://paginas.fe.up.pt/~mobbio2013/>.
29. Hosseini MS, Araabi BN, Soltanian-Zadeh H. Pigment Melanin: Pattern for Iris Recognition. IEEE Transactions on Instrumentation and Measurement 2010, 59: 792-804.
30. Waleed S.-A. Fathy, Hanaa S.Ali .Entropy with Local Binary Patterns for Efficient Iris Liveness Detection. Wireless Personal Communications, 102(3), 2331-2344, 2018
31. Torrey L. and Shavlik J. "Transfer learning," in Handbook of Research on Machine Learning Applications, 2009.
32. Simonyan K. and Zisserman A. "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
33. Jia Y., Shelhamer E., Donahue J. Karayev, S, Long J., Girshick R., Guadarrama S., and Darrell T., "Caffe: Convolutional architecture for fast feature embedding," in Proc. of the 22nd ACMMM2014, pp. 675–678.
34. Proença H., Filipe S., Santos R., Oliveira J., and Alexandre L. A., "The ubiris.v2: A database of visible wavelength iris images captured on-themove and at-a-distance," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, no. 8, pp. 1529–1535, 2010.
35. Multi-class iris segmentation ground truth: <https://dasec.h-da.de/research/biometrics/MCIS/>

36. Lowe DG Distinctive image features from scale-invariant key points. *International Journal of Computer Vision* 2004, 60: 91-110.
37. Liao S., Zhu X., Lei Z., Zhang L., and Li S. Z. Learning multi-scale block local binary patterns for face recognition. In *Proc. ICB'07*, pages 828–837, 2007.