UNIVERSIDAD DE MÁLAGA

PhD Thesis - Tesis doctoral
Tesis por compendio de publicaciones

Real-time embedded eye detection system

Camilo Andrés Ruiz Beltrán

Febrero 2024

Departamento de Tecnología Electrónica
Programa de doctorado: Ingeniería de Telecomunicación
Supervisado por:
Dr. Martín González García
Dra. Rebeca Marfil Robles

AUTOR: Camilo Andrés Ruiz Beltrán

https://orcid.org/0000-0001-8270-2062

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña RUIZ BELTRÁN, CAMILO ANDRÉS
Estudiante del programa de doctorado INGENIERÍA DE TELECOMUNICACIÓN de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: REAL-TIME EMBEDDED EYE DETECTION SYSTEM

Realizada bajo la tutorización de BANDERA RUBIO, ANTONIO JESUS y dirección de GONZALEZ GARCIA, MARTIN; MARFIL ROBLES, REBECA (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 5 de ABRIL de 2024

Fdo.: RUIZ BELTRÁN, CAMILO ANDRÉS Doctorando/a	Fdo.: BANDERA RUBIO, ANTONIO JESUS Tutor/a
Fdo.: GONZALEZ GARCIA, MARTIN; MARFIL ROBLES, REBECA Director/es de tesis	

UNIVERSIDAD DE MÁLAGA
DEPARTAMENTO DE TECNOLOGÍA ELECTRÓNICA

Los Drs. D. Martín González García y D^a. Rebeca Marfil Robles, profesores del Departamento de Tecnología Electrónica de la E.T.S.I. de Telecomunicación de la Universidad de Málaga, Certifican que D. Camilo Andrés Ruiz Beltrán, Ingeniero Electrónico, ha realizado en el Departamento de Tecnología Electrónica de la Universidad de Málaga en el programa de doctorado Ingeniería de Telecomunicación, bajo nuestra dirección, el trabajo de investigación correspondiente a su Tesis Doctoral titulada:

"Real-time embedded eye detection system"

Revisado el presente trabajo, estiman que puede ser presentado al tribunal que ha de juzgarlo, y autorizan la presentación de esta Tesis Doctoral en la Universidad de Málaga. Además certifican que las publicaciones en coautoría que avalan la tesis no han sido utilizadas en tesis anteriores.

Málaga, 24 de enero de 2024

Dr. D. Martín González García
Prof. Dpto. Tecnología Electrónica

Dra. D^a Rebeca Marfil Robles
Prof. Dpto. Tecnología Electrónica

UNIVERSIDAD DE MÁLAGA
DEPARTAMENTO DE TECNOLOGÍA ELECTRÓNICA

El Dr. Antonio Jesús Bandera Rubio, Profesor Titular de Universidad, perteneciente al Departamento de Tecnología Electrónica de la E.T.S.I de Ingeniería de Telecomunicación de la Universidad de Málaga, Certifica que D. Camilo Andrés Ruiz Beltrán, Ingeniero Electrónico, ha realizado en el Departamento de Tecnología Electrónica de la Universidad de Málaga en el programa de doctorado Ingeniería de Telecomunicación, bajo su tutorización, el trabajo de investigación correspondiente a su Tesis Doctoral titulada:

"Real-time embedded eye detection system"

Málaga, 24 de enero de 2024

Dr. D. Antonio Jesús Bandera Rubio
Prof. Dpto. Tecnología Electrónica

Dedico esta tesis a todas las personas que aprecio en esta vida y que tanto me han apoyado.

Contents

Abstract

Biometric identification by iris recognition is based on the analysis of the iris pattern using mathematical techniques. Although it is a relatively recent technique (the first automatic identification system was developed and patented by John Daugman in the last decade of the 20th century), its excellent identification characteristics have led to its rapid evolution. Thus, using the most recent developments, it has become a mature technique. In these systems, finding the exact position of an eye is crucial to extract the iris region quickly and correctly. Unfortunately, this is very difficult to achieve without the active cooperation of the user, so designing systems where the camera captures the user's image from a distance and without the user stopping to frame the eye in the image remains a challenge. In iris recognition at a distance (IAAD) systems based on the use of a single high-resolution camera, the system shall be able to process these high-resolution images at high speed. The captured image must have enough resolution and enough contrast, this can be achieved using a combination of infrared lightning, small aperture and short exposition time, however this approach results in a shallow the depth of field of the camera (i.e., the distance around the image plane for which the image sensor is focused). Also, because the person to be identified is moving, it is difficult to time the camera shutter release to coincide with the moment when the iris is in the depth of field and therefore in focus. Only by processing many images per second can the system try to ensure that one of the images has captured this moment. On the other hand, current commercial and research systems for addressing this task use software frameworks that require a dedicated computer, whose power consumption, size and price are significantly large. In this Thesis, different versions of an eye detector have been designed and implemented, always with the premise that they should offer a high detection rate of true positives and that they should run in real time on an edge device. The first of the proposals falls within the framework of what we now consider to be traditional computer vision, and proposes a massive parallelisation of the Viola and Jones algorithm for object detection. The second implements a Deep Learning approach, in particular, a version of the popular one-stage strategy YOLO. Both proposals have been implemented on a Multi-processor system-on-chip (MPSoC) platform, and offer processing speeds that exceed those of the image sensor used in the practical implementation of the system. In addition, both proposals took into account the need to discard images that were not of sufficient quality for the extracted irises to be used to identify the user. The work carried out could therefore be divided into three main chapters: the implementation of a functional design based on fully parallelised Haar features and an AdaBoost classifier; the inclusion in this system of a module for analysing the quality of the detected eye images; and the implementation of a new design based on YOLO. These three main chapters have been published separately in international journals, so, as the structure of this Doctoral Thesis should coincide with the division into contributions, it was decided to present the thesis as a compendium of publications.

Agradecimientos

Ahora que la tesis está terminada miro hacia el pasado y pienso en todas las cosas que han tenido que alinearse y todas las personas que me han ayudado de alguna manera y por eso quero expresarles mi gratitud. A mis padres, Omar y Gladys, que siempre me dado apoyo y amor incondicional. Estoy seguro que les llena de orgullo verme cumplir mis objetivos. A mi amada Alba, no puedo agradecerte lo suficiente. Los últimos años han sido duros, y aún así hemos podido seguir persiguiendo nuestras metas. Tu compañía y amor llenan mis días de felicidad. A Ana y a José, me habéis acogido en vuestra familia como si fuera vuestro hijo y siempre estaré agradecido por vuestra generosidad. He podido terminar este trabajo gracias a todos vosotros.

Naturalmente no puedo olvidarme de la invaluable ayuda del tutor de esta tesis, Dr. Antonio Bandera y de los directores, Dra. Rebeca Marfil y Dr. Martín Gonzáles. Me habéis guiado y ayudado con vuestra sabiduría y experiencia, no solo para terminar la tesis sino en muchos más aspectos en la vida. Siempre tendréis mi gratitud.

Camilo Andrés Ruiz Beltrán
Málaga
Enero 2024

Esta tesis fue parcialmente apoyada por el Proyecto Técnico Integrado MIRoN financiado, a su vez, por el proyecto UE RobMoSys (H20202-732410), los proyectos LYNX e HIRIS, ambos financiados por el Centro para el Desarrollo Tecnológico Industrial, E.P.E. (CDTI-E.P.E.), y el proyecto CPP2021-008931 (DIMAS), financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea NextGenerationEU/PRTR. Además, el doctorando ha formado parte de los equipos de trabajo en proyectos como el RTI2018-099522-B-C41, financiado por el Ministerio de Ciencia, Innovación y Universidades y fondos FEDER, y PDC2022-133597-C42. TED2021-131739B-C21 y PID2022-137344OB-C32, financiados por MCIN/AEI/10.13039/501100011033 y por NextGenerationEU/PRTR de la Unión Europea (para las dos primeras subvenciones), y "FEDER Una forma de hacer Europa" (para la tercera subvención).

Acknowledgments

Now that this thesis is finally ready, I look back and think about all the things that have aligned and all the people that helped to allow the completion of this work. Therefore, I want to express gratitude to everyone involved. My parents, Omar and Gladys, have given me all their support and unconditional love, I am sure they are happy and proud to see me reaching my goals. My dearest Alba, I cannot thank you enough. The last few years have been difficult, but we have been able to keep pursuing our ideas, your company and love fill my days with joy. Ana and Jose I'll always be thankful for your generosity. You embraced me into your family and treated me as another son. I was able to finish this work because of all of you.

Of course, I cannot forget about the irreplaceable help of the tutor of this thesis, Dr. Antonio Bandera and the directors, Dr. Rebeca Marfil and Dr. Martín Gonzáles. They guided and helped me with their expertise and experience, not only towards the completion of this thesis, but also in many more aspects. They will always have my gratitude.

Camilo Andrés Ruiz Beltrán
Málaga
January 2024

This thesis was partially supported by the MIRoN Integrated Technical Project funded, in turn, by the EU RobMoSys project (H20202-732410), the LYNX and HIRIS projects, both funded by the Centro para el Desarrollo Tecnológico Industrial, E.P.E. (CDTI-E.P.E.), and the CPP2021-008931 (DIMAS) project, funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. Moreover, the PhD student has been part of the working teams in projects such as the RTI2018-099522-B-C41, funded by the Spanish Ministerio de Ciencia, Innovación y Universidades and FEDER funds, and PDC2022-133597-C42, TED2021-131739B-C21 and PID2022-137344OB-C32, funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR (for the first two grants), and "ERDF A way of making Europe" (for the third grant).

Resumen

Introducción

Las personas usamos todos los días el sentido de la vista para obtener información de nuestro alrededor, observando objetos y apreciando sus detalles. Es una habilidad que, mientras tengamos nuestro sentido de la vista intacto, usaremos de forma continuada a lo largo de nuestras vidas. La supuesta facilidad con la que nosotros resolvemos problemas usando visión ha motivado la búsqueda de alternativas artificiales. No obstante, los algoritmos para emular el funcionamiento de la visión humana en una máquina son complejos, siendo tradicionalmente uno de los principales obstáculos la robustez: nuestra vista identifica un mismo objeto, aunque existan cambios en la iluminación, orientación, fondo, y demás factores aleatorios que afecten lo que vemos en el entorno que nos rodea, y que añaden una gran variabilidad a la imagen que percibimos. Sólo en los últimos años, con la popularización de los algoritmos de aprendizaje profundo y las redes neuronales convolucionales, puede afirmarse que el problema está cerca de ser resuelto.

Se puede afirmar que las aportaciones compartidas por la enorme comunidad de investigadores han permitido que, en las últimas décadas, la visión artificial pase de ser una tecnología de laboratorio a ser útil en escenarios reales y específicos. Así existen, por ejemplo, soluciones comerciales para sistemas de vigilancia, sistemas de control de calidad en fábricas, sistemas de seguridad, o sistemas de navegación, entre otros. En muchas ocasiones, estas soluciones comerciales basadas en visión incluyen algoritmos destinados al procesamiento de información referida a la persona. Uno de los temas a los que se ha prestado más atención es el referido con la detección de caras, ya que el rostro es una fuente muy importante de información, permitiendo extraer de este, por ejemplo, la edad, el género, el estado emocional e información biométrica de un sujeto. Debido a esta necesidad por parte de gran variedad de sistemas, de algoritmos robustos de detección de caras, la investigación en esta área ha recibido una atención considerable en los últimos años, habiéndose publicado aproximadamente 820.000 artículos de este tema en los últimos cinco años según Google Scholar. Además, una vez extraída la cara de un sujeto, es posible determinar el área de los ojos y obtener de ellos información que puede ser usada para detectar la dirección de la mirada, para control de acceso, para detectar el estado de vigilia, etc. La detección de los ojos de una persona es una tarea fundamental en aplicaciones tan importantes como el reconocimiento del iris en la identificación biométrica, o la detección de fatiga en los sistemas de asistencia a la conducción. Resolver esta tarea en el marco específico de la identificación por reconocimiento de iris a distancia será el objetivo fundamental de esta Tesis Doctoral.

Marco de la Tesis

La identificación biométrica a larga distancia y con personas en movimiento es un reto complejo, por un lado la dificultad de que el elemento utilizado para la identificación es pequeño, en este caso, el iris, que es el tejido altamente texturizado en forma de anillo visible externamente presente en el ojo. Cada iris tiene un patrón de textura único que permite la identificación exclusiva de una persona. Esta característica hace que la identificación por reconocimiento del iris sea una solución popular y ampliamente aplicable, especialmente en su versión de identificación del iris a distancia (*Iris At A Distance*, IAAD), que se utiliza en diversos campos como el control de fronteras, la vigilancia, etc. Por otra parte otra dificultad es cuando el sistema debe capturar el iris de la persona en movimiento por lo que el sistema debe funcionar suficientemente rapido para capturar alguna imagen útil.

El sistema propuesto en la presente Tesis Doctoral se ha integrado en una arquitectura capaz de identificar personas mediante el reconocimiento del iris. Dicha arquitectura consiste en una unidad de captura y procesamiento de imágenes del iris que se sitúa en un poste, frente a un arco o punto de acceso. La óptica y el sensor empleados por el sistema permiten capturar imágenes del iris enfocadas a aproximadamente 1,7 m del lugar donde se ubica el sensor. El tiempo de exposición es muy bajo y permite evitar la distorsión por movimiento cuando la persona camina a una velocidad normal (1-2 m por segundo). Sin embargo, la profundidad de campo es muy escasa (sólo unos 10-15 cm), por lo que, para garantizar la captura de una imagen enfocada, el sistema se ve obligado a trabajar a la velocidad máxima del sensor de imagen utilizado (en su última versión, un Teledyne e2v EMERALD 16MP, que proporciona hasta 47 fotogramas por segundo (fps)). El poste incluye tanto el sensor como el sistema de captura y procesamiento, así como la iluminación (51 W mediante LED de alta potencia). Si la distancia entre el sensor y el sujeto aumenta, o disminuye, los parámetros ópticos del sistema (principalmente la distancia focal del objetivo) deben ajustarse para mantener la resolución de captura (un mínimo de 200-250 píxeles/centímetro) y aumentar, en la medida de lo posible, la profundidad de campo. El sistema de captura y procesado está conectado por Ethernet a un ordenador externo, que se encarga de extraer el patrón de iris normalizado y compararlo con la base de datos para cerrar la identificación. Este sistema se desplegará para cubrir un punto de acceso parcialmente controlado, en el que los sujetos deberán caminar a un ritmo normal y deberán evitar cualquier comportamiento que impida la adquisición de la imagen del iris. Con estos sistemas, uno de los grandes problemas es el ya mencionado de que la profundidad de campo es muy limitada, por lo que para captar al menos un par de imágenes de los iris con la calidad de enfoque necesaria, será necesario procesar muchas imágenes por segundo (en nuestro caso, el máximo que proporciona el sensor (los mencionados 47 fps)). El problema es que, como el sistema dispone de unos 2-3 s de grabación por usuario en los que se pueden detectar los ojos, el número de imágenes que habrá que enviar al ordenador externo para su procesamiento puede superar las 250 imágenes. En un caso normal, el ordenador externo no puede procesar este volumen de información antes de que el usuario haya abandonado el punto de acceso. Una solución a este problema es filtrar el gran número de imágenes en las que el iris no está enfocado. Esto supone descartar casi el 97% de las imágenes del ojo captadas por el sistema (Ruiz-Beltrán et al, 2023b).

Por otra parte, capturar y preprocesar un gran volumen de imágenes de entrada requiere el uso de un dispositivo con alta capacidad de computación (*edge-computing*). En la actualidad, las opciones se presentan principalmente en forma de unidades de procesamiento gráfico (GPU), circuitos integrados de aplicación específica (ASIC), o matrices de puertas lógicas programables (FPGA). Debido al elevado consumo y tamaño de las GPU y a la escasa flexibilidad de los ASIC, los FPGA suelen ser la opción más interesante. Además, si el enfoque tradicional de

desarrollo de FPGAs utilizando lenguajes de hardware de bajo nivel (como Verilog y VHDL) suele llevar mucho tiempo y ser muy ineficiente, el uso de herramientas de síntesis de lenguajes de alto nivel (HLS) permite a los actuales desarrolladores programar soluciones de hardware utilizando C/C++ y OpenCL. Esto mejora significativamente la eficiencia en los desarrollos de FPGAs. Por último, las FPGA se integran hoy en día en sistemas en chip multiprocesador (*Multiprocessor System-on-Chip*, MPSoC), en los que se combinan parte lógica y software. Estos MPSoC ofrecen así las capacidades de aceleración de la FPGA y las capacidades computacionales que le permiten trabajar como un sistema autónomo independiente, que no tiene que estar conectado a un ordenador/controlador externo.

Objetivos y metodología

El trabajo realizado en el marco de la presente Tesis Doctoral ha tenido como objetivo el diseño e implementación de un detector de ojos que funcione en tiempo real usando visión artificial. Además, se han diseñado módulos adicionales, que permiten tanto disponer de un canal completo de vídeo, como el analizar a nivel básico estas regiones para determinar su nivel de contraste. El hecho de empotrar la solución diseñada ha generado interés tanto a nivel de investigación, con distintas propuestas que hacen uso de los recursos computacionales de los MPSoC, como a nivel de aplicación, sirviendo como base para una relevante línea de transferencia a la empresa, en la que se enmarcan varios contratos y proyectos de colaboración con la empresa SHS Consultores SL por valor superior a los 500K EUR, todos ellos en el marco de la identificación biométrica por reconocimiento del iris.

Al usar visión artificial, la tarea consiste en extraer, de la imagen de entrada, las subimágenes en las que se ubican los ojos de la persona o personas que aparecen en ella. Estos recortes de la imagen de entrada deberán ser enviados a un segundo sistema, que será el encargado de extraer el patrón de iris y abordar su reconocimiento. Esta tarea de detección de ojos se puede realizar mediante el uso de técnicas más heurísticas y tradicionales, como por ejemplo el algoritmo basado en características Haar y clasificador AdaBoost propuesto por Viola y Jones (2001), o usando técnicas de aprendizaje profundo, como pueden ser las redes neuronales convolucionales (*Convolutional Neural Networks*, CNN). El entorno en el que funcionará el detector hace que el sistema deba ser empotrable con unos requisitos críticos de tiempo real y bajo consumo. Para cumplir estos requisitos se exploraron inicialmente diferentes alternativas en cuanto al hardware, pero finalmente nos decidimos por llevarlo a cabo basándonos en las FPGA (*Field Programmable Gate Array*). En los artículos que se anexan en esta memoria se analiza la viabilidad de las distintas técnicas en comparación con el resto y se evalua su comportamiento en tiempo y consumo. Por otra parte, como se ha comentado en el Apartado anterior, el sistema completo de detección de ojos propuesto en esta Tesis ha sido diseñado para poder obtener las imágenes del ojo a distancia. Este factor es especialmente crítico en el caso de la identificación por iris en zonas de paso donde no se pide al usuario que pare su marcha, obligando a trabajar con un sensor con resolución suficiente para capturar imágenes con una calidad mínima que permita la identificación desde una distancia de 1 a 2 metros del sujeto.

La metodología a seguir para la implementación de este sistema consta de tres hilos de trabajo, que se han ejecutado en paralelo durante la mayor parte del tiempo del desarrollo de esta Tesis Doctoral:

- En un primer hilo se aborda el estudio de las distintas técnicas de visión artificial que tienen como finalidad la detección del rostro y/o los ojos para conocer su funcionamiento e identificar ventajas y desventajas. Es importante destacar que estamos hablando de una fase que se ha prolongado durante todo el periodo de realización de esta Tesis Doctoral. Si

inicialmente las opciones que podían ofrecer alta velocidad y tratar con imágenes de alta resolución no solían basarse en redes neuronales, sino que empleaban aproximaciones más clásicas, esto cambiará radicalmente con la propuesta de soluciones más lígeras de la popular estructura convolucional YOLO (*You Only Look Once*) (Redmon et al, 2016). Como se analiza con cierta profundidad en Ruiz-Beltrán et al (2023b), las primeras implementaciones que podrían considerarse válidas se basan en la Tiny-YOLO v3 y datan del año 2018. Habrá que esperar unos años para que ofrezcan resultados que puedan ser interesantes para nuestro marco de aplicación (Oh et al, 2020; Zhang et al, 2021; Esen et al, 2021). Usando estas últimas propuestas como base se implementará la última versión de nuestro sistema de detección de ojos.

- El segundo hilo corre en paralelo con la primera, y ha supuesto poner en funcionamiento una plataforma hardware que permita capturar las subimágenes de ojos con la calidad suficiente para conseguir la identificación por reconocimiento de iris a distancia y con personas en movimiento. Al igual que ocurre al describir la fase anterior, han sido numerosos los cambios experimentados en la tecnología en relativamente poco tiempo. Decididos a implementar toda la parte crítica del sistema en FPGA, desde el inicio nos decantamos por emplear un AP (*All Programmable*) SoC (*System-On-Chip*). Estas plataformas incluían, junto a la parte programable (la FPGA), una parte software (microprocesador) que permitía implementar el control global de la actuación del sistema, incluyendo su conexión con el exterior (parcialmente con el sensor, pues el *core* que finalmente lee los datos se sintetiza en la FPGA, como con el módulo encargado de procesar las imágenes de ojos para obtener el patrón del iris y proceder al reconocimiento). Después de haber probado con distintas arquitecturas, nuestra propuesta se basará en la plataforma Zynq Ultrascale+ de AMD/Xilinx, un MPSoC que cuenta con microprocesador, GPU y FPGA en el mismo empaquetado. De las tres fases, esta segunda es la que se cerró en primer lugar, y, con cambios menores, se lleva trabajando con un mismo hardware durante los últimos dos años.

- El tercero de los hilos supone el diseño, implementación y evaluación de las distintas propuestas que se han ido generando a lo largo del desarrollo de esta Tesis Doctoral. Las primeras implementaciones buscaron conseguir el requisito de la alta velocidad (más de 40 fotogramas por segundo (fps)) personalizando el algoritmo propuesto por Viola y Jones (2001). La descripción de esta propuesta se presenta en Ruiz-Beltrán et al (2022). Este sistema cumple los requisitos de alta velocidad de procesamiento y detecta los ojos presentes en las imágenes, pero adolece de dos problemas: presenta una alta detección de falsos negativos y no descarta por si mismo aquellas detecciones que tienen poco contraste. Ambos problemas implican que el flujo de detecciones, generado cuando una persona cruza frente a nuestro sistema, sea excesivamente alto (si la persona tarda unos 2-3 segundos en cruzar y se capturan más de 40 fps, el número de ojos detectados podría rondar los 250). Este número satura los siguientes módulos encargados de obtener el patrón de iris o proceder al reconocimiento. La solución consistió en incluir en la arquitectura un estimador del contraste de la subimagen detectada, pues muchas de estas detecciones se asocian a ojos capturados muy lejos o cerca del punto de enfoque del sensor de imagen. Las subimágenes de ojos que no ofrecen el nivel de contraste exigido podían ser eliminados por nuestro propio sistema (Ruiz-Beltrán et al, 2023a). Esta propuesta se integra correctamente con el resto de la arquitectura, reduciendo este flujo de subimágenes de salida. Sin embargo, su inclusión supone reducir la velocidad de procesamiento (que sigue siendo, en cualquier caso, superior a los 40 fps), pero no reducir la alta tasa de falsos positivos. La última tarea abordada en este hilo de trabajo ha sido el diseño e implementación de un detector de ojos basado en la red neuronal Tiny YOLO v3 (Ruiz-Beltrán et al, 2023b). Esta tarea ha

supuesto cambios importantes en la forma de desplegar el software en la plataforma (por ejemplo, de usar versiones *bare metal*[1] en la parte software, se ha pasado a depender de la instalación de un PetaLinux[2]).

Contribuciones

El despliegue de un sistema completo de identificación remota del iris puede facilitarse si se minimiza el peso y consumo de energía de los dispositivos empleados, sin que esto suponga, en ningún caso, reducir la eficiencia en términos de velocidad de procesamiento y rendimiento. Además, dado el gran número de ojos que pueden ser detectados cuando el sistema trabaja con personas en movimiento, es importante que sólo se procesen aquellos fotogramas que contengan imágenes de ojos correctamente enfocados. Para su integración en este tipo de sistemas, esta Tesis Doctoral describe un par de implementaciones de detectores de ojos, basadas en la paralelización del popular algoritmo de Viola y Jones (2001) y en el uso de la CNN Tiny-YOLO v3, y satisfactoriamente empotradas en un MPSoC. El resultado a conseguir se puede definir, básicamente, como una cámara inteligente, que devuelva como salida imágenes de alta resolución que contienen ojos correctamente enfocados. En resumen, las contribuciones de esta Tesis Doctoral son:

- La introducción de una arquitectura completa para la detección de ojos correctamente enfocados, sintetizada en un MPSoC, y diseñada para soportar el reconocimiento del iris a distancia. El marco incluye todos los módulos necesarios para el redimensionamiento de la imagen y el filtrado de paso alto, la detección de ojos y el recorte final de las imágenes de los ojos a partir de la imagen de entrada original. Como se ha comentado en distintos puntos de este Resumen, se han implementado y evaluado dos opciones. La primera de ellas sigue una aproximación más clásica a la resolución de un problema en visión artificial, y supone la paralelización del algoritmo propuesto por Viola y Jones (2001). El enfoque propuesto estima la región del ojo directamente, sin requerir la localización del rostro humano, asumiendo que la apariencia y geometría del ojo es distinguible y los extractores de características, como el basado en características tipo Haar, pueden resolver con éxito el problema. Nuestro esfuerzo ha consistido en diseñar un enfoque de procesamiento de imágenes altamente paralelo y de una sola pasada, que enlaza íntimamente los procesos de caracterización y clasificación. El paso de clasificación en el enfoque Viola-Jones fue originalmente diseñado para ser ejecutado secuencialmente, priorizando que una respuesta negativa de un clasificador débil permitiera abortar el resto del proceso de identificación. Esta idea de reducir el número de clasificadores débiles evaluados también es inherente a la implementación del clasificador en forma de árbol de decisión. En nuestro diseño, todos los clasificadores débiles son siempre evaluados, pero este no es el problema para alcanzar un alto valor de eficiencia ya que estas evaluaciones son todas ellas abordadas en paralelo. Es importante señalar que este clasificador puede procesar hasta 752 fps, sin reducir la tasa de éxito del enfoque Viola-Jones para la detección de ojos. La segunda propuesta de detector de ojos se basa en el uso de redes neuronales convolucionales. Como se ha comentado, para conseguir una implementación rápida, nos decidimos por implementar una de las versiones lígeras de YOLO, una CNN para detección de objetos que no asume la existencia inicial de un conjunto de propuestas de regiones (Redmon et al, 2016). Desde su creación, YOLO ha evolucionado a través de muchas iteraciones (YOLOv1, YOLO9000, YOLOv3, PP-YOLO,...), pero una de las versiones más populares para ser sintetizada en FPGA ha sido

[1] La programación *bare metal* es un tipo de programación de bajo nivel que se codifica en un sistema a nivel de hardware. Funciona sin capa de abstracción ni sistema operativo (SO).

[2] https://docs.xilinx.com/r/en-US/ug1144-petalinux-tools-reference-guide/Introduction

la Tiny-YOLOv3 (Redmon y Farhadi, 2018). Para acelerar el proceso de descripción del hardware, Tiny-YOLOv3 se ha sintetizado en la FPGA utilizando la Unidad de Aprendizaje Profundo (*Deep Learning Unit*, DPU) ofrecida por AMD/Xilinx. La ventaja de usar DPU reside en su capacidad para acelerar, entre otros, el cálculo convolucional y de la función de activación, con parámetros que son configurables.

- A ambos esquemas de clasificación se le ha dotado de la capacidad para descartar imágenes que contuvieran ojos no enfocados, siguiendo siempre la hipótesis de que, como señala Daugman (2004), el efecto del desenfoque consiste principalmente en atenuar las frecuencias más altas de la imagen. En la primera de las propuestas de detector, para dotar de esta capacidad de descarte a la arquitectura se ha propuesto e implementado un módulo que estima el contraste de la imagen, lo que permite descartar las imágenes incorrectamente enfocadas. En la segunda de las propuestas, basada en el uso de CNN, se ha planteado un método sencillo pero eficaz para conseguir este descarte, basado en el uso no sólo de la imagen capturada sino también de una versión filtrada de paso alto de la misma como entradas al sistema. Las pruebas realizadas demuestran que, al considerar esta información en los pasos de entrenamiento y clasificación, el sistema diseñado detecta únicamente imágenes oculares enfocadas.

Resulta además importante destacar la extensa evaluación de las propuestas emanadas de esta Tesis Doctoral en un entorno real de trabajo, con hardware seleccionado o diseñado a medida para este escenario. El sistema ha demostrado su capacidad para captar a distancia las imágenes oculares de personas en movimiento, descartando las afectadas por desenfoque.

Proyectos y contratos de transferencia

Como se ha comentado al describir los Objetivos de esta Tesis, su génesis entroca en parte con el interés mostrado por la empresa española SHS Consultores SL por los resultados obtenidos. Sin embargo, el trabajo ha sido también ha sido importante en otros contratos de transferencia o proyectos de investigación. A continuación se analizan brevemente estos trabajos, en los que he estado implicado en mayor o menor medida.

- **LYNX**. El objetivo del proyecto LYNX era el seguimiento e identificación de especies protegidas en entornos de especial interés medioambiental mediante la aplicación de tecnologías que no interfieran en la actividad natural de la fauna. Se buscaba así facilitar el seguimiento de estas especies mediante técnicas no invasivas y que interfirieran lo mínimo posible con el animal y su entorno, mejorando las actuales soluciones basadas en anillado físico o conteo manual mediante técnicas de fototrampeo. Debido al elevado grado de amenaza que presenta en la actualidad y su alto valor biológico, el objeto fundamental de este proyecto era el Lince Ibérico (Lynx Pardinus), especie cuya principal población se concentra en los entornos de Sierra Morena y en el entorno del Parque Natural de Doñana, situado en la Comunidad Autónoma de Andalucía. En este escenario, la colaboración de nuestro grupo con la empresa Magtel Operaciones se encuadraba en dos de las tareas más ambiciosas del proyecto: el diseño y desarrollo de una cámara de fototrampeo inteligente, capaz de filtrar, del conjunto de imágenes capturadas, aquellas que, con alta probabilidad, contenían un lince; y la implementación de un algoritmo que permitiera la identificación unívoca de los animales detectados. En paralelo, el diseño de la cámara debería resolver algunos de los problemas que presenta actualmente su despliegue, permitiendo la recogida de los datos en remoto.

Estos dos objetivos engarzaban perfectamente en las líneas de trabajo que nuestro grupo venía desarrollando en los últimos años: la visión empotrada y el procesamiento de imagen. Por tanto, el trabajo se abordó con dos equipos de expertos, que desarrollaran su labor en continuo contacto para que las necesidades de cada uno de los grupos fueran suplidas por el otro. Así, en la primera línea de trabajo, la cámara desarrollada usando tecnología MPSoC, fue dotada con el disparo desde un sensor térmico y un hardware a medida, sintetizado en la parte lógica, que permite emplear características de textura LBP (*Local Binary Pattern*) y un clasificador SVM (*Support Vector Machine*) para discriminar la textura del lince del resto de la escena. La versión inicial de los algoritmos de caracterización y clasificación, y el propio entrenamiento del clasificador, fueron proporcionados por el grupo de procesamiento de imagen. En la segunda línea de trabajo, la identificación individual de linces empleó el patrón de manchas de éstos como fuente de información, y utiliza características visuales, junto a información estructurada de distribución espacial, para caracterizar específicamente a cada animal. Las cámaras desarrolladas contaron, finalmente, con la capacidad de conectarse a redes de telefonía móvil para, de esta forma, poder acceder a un servidor central, desde el que se recogen datos sobre el estado de las mismas, así como las propias imágenes capturadas. Todo el sistema fue desplegado en entornos reales, con cámaras ubicadas en Córdoba o Málaga.

- **HIRIS**. Nuestro grupo de investigación había trabajado anteriormente con la empresa SHS Consultores en el diseño de un sistema de reconocimiento por identificación de iris. Este sistema trabajaba a distancia y con personas en movimiento y, aunque ya empleaba tecnología FPGA para el diseño del detector de ojos, tanto la resolución del sensor como la velocidad de captura no eran las adecuadas para asegurar el correcto funcionamiento del mismo. El proyecto HIRIS persiguía incluir las fases iniciales del proceso de reconocimiento (en concreto la detección del ojo) en tecnología MPSoC, usando un sensor de 16 Mpx para la captura de la imagen y buscando velocidades de procesamiento que permitieran obtener el máximo rendimiento del sensor. Además, en nuestro caso, suponía una primera toma de contacto con la aplicación de algoritmos antifraude a las imágenes de ojos detectadas. El trabajo en este proyecto dió lugar a las dos primeras publicaciones que forman parte de este compendio (Ruiz-Beltrán et al, 2022, 2023a).

- **DIMAS**. El objetivo del proyecto DIMAS es el diseño y desarrollo de un sistema de reconocimiento de iris en movimiento, cuyas principales características sean ser embebido en tecnología MPSoC y ser mejorado con aquellas características relacionadas con la defensa frente a ataques de presentación y comportamiento fiable. Liderado por SHS Consultores SL y actualmente en curso, el proyecto supone una oportunidad para adquirir nuevos conocimientos relacionados con los retos actuales en el reconocimiento del iris, pero también es una nueva oportunidad para transferir conocimiento científico y tecnológico de nuestro grupo de investigación a la industria. El resultado esperado será una versión mejorada del producto AIRIM, ofrecido actualmente por SHS Consultores. Esta versión mejorada incluirá el sensor de 16 Mpx que se está utilizando en las últimas versiones del sistema, óptica adaptativa, detección de ojos usando CNN y defensa frente a ataques por el uso de lentes de contacto texturizadas. El prototipo será evaluado en un piloto a gran escala en las instalaciones de los socios de este Consorcio en Sevilla y Málaga. La tercera de las publicaciones incluídas en la presente Tesis es el fruto inicial del trabajo llevado a cabo en este proyecto (Ruiz-Beltrán et al, 2023b).

Aparte de estos proyectos, en los que he estado contratado y que guardan estrecha relación con esta Tesis Doctoral, el trabajo llevado a cabo en esta Tesis ha sido portado parcialmente

a diseños empleados en otros proyectos, en los que he formado parte del equipo de trabajo o he estado también contratado. En concreto, estos proyectos son el MIRoN Integrated Technical Project financiado a su vez en el marco del EU RobMoSys project (H20202-732410), el proyecto RTI2018-099522-B-C41, financiado por el Ministerio de Ciencia, Innovación y Universidades y FEDER funds, y los PDC2022-133597-C42, TED2021-131739B-C21 y PID2022-137344OB-C32, financiados por MCIN/AEI/10.13039/501100011033 y por la Unión Europea a través de los programas NextGenerationEU/PRTR (para los dos primeros proyectos citados), y "ERDF A way of making Europe" (para el tercero).

Revistas

- Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Marfil, R.; Bandera, A. **FPGA-Based CNN for Eye Detection in an Iris Recognition at a Distance System**. Electronics 2023, 12, 4713. https://doi.org/10.3390/electronics12224713
- Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Marfil, R.; Bandera, A. **Real-Time Embedded Eye Image Defocus Estimation for Iris Biometrics**. Sensors 2023, 23, 7491. https://doi.org/10.3390/s23177491
- Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Sánchez-Pedraza, A.; Rodríguez-Fernández, J.A.; Bandera, A. **Real-time Embedded Eye Detection System**. Expert Systems with Applications 2022, 194, 116505. https://doi.org/10.1016/j.eswa.2022.116505

Conferencias

- Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Rodríguez-Fernández, J.A.; Bandera, A.; Sánchez-Pedraza, A. **Real-time iris image quality evaluation implemented in Ultrascale MPSoC**. In Proceedings of the XXXVIII Conference on Design of Circuits and Integrated Systems (DCIS 2023), Málaga, Spain, 15-17 November 2023
- Ruiz-Beltrán, C.A.; Bandera, A.; González-García, M.; Marfil, R. **Real-time embedded eye detection and analysis framework**. In Proceedings of the XXXVIII Conference on Design of Circuits and Integrated Systems (DCIS 2023), Málaga, Spain, 15-17 November 2023
- Romero-Garcés, A.; Ruiz-Beltrán, C.; Marfil, R.; Bandera, A. **Lightweight Cosmetic Contact Lens Detection System for Iris Recognition at a Distance**. In Proceedings of the 18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023), Salamanca, Spain, 5–7 September 2023; García Bringas, P., Pérez García, H., Martínez de Pisón, F.J., Martínez Álvarez, F., Troncoso Lora, A., Herrero, Á., Calvo Rolle, J.L., Quintián, H., Corchado, E., Eds.; Springer: Cham, Switzerland, 2023; pp. 246–255

Estructura de la Tesis

Al presentarse esta Tesis Doctoral como compendio de publicaciones, y una vez se completa el presente resumen en castellano, el cuerpo de la misma se organiza en dos partes diferenciadas. En la primera de las partes (*Thesis Description*) se hace una breve introducción en inglés de la Tesis Doctoral. Su organización es muy similar a la que adopta el presente resumen. La segunda parte (*Included papers*) incluye los tres artículos que dan forma al trabajo realizado en la Tesis

Doctoral. A continuación se realiza un esbozo de los artículos incluidos en esta segunda parte de la Tesis Doctoral, así como las contribuciones del autor en cada uno de ellos.

Artículo A: Sintetizado de una versión paralelizada del algoritmo Viola-Jones en un MPSoC

Descripción: Este artículo presenta una solución embebida basada en hardware para la detección de ojos en tiempo real. Desde un punto de vista algorítmico, la propuesta supone un rediseño del popular método Viola-Jones, consiguiéndose una implementación totalmente paralelizada del procesamiento de imágenes (caracterización y clasificación), que se ejecuta en una sola pasada. Sintetizada e implementada en un MPSoC, esta propuesta procesa más de 88 imágenes por segundo, tardando el clasificador menos de 2 ms por imagen. La validación experimental se abordó con éxito en un sistema de identificación por reconocimiento del iris real, en el que las personas pasan sin detenerse y andando a una cadencia de paso normal. En este caso, el prototipo emplea un sensor de imagen digital CMOS que proporciona imágenes de 16 Mpx, y genera como salida un flujo de ojos detectados como imágenes de 640 x 480. Los experimentos para determinar la precisión del sistema propuesto emplean la base de datos CASIA-Iris-distance V4, obteniéndose con ella un valor de acierto en la detección del 100%.

Contribución del autor: Existía una versión inicial del algoritmo, que trabajaba con un sensor de 5 Mpx a una velocidad que rondaba los 25 fotogramas por segundo. Mi trabajo consistió en rediseñar la arquitectura y sintetizarla en el MPSoC UltraScale+ de AMD/Xilinx hasta conseguir los valores que se han descrito anteriormente.

Artículo B: Incorporación de un sistema de evaluación del contraste a la arquitectura hardware

Descripción: Este trabajo describe la integración, en la parte lógica de la arquitectura ya diseñada en el trabajo anterior, del bloque funcional para evaluar el nivel de desenfoque de las imágenes capturadas. De esta forma, el sistema podrá descartar las imágenes que no tengan la calidad de enfoque requerida en los pasos de procesamiento posteriores. Si anteriormente se había usado Vivado como entorno de trabajo, la nueva propuesta se implementó usando Vitis High Level Synthesis (HLS), consiguiéndose procesar más de 57 imágenes por segundo. Utilizando, para su validación, una versión ampliada de la base de datos CASIA-Iris-distance V4, la evaluación experimental muestra que el marco propuesto es capaz de descartar con éxito imágenes de ojos desenfocados. Pero lo más relevante es que, en una implementación real, esta propuesta permite descartar hasta el 97% de las imágenes de ojos desenfocados, que no tendrán que ser procesadas por los bloques de segmentación y extracción de patrones normalizados del iris.

Contribución del autor: El trabajo técnico propuesto en este artículo ha sido completamente llevado a cabo por el autor de esta Tesis, trabajando sobre la base de la implementación inicial de estimación de contraste usando máscaras de convolución propuestas en la literatura por distintos autores.

Artículo C: Sintetizado de la red neuronal YOLOv3 Tiny en el MPSoC

Descripción: En este artículo, se describe el sintetizado de un sistema de detección de ojos basado en la CNN YOLOV3 Tiny en un MPSoC Zynq XCZU4EV UltraScale+. Este detector de ojos no solo es capaz de procesar imágenes de alta resolución capturadas a gran velocidad, sino que descarta aquellas que están seriamente afectadas por desenfoque. Para ello, la red se

entrena únicamente con imágenes de ojos correctamente enfocados. Además, aprovechando la ventaja de las redes neuronales de poder trabajar con entradas multicanal, las entradas de la CNN serán la imagen de nivel de gris y una versión filtrada de paso alto, utilizada normalmente para determinar si el iris está enfocado o no. El sistema completo sintetiza otros núcleos e implementa la CNN utilizando la denominada Unidad de Procesado de Aprendizaje Profundo (DPU), el IP-core propuesto por AMD/Xilinx. En comparación con diseños anteriores que sintetizan CNNs en FPGA, la DPU optimiza las típicas funciones que usan los algoritmos de aprendizaje profundo, lo que le permite acelerar la inferencia de la red neuronal. Como en los casos anteriores, la validación experimental se ha abordado con éxito en un escenario del mundo real, trabajando correctamente con personas en movimiento, y demostrando que es posible detectar únicamente las imágenes de ojos que están enfocadas. En las pruebas llevadas a cabo, el prototipo descarta correctamente hasta el 95% de los ojos presentes en las imágenes de entrada por no estar correctamente enfocados.

Contribución del autor: En esta propuesta el autor ha sido el responsable de todas las etapas del proceso de diseño, implementación, y evaluación. Se han analizado distintas implementaciones de CNN, decidiéndose finalmente por la YOLO v3 Tiny. Se ha diseñado la arquitectura completa del sistema, integrando en ella el core DPU de Xilinx/AMD.

Conclusiones y Trabajo Futuro

Las dos aproximaciones propuestas en esta Tesis permiten que la fase de detección de ojos funcione a una velocidad mayor que la del sensor de imagen empleado en la implementación real (el Teledyne e2v EMERALD 16MP), con lo que se pueden procesar los 47 fotogramas que el sensor proporciona por segundo. Esto permite que el sistema real, que trabaja con una profundidad de campo muy pequeña, pueda sin embargo identificar correctamente a las personas que pasan andando por la zona de paso. La implementación basada en la YOLOV3 Tiny es muy robusta, con un porcentaje de detección cercano al 100% en el entorno real y una prácticamente nula detección de falsos positivos.

En las últimas implementaciones se despliega en la parte software del MPSoC un Petalinux. La existencia de este sistema operativo debería favorecer el empleo de todos los recursos que, junto a la parte lógica FPGA, ofrece el MPSoC. Usando la ARM MALI 400MP GPU y las Unidades de Procesamiento en Tiempo Real (*Real-time processing unit*, (RPU)) de la UltraScale+, el trabajo futuro debería centrarse tanto en extraer el patrón normalizado del iris como en implementar los algoritmos de detección de fraude en la misma arquitectura sintetizada en el MPSoC.

Referencias

References

Viola, P.; Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 (pp. I–I). volume 1. doi:10.1109/CVPR.2001.990517.

Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Marfil, R.; Bandera, A. (2023b). FPGA-Based CNN for Eye Detection in an Iris Recognition at a Distance System. Electronics 12, 4713. https://doi.org/10.3390/electronics12224713

Oh, S., You, J.-H., and Kim, Y.-K. (2020). Implementation of compressed yolov3-tiny on FPGA-SOC. In 2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), pages 1–4

Zhang, H.; Jiang, J.; Fu, Y.; Chang, Y. (2021). YOLOv3-tiny object detection soc based on FPGA platform. In 2021 6th International Conference on Integrated Circuits and Microsystems (ICICM), pages 291–294

Esen, F.; Degirmenci, A.; Karal, O. (2021). Implementation of the object detection algorithm (YOLOv3) on FPGA. In 2021 Innovations in Intelligent Systems and Applications Conference (ASYU), pages 1–6

Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Sánchez-Pedraza, A.; Rodríguez-Fernández, J.A.; Bandera, A. (2022) Real-time Embedded Eye Detection System. Expert Systems with Applications 194, 116505. https://doi.org/10.1016/j.eswa.2022.116505

Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Marfil, R.; Bandera, A. (2023a) Real-Time Embedded Eye Image Defocus Estimation for Iris Biometrics. Sensors 23, 7491. https://doi.org/10.3390/s23177491

Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788. doi:10.1109/CVPR.2016.91

Redmon, J.; Farhadi, A. (2018) YOLOv3: An Incremental Improvement. CoRR, abs/1804.02767

Daugman, J. (2004) How iris recognition works. IEEE Transactions on Circuits and Systems for Video Technology 14, 21–30

Part I

Thesis Description

Introduction

Everyday many living beings use the sight sense to obtain information from the surrounding environment, focusing on interesting objects such as resources or turning the attention to visual clues that can provide useful data. It is a ubiquitous ability that we as humans use throughout our entire live as long our eyes are healthy. The apparent simplicity to solve problems by means of our sight sense motivated the search of artificial alternatives. However, the algorithms to emulate the human vision are complex. Furthermore, there is a big challenge for reliability. Thus, for example, we are able to identify an object overcoming problems such as varying light conditions, confusing background, varying perspective and many other variables that must be mitigated or accounted in an artificial vision system.

Artificial vision, commonly known as computer vision, is an interdisciplinary field that allows machines to interpret and make decisions based on visual data. This technology aims to emulate human vision by using computers to understand and extract useful information from images or videos. A key application of this technology is image recognition, where computer algorithms are developed to identify and classify objects within images. This is useful in many fields, such as biometrics, security, industrial automation, autonomous vehicles, medical imaging and many others, with particular interest in applications that require continuous observation. Researchers' contributions allowed machine vision to move out of the laboratory and into concrete applications in real-life scenarios. Clearly, Deep learning has had a decisive influence on this. This has led to commercial solutions, such as surveillance systems, quality control in factories or autonomous navigation systems, among others. Many of these systems require as a first step the detection of a specific Region Of Interest (ROI), be it a person or a face, a car or a boat, a possible problem on a mammogram... Significantly, in a large number of these applications, this ROI is the person himself, the face, or a relevant element of it. This is the case in iris and face recognition, gaze tracking, behaviour and expression interpretation, natural language processing or driver fatigue detection. In many of these scenarios, the eyes contain sufficient information and therefore an eye detection can be the first step to accomplish. Clear examples are iris recognition for biometric identification or drowsiness detection in driver monitoring systems.

Background and motivation

In 2008, the company SHS Consultores SL contacted the then main researcher in charge of the Integrated Systems Engineering group at the University of Malaga. The proposal was to collaborate with them on a CTIC project funded by the Spanish Government to develop solutions for the mass identification of users in transit areas (the INTEGRA project). The INTEGRA project proposed the development of different functional prototypes, including the development of an access control system for iris identification. The most relevant feature of this prototype was the ability to capture iris images without requiring the active participation of the people crossing

the access, so that they could walk in front of the capture cameras. This required the capture system to use very low exposure times which, together with the fact that the aperture had to be reduced to achieve focused images, meant that there was very little illumination entering the sensor. This scenario constituted a relevant challenge for the research group in this project, which forced them to design a capture architecture that synchronised the triggering of different high-resolution cameras with that of a high-power infrared illumination source. The safety issues involved in the use of this type of illumination meant that the triggering time of the spotlight had to be reduced as much as possible, without losing illumination while the exposure on the sensor was active. To prevent the machine vision system from continuously capturing images, thus saturating the subsequent processing modules, the system incorporated a stereo vision system. This system detected the areas of the image where the eyes of people approaching the access were located, and which allowed the cameras to trigger only when there was a person in the area where there was a high probability of capturing the sub-images containing the iris information with the appropriate focus and resolution characteristics. This module was present in early versions of the designed prototype, however its need was eliminated when, with the development of optimised image processing algorithms, the detection of eye zones in the input images could be carried out very efficiently. The design and implementation of these algorithms was the second task to be tackled by the research group in the framework of that project. The idea was to optimise the detection process based on Haar-type visual features and classifiers based on the AdaBoost strategy (Viola and Jones, 2001). The OpenCV library was employed to provide this framework, as the algorithm can run in the computer interfacing each one of the CCD cameras deployed in the prototype.

The original contract initiated in 2008 was completed with an addendum in 2012. Thus, once the system configuration had been completed and correctly evaluated, the research group's work on the project focused, from January 2012 onwards, on the integration of the different software components that coexisted in the application. Thus, a final application was developed based on two processing elements, each of which managed the information provided by two capture cameras, and in which one of the elements acted as the master and provided the final output of the complete system. This layout of computers and cameras was fully modular, and could be expanded in the future to accommodate a larger number of processing elements and cameras if a larger field of view was to be covered. In this configuration, once a person was detected, the capture cameras were triggered and possible eyes were detected in the captured images. Since many of these images might not contain any valid eyes, a fast discard test algorithm was designed, which allowed a considerable reduction of the set of images to be passed to the recogniser without, however, losing any valid images. The set of valid images was then analysed by the recogniser, which returned an identifier if the person was recognised or a special character indicating that the person was not in the system. The system allowed the entire capture and recognition process to take no more than two seconds, thus adhering to the requirements initially set for the application.

At that time, and for the next few years, cameras with CMOS technology did not provide adequate image quality to solve the identification problem. There were no MPSoC platforms that could efficiently integrate FPGA and microprocessor, nor was there access to image sensors that offered a convenient interface. The work of designing and shaping a camera that could work with the appropriate speed or resolution requirements, was beyond the scope of this first project, although the results would be improved in a second CTIC project (the DIBIMAS one, started in 2010). As the most significant milestone, the final result of these projects constituted the germ of a product currently developed and offered by the company SHS Consultores SL. The AIRIM system (Automatic System for Iris Recognition in Motion over long distances) is offered as a solution capable of identifying users who continuously pass through a virtual corridor, without the need to stop.

The work with SHS Consultores SL continued in 2014, in the framework of a RetosColaboración project. The FGACCESS project aims to develop a novel high-security automatic access control system that is unnoticed by users. It was thus a continuation of the work carried out in the INTEGRA and DIBIMAS projects, and a continuation of the collaboration in transfer with SHS Consultores SL. The group has gained experience in the use of All-Programmable on-chip platforms (AP SoC), which include both logic (FPGA) and software (microprocessor) parts. Thus, the participation of our research group involved synthesising the iris (eye) detector in the AP SoC, the system would use a raw image sensor to achieve the capture task. This would allow the development of a custom camera, which would not provide high-resolution images but only sub-images that potentially included a ROI, in this case, the eyes of people passing through the access control. In this way, the entire system was greatly relieved of computational work, as the camera itself did not only the cropping of the areas with eyes, but also the filtering of images that did not contain eyes, which were no longer processed by the rest of the system. In addition, to reduce the need for the high illumination required by the system developed in the INTEGRA project, the company worked with Anafocus to design high-sensitivity infrared sensors. These sensors, modified versions of the Lince5M already marketed by Anafocus, allowed the system to work with very low infrared illumination, completely eliminating security problems.

The architecture initially proposed within the framework of the FGACCESS project comprised (i) a system to detect the presence of the person, using sensors that will provide images of the environment with low resolution but covering a high field of vision; (ii) a motorization system, capable of moving the capture set according to this information; and (iii) a second capture system, responsible for providing the irises of the people passing in front of the cameras with sufficient quality to allow their recognition, this being its main verification test. The set is joined by (iv) a control system and information concentrator; and (v) the lighting system. However, as the project progressed, this architecture evolved towards a simpler scheme, in which the detection, motorization and concentrator modules have been eliminated. Thus, the scheme would be a single detection and capture module, organized as a sensor array. The array distribution will allow eliminating the motorization module, as the necessary field of view will be covered without having to move the sensor column. Each of these sensors is connected to an FPGA, which will be in charge of implementing the front-end of each sensor, configuring it and capturing its information. As previously mentioned, the FPGA chosen for this implementation will be the one included in the AP SoC Zynq-7020. In that case, we will therefore have, together with the logic part, a microprocessor, in this case a Dual ARM® Cortex-A9. This made it possible to provide this front-end with greater versatility, gaining in responsibility in the final architecture of the project and freeing the Concentrator of functionalities. In order to include the eye detector in the AP SoC, the Zynq-7020 had to be replaced by a 7030. The system thus integrated allowed to capture about 19-20 fps.

The detection algorithm is still based on Haar features and an AdaBoost classifier. For integration into the logic part of the AP SoC, the proposal was parallelized, with detection being carried out at a single scale and originally using RTL (register-transfer level) encoding. In order to speed up the design, HLS (high-level synthesis) was used, and the lack of FPGA resources, necessary to achieve a higher processing speed, was solved by including a second FPGA in the design (an Artix-7 from Xilinx).

In 2017 the collaboration in a new project with SHS Consultores SL starts. My participation in this whole story begins within the framework of this project. The goal of the HIRIS project was to integrate to the design a higher resolution sensor, the 16Mpx Anafocus EMERALD, and get to increase the processing speed to be able to process the 47 fps that this sensor can provide. The work on this project resulted in the first two publications that are part of this compendium (Ruiz-Beltrán et al., 2022, 2023b). The eye detector is still based on the parallelized approach

of the Viola and Jones algorithm (Viola and Jones, 2001), optimized using HLS to achieve very high processing speeds. The problem that many of the captured images were of out-of-focus images is solved by adding a contrast estimation module to the system, which will allow these invalid images to be discarded. The processing speed remains above 47 fps. We can state that the scheme of our algorithm follows the traditional one in computer vision, with its characterization and classification phases, totally merged in our case to achieve the required processing speed. Although work had already been done with Convolutional Neural Networks (CNN), synthesizing them on the Multi-Processor System-on-Chip (MPSoC) platforms being used was not an obvious job. In addition, the processing speed we needed was not within the reach of the CNN implementations we were familiar with.

The work in HIRIS is continued in the DIMAS project, a new collaborative project with the company SHS Consultores SL. The objective of the DIMAS project is the design and development of a iris recognition at a distance system, whose main features are to be embedded in MPSoC technology and to be enhanced with those features related to defense against presentation attacks and reliable behavior, those attacks can vary from simple techniques such as using a printed face or more complex such as using textured contact lenses to impersonate another user. Led by SHS Consultores SL and currently in progress, the project is an opportunity to acquire new knowledge related to the current challenges in iris recognition, but it is also a new opportunity to transfer scientific and technological knowledge from our research group to the industry. The expected result will be an improved version of the AIRIM product, currently offered by SHS Consultores. This improved version will include the 16Mpx sensor that is being used in the latest versions of the system, adaptive optics, eye detection using CNN and defense against attacks by the use of textured contact lenses. CNNs are the current state-of-the-art solution to image-processing tasks. Unfortunately, the excellent recognition accuracy of CNNs comes at the cost of very high computational complexity, and one of the current challenges is managing the power, delay and physical size limitations of hardware solutions dedicated to accelerating their inference process. An eye detection system based on CNN was developed to work in the same Zynq XCZU4EV UltraScale+ multiprocessor system-on-chip (MPSoC). The complete system synthesizes other cores and implements CNN using the so-called Deep Learning Processor Unit (DPU), the intellectual property (IP) block released by AMD/Xilinx. Compared to previous hardware designs for implementing FPGA-based CNNs, the DPU IP supports extensive deep learning core functions, and developers can leverage DPUs to conveniently accelerate CNN inference. This new proposal is evaluated against previous versions of the system, based on Haar features and an AdaBoost classifier, the use of CNN allows for greater robustness, maintaining the high rate of true positives but drastically reducing the rate of false positives. Moreover, exploiting the neural network's advantage of being able to work with multi-channel input, the inputs to the CNN will be the grey level image and a high-pass filtered version, typically used to determine whether the iris is in focus or not. Thus, the detector only provides as output images eyes that are in focus, discarding all those seriously affected by defocus blur, this work resulted in the third paper that shapes this compendium Thesis (Ruiz-Beltrán et al., 2023a).

Objectives

The work carried out in this PhD Thesis aims to develop and implement an eye detector that works in real time using artificial vision. Specifically, the objective is to detect the iris of a subject moving along a given trajectory. This is a case of iris detection at a distance (IAAD). For this application, the performance and speed of the system are essential, as the detection and analysis of the eye must be performed using high-resolution, undistorted images at a high

frame rate. To achieve this goal, additional hardware and software modules have been developed to analyse the detected images and assess their quality, e.g. to determine the contrast level of the region of interest. The system capability to provide enough and accurate eye images from a distance between 1 to 2 meters and with the subject walking is also evaluated, because this capacity is fundamental to allow reliable identification within the application scenario. Such captured images must have enough contrast and resolution to be further processed to extract information. In this application, the goal is the identification the iris pattern, which must have a minimum quality to be viable ($190 \sim 250$ pixels per centimeter).

A ROI must be detected and extracted from the original video stream. This ROI is an image that contains one of the eyes of the subject. This can be accomplished using heuristic techniques such as the algorithm based on Haar features and AdaBoost classifier proposed by Viola and Jones (2001). The task can also be accomplished using Deep Learning techniques such as CNN. In this work both techniques are explored, implemented and compared.

The proposed system is embedded because it allows to deploy multiple systems in parallel, mounted on a gate, to cover a wider detection area while maintaining a reasonable cost. Furthermore, the system will be used in an IAAD scenario imposing strict timing constraints since it is a real time application. Hardware solutions were explored such as microprocessor, FPGA (Field Programmable Gate Array) or GPU (Graphic Processing Unit) assessing their viability, timing and power consumption.

Contributions

The deployment of a complete remote iris identification system can be facilitated by minimising the weight and power consumption of the devices used, without reducing efficiency in terms of processing speed and performance. In addition, given the large number of eyes that can be detected when the system works with moving people, it is important that only frames containing images of properly focused eyes are processed. For integration into such systems, this PhD Thesis describes a couple of eye detector implementations, based on the parallelisation of the popular Viola and Jones (2001)'s algorithm and the use of the CNN Tiny-YOLO v3, and successfully embedded in an MPSoC. The result to be achieved can basically be defined as an intelligent camera, which returns as output high resolution images containing correctly focused eyes. In summary, the contributions of this PhD Thesis are:

- The introduction of a complete architecture for correctly focused eye detection, synthesised in an MPSoC, and designed to support remote iris recognition. The framework includes all the necessary modules for image resizing and high-pass filtering, eye detection and final cropping of the eye images from the original input image. As discussed at various points in this Summary, two options have been implemented and evaluated. The first one follows a more classical approach to solving a computer vision problem, and involves the parallelisation of the algorithm proposed by Viola and Jones (2001). The proposed approach estimates the eye region directly, without requiring the localisation of the human face, assuming that the appearance and geometry of the eye is distinguishable and feature extractors, such as the Haar-type feature extractor, can successfully solve the problem. Our effort has been to design a highly parallel, single-pass image processing approach that intimately links the characterisation and classification processes. The classification step in the Viola-Jones approach was originally designed to be executed sequentially, prioritising that a negative response from a weak classifier would allow the rest of the identification process to be aborted. This idea of reducing the number of weak classifiers evaluated is also inherent in the decision tree implementation of the classifier. In our design, all weak

classifiers are always evaluated in parallel to achieve a high efficiency value. It is important to note that this classifier can process up to 752 fps, without reducing the success rate of the Viola-Jones approach for eye detection. During the development, in the first step an AdaBoost classifier was trained using the CASIA database (Center for Biometrics and Security Research, 2010). Then, the preprocessing was modified to work with the video stream entirely on the FPGA. Moreover, the classifier was modified to compute all the stages in parallel to increase the throughput. Finally each classifier was modified to use Look Up Tables (LUT) to store the result of some mathematical operations resulting on an increase of the speed at the expense of memory. The entire image preprocessing and detection is carried out on the FPGA, the rest of the tasks such as camera configuration, image cropping and Ethernet connectivity are made on the ARM microprocessor within the Ultrascale+ MPSoC. The details of this work are explained on the first two publications in this Compendium (Ruiz-Beltrán et al., 2022, 2023a).

The second eye detector proposal is based on the use of convolutional neural networks. As discussed, in order to achieve a fast implementation, we decided to implement one of the lightweight versions of YOLO, a CNN for object detection that does not assume the initial existence of a set of proposed region proposals (Redmon et al., 2016). Since its creation, YOLO has evolved through many iterations (YOLOv1, YOLO9000, YOLOv3, PP-YOLO,...), but one of the most popular versions to be synthesised on FPGA has been the Tiny-YOLOv3 (Redmon and Farhadi, 2018).

Parallelism and ease of programming are possibly the features that justify the use of FPGA to accelerate CNN inference. In our implementation, to further accelerate the process of designing and synthesising a CNN in the logic part of the MPSoC, the AMD Deep Learning Processor Unit was used (DPU) (Zhu et al., 2020). The DPU is a programmable engine, implemented in the FPGA to accelerate neuronal network throughput (XILINX, 2023a). Thus, in our design, a Tiny YOLO-v3 eye detection was deployed in a Zynq UltraScale+ MPSoC XCZU4EV using a single DPU 1600 core with Softmax feature. The computational performance and acceleration effect of DPU where analysed. Xilinx Vitis AI was used as development environment (XILINX, 2023c). This work is further explained in our publication Ruiz-Beltrán et al. (2023b).

- Both classification schemes have been endowed with the ability to discard images containing unfocused eyes, always following the assumption that, as pointed out by Daugman (2004), the effect of defocus is mainly to attenuate the higher frequencies of the image. In the first of the detector proposals, in order to provide the architecture with this discarding capacity, a module has been proposed and implemented that estimates the contrast of the image, which allows incorrectly focused images to be discarded. In the second proposal, based on the use of CNN, a simple but effective method to achieve this discarding has been proposed, based on the use not only of the captured image but also of a high-pass filtered version of it as inputs to the system. Tests show that, by considering this information in the training and classification steps, the designed system detects only focused eye images.

It is also important to emphasise the extensive evaluation of the proposals arising from this PhD Thesis in a real working environment, with hardware selected or custom-designed for this scenario. The system has demonstrated its ability to remotely capture the eye images of people in movement, discarding those affected by defocusing.

Publications

The present PhD Thesis encompasses the following publications:

Journals

- Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Marfil, R.; Bandera, A. **FPGA-Based CNN for Eye Detection in an Iris Recognition at a Distance System**. Electronics 2023, 12, 4713. https://doi.org/10.3390/electronics12224713

- Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Marfil, R.; Bandera, A. **Real-Time Embedded Eye Image Defocus Estimation for Iris Biometrics**. Sensors 2023, 23, 7491. https://doi.org/10.3390/s23177491

- Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Sánchez-Pedraza, A.; Rodríguez-Fernández, J.A.; Bandera, A. **Real-time Embedded Eye Detection System**. Expert Systems with Applications 2022, 194, 116505. https://doi.org/10.1016/j.eswa.2022.116505

Conferences

- Ruiz-Beltrán, C.A.; Romero-Garcés, A.; González-García, M.; Rodríguez-Fernández, J.A.; Bandera, A.; Sánchez-Pedraza, A. **Real-time iris image quality evaluation implemented in Ultrascale MPSoC**. In Proceedings of the XXXVIII Conference on Design of Circuits and Integrated Systems (DCIS 2023), Málaga, Spain, 15-17 November 2023

- Ruiz-Beltrán, C.A.; Bandera, A.; González-García, M.; Marfil, R. **Real-time embedded eye detection and analysis framework**. In Proceedings of the XXXVIII Conference on Design of Circuits and Integrated Systems (DCIS 2023), Málaga, Spain, 15-17 November 2023

- Romero-Garcés, A.; Ruiz-Beltrán, C.; Marfil, R.; Bandera, A. **Lightweight Cosmetic Contact Lens Detection System for Iris Recognition at a Distance**. In Proceedings of the 18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023), Salamanca, Spain, 5–7 September 2023; García Bringas, P., Pérez García, H., Martínez de Pisón, F.J., Martínez Álvarez, F., Troncoso Lora, A., Herrero, Á., Calvo Rolle, J.L., Quintián, H., Corchado, E., Eds.; Springer: Cham, Switzerland, 2023; pp. 246–255

Thesis framework

This PhD Thesis is the result of five years of work by the author as a member of the Engineering of Integrated Systems (ISIS) research group [3], part of the Department of Electronic Technology of the University of Malaga. This research has been funded by various contracts, in the framework of which innovative work has been carried out, with a clear practical character. During this period, the author successfully completed the PhD Programme in Telecommunication Engineering (by the University of Malaga), coordinated by the Departments of Communication Engineering and Electronic Technology, where I obtained a strong background knowledge concerning the fundamental pillars of embedded vision. I also completed a three months research stay at the School

[3] https://www.uma.es/departamento-de-tecnologia-electronica/info/55617/dte_grupo_isis/

of Innovation, Design and Engineering, of the University of Mälardalen (Sweden), in 2022, under the supervision of Prof. Dr. Martin Ekstrom.

As mentioned in Section 1.1, the genesis of this PhD Thesis is partly due to the interest shown by the Spanish company SHS Consultores SL in the results obtained. However, the work has also been important in other transfer contracts or research projects. These works, in which I have been involved to a greater or lesser extent, are briefly discussed below.

- **LYNX**. The objective of the LYNX project was to monitor and identify protected species in environments of special environmental interest through the application of technologies that do not interfere with the natural activity of the fauna. The aim was to facilitate the monitoring of these species by means of non-invasive techniques that interfere as little as possible with the animal and its environment, improving the current solutions based on physical ringing or manual counting by means of photo-trapping techniques. Due to the high degree of threat it currently presents and its high biological value, the fundamental object of this project was the Iberian Lynx (Lynx Pardinus), a species whose main population is concentrated in the surroundings of Sierra Morena and in the area of the Doñana Natural Park, located in the Autonomous Community of Andalusia. In this scenario, the collaboration of our group with the company Magtel Operaciones was framed in two of the most ambitious tasks of the project: the design and development of an intelligent photo-trapping camera, capable of filtering, from the set of captured images, those that, with high probability, contained a lynx; and the implementation of an algorithm that would allow the univocal identification of the detected animals. In parallel, the design of the camera should solve some of the problems currently encountered in its deployment, allowing remote data collection.

  These two objectives fit perfectly with the lines of work that our group has been developing in recent years: embedded vision and image processing. Therefore, the work was approached with two teams of experts, working in continuous contact so that the needs of each of the groups would be met by the other. Thus, in the first line of work, the camera developed using MPSoC technology was equipped with triggering from a thermal sensor and a custom hardware, synthesised in the logic part, which allows LBP (Local Binary Pattern) texture features and an SVM (Support Vector Machine) classifier to discriminate the texture of the lynx from the rest of the scene. The initial version of the characterisation and classification algorithms, and the classifier training itself, were provided by the image processing group. In the second line of work, individual lynx identification used the lynx spotting pattern as a source of information, and uses visual features, together with structured spatial distribution information, to specifically characterise each animal. Finally, the cameras developed were able to connect to mobile phone networks to access a central server, from which data on the status of the cameras, as well as the captured images themselves, are collected. The entire system was deployed in real environments, with cameras located in Cordoba and Malaga.

- **HIRIS**. Our research group had previously worked with the company SHS Consultores on the design of an iris identification recognition system. This system worked at a distance and with people in movement and, although it already used FPGA technology for the design of the eye detector, both the resolution of the sensor and the capture speed were not adequate to ensure its correct operation. The HIRIS project aimed to include the initial phases of the recognition process (specifically the detection of the eye) in MPSoC technology, using a 16 Mpx sensor to capture the image and seeking processing speeds that would allow maximum performance of the sensor. Moreover, in our case, it was a first contact with the application of anti-fraud algorithms to the detected eye images. The work on this project

gave rise to the first two publications that form part of this compendium (Ruiz-Beltrán et al., 2022, 2023b).

- **DIMAS**. The objective of the DIMAS project is the design and development of a system for iris recognition in motion, whose main characteristics are to be embedded in MPSoC technology and to be improved with those features related to defence against presentation attacks and reliable behaviour. Led by SHS Consultores SL and currently ongoing, the project is an opportunity to acquire new knowledge related to the current challenges in iris recognition, but it is also a new opportunity to transfer scientific and technological knowledge from our research group to the industry. The expected result will be an improved version of the AIRIM product, currently offered by SHS Consultants. This improved version will include the 16 Mpx sensor that is being used in the latest versions of the system, adaptive optics, eye detection using CNN and defence against attacks by the use of textured contact lenses. The prototype will be evaluated in a large-scale pilot at the Consortium partners' facilities in Seville and Malaga. The third of the publications included in this Thesis is the initial fruit of the work carried out in this project (Ruiz-Beltrán et al., 2023a).

Apart from these projects, in which I have been contracted and which are closely related to this Doctoral Thesis, the work carried out in this Thesis has been partially ported to designs used in other projects, in which I have been part of the work team or I have also been contracted. Specifically, these projects are the MIRoN Integrated Technical Project funded under the EU RobMoSys project (H20202-732410), the RTI2018-099522-B-C41 project, funded by the Ministry of Science, Innovation and Universities and FEDER funds, and the PDC2022-133597-C42, TED2021-131739B-C21 and PID2022-137344OB-C32 projects, funded by MCIN/AEI/10.13039/501100011033 and by the European Union through the NextGenerationEU/PRTR programmes (for the first two projects mentioned above), and 'ERDF A way of making Europe' (for the third).

Methodology

The methodology followed for the work carried out in this PhD Thesis consists of three main work threads, which have been executed in parallel during most of the development time of this doctoral Thesis covering the algorithms, hardware platform and implementation:

- A first thread addresses the study of various artificial vision techniques aimed at detecting the eyes, understanding their operation and identifying advantages and disadvantages. The chosen ones were the aforementioned proposal by Viola and Jones (2001) and the Tiny YOLOv3 (Adarsh et al., 2020). It is important to highlight that this work thread has been carried out throughout the entire period of this doctoral Thesis. Initially, options that could offer high speed and deal with high-resolution images did not usually rely on neural networks but employed more classical approaches. This will change radically with the proposal of lighter solutions from the popular convolutional structure YOLO (*You Only Look Once*) (Redmon et al., 2016). As analyzed in some depth by Ruiz-Beltrán et al. (2023b), the first implementations that could be considered valid are based on Tiny-YOLO v3 and date back to the year 2018. It will take a few years for them to offer results that may be interesting for our application framework (Oh et al., 2020; Zhang et al., 2021; Esen et al., 2021). Using these latest proposals as a foundation, our second version of the eye detection system will be implemented.

- The second thread runs in parallel with the first one and involves setting up a hardware platform capable of capturing eye sub-images with enough quality to achieve iris recognition at a distance with people in motion and the hardware must have size an power constraints that allow it to be mounted on a gate and with enough frame rate to avoid motion blur. Similar to the description of the previous thread, there have been numerous changes in technology over a relatively short period. Finally, the decision was to synthesize the critical part of the system using FPGA but implementing the interfaces with the rest of the framework in a computer-like device. Therefore, an AP (*All Programmable*) SoC (*System-On-Chip*) was chosen from the beginning. These platforms include, alongside the programmable part (FPGA), a software part (microprocessor) that allows the implementation of global control over the system's operation, including its connection to the outside (partially with the sensor, as the core that finally reads the data is synthesized in the FPGA, as well as with the module responsible for processing eye images to obtain the iris pattern and proceed with recognition). After experimenting with different architectures, our proposal will be based on the Zynq Ultrascale+ platform from AMD/Xilinx, an MPSoC that features a microprocessor, GPU, and FPGA in the same packaging, It provides enough resources to deploy the Viola and Jones solution and it is also compatible with the Vitis AI framework to deploy DPU accelerator for CNN. Of the three main threads, this one was finished first, and with minor changes, we have been working with the same hardware for the last two years. In this case a Ultrascale+ device was selected within a System on Module (SoM) manufactured and sold by Trenz Electronic, the TE0820 4EV SoM. The module needs to be mounted in a compatible carrier board, which provides power and interfaces with the necessary peripherals. A carrier from the same manufacturer was selected: the TE0701. In a pursue to reduce costs and complexity, a custom carrier board was designed with only the peripherals used on the project. This board was manufactured and successfully tested.

- The third thread involves the design, implementation, and evaluation of various proposals generated throughout the development of this doctoral thesis in the first work thread. The implementation is aimed to meet the requirement of high speed (more than 40 frames per second (FPS)). One successful implementation resulted by customizing the algorithm proposed by Viola and Jones (2001). The description of this proposal is presented in Ruiz-Beltrán et al. (2022). This system meets the high-speed processing requirements and detects eyes in the images but suffers from two problems: it has a high false-negative detection rate and does not automatically reject detections with low contrast. Both issues imply that the flow of detections generated when a person crosses in front of our system is excessively high (if a person takes from 2 to 3 seconds to cross, and more than 40 fps are captured, the number of detected eyes could be around 250). This spike of data saturates the subsequent modules responsible for obtaining the iris pattern and proceeding with recognition. The solution was to include in the architecture a contrast estimator of the detected sub-image, this is because many of these detections are associated with eyes captured very far or near the focus point of the image sensor. Sub-images of eyes that do not offer the required contrast level could be eliminated by our system (Ruiz-Beltrán et al., 2023b). This proposal integrates correctly with the rest of the architecture, reducing this output sub-image flow. However, its inclusion results in a reduction in processing speed (which remains, in any case, above 40 fps), but does not reduce the high rate of false positives. The last task addressed in this thread has been the design and implementation of an eye detector based on the Tiny YOLO v3 neural network (Ruiz-Beltrán et al., 2023a). This task has entailed significant changes in the deployment of the software on the platform

(for example, moving from using *bare metal*[4] versions in the software part to depending on the installation of PetaLinux (XILINX, 2023b)).

Thesis outline

Besides the introductory chapter, **Thesis description**, the remaining ones in the first part of this thesis, are organized as follows:

- Theoretical background:

  The challenge of identifying regions of interest (ROI) in an image is a common problem in computer vision. Traditional methods to find it typically involve two main phases, feature Detection and classification. The Deep Learning methods, particularly, the so-called Convolutional Neural Networks involve training a model and deploying it. In reviewing state-of-the-art methods, we discuss the following approaches:

  - Color analysis
  - Edge detection
  - Feature extraction based
  - Deep Learning

- Summary of included papers

  This Doctoral Thesis is presented as a compilation of publications, three papers take part of it:

  - A parallelized version of the Viola-Jones algorithm synthesized to deploy it on an MPSoC
  - Integrate a contrast evaluation core into the hardware architecture
  - Deployment of a YOLOv3 Tiny CNN on the MPSoC

- Conclusions and future work

  The PhD thesis examines two eye detection methods: Viola-Jones and Tiny YOLO v3. While Viola-Jones is extremely fast, it has issues with false positives and lacks flexibility. Tiny YOLO v3, though slower, offers high accuracy and adaptability, making it more suitable for Iris At A Distance applications since it supports retraining and hardware reconfiguration without needing a complete system overhaul. Future work will integrate iris normalization and fraud detection algorithms within the MPSoC architecture to reduce server load and bandwidth usage, leveraging its full range of resources for enhanced performance, security, and scalability.

[4] *Bare metal* programming is a low level approach that works directly with the hardware registers, it works without abstraction layer and without operating (OS).

Theoretical background

The problem of finding the region of interest in the input image is a recurrent issue in computer vision. In our implementation this will be the goal: to detect the regions in the image where the eyes of the person(s) are located. This constitutes a search space optimization problem, where the Region Of Non-Interest (RONI) is ignored (Sharifara et al., 2017). As mentioned above, the methods already considered as traditional follow a scheme with at least two distinct phases. In the first phase, a characteristic is sought that allows the object to be labelled (for example, a specific colour or shape), while in the second phase a classifier is used to distinguish, using these characteristics, the object from other similar objects. In between these two phases, a compression process is sometimes included, in which the features are projected into a space of smaller dimensions. In any case, the whole process is highly conditioned by the experience of the designer himself, who controls at all times what is being extracted from the image and how it is processed. In this brief review of the state of the art in such methods, we will discuss methods based on colour study and face detection to search for eyes in faces; methods based on edge extraction; and methods based on searching for features in the image, extracted from the brightness levels of the pixels in the image, and statistically evaluated by the classifier to determine whether or not a region of the image is an eye.

Color analysis based

In the case of the eye detection, a first step could be to find a face and then use the natural expected location of the eyes. One method to find a face is based on the assumption that the subject is clothed. In that case the search of the face can be simplified to skin detection using the Image color. This method has two phases, the first involves training the system to detect the skin color by means of thresholds, fuzzy logic or other algorithms. The second phase is the actual face detection in which the image is partitioned according to the color regions (Pujol et al., 2017). The process is schematised in Figure 1.

Once the face area is detected, the next task is to extract features such as the eyes or the mouth. This can be made by using the natural location of these within the face or can be also be achieved by detecting different colors within the face. For example, the eyes have the white conjunctiva, the nose tends to be brighter and the mouth has lips with a darker color. In general, landmarks can be differentiated as zones without skin color within the face (Ji et al., 2018).

The color analysis can be carried out on different color spaces to increase the reliability of the system. In this case, after this preliminary analysis, the results can be merged as shown in Figure 2 (Singh et al., 2003). Since the face is a region of the image that has a homogeneous skin color, this method has as advantages that the color is not affected by the face orientation. The algorithm is robust to partial occlusions and this detection allows for a fast detection (Liu and Peng, 2010).

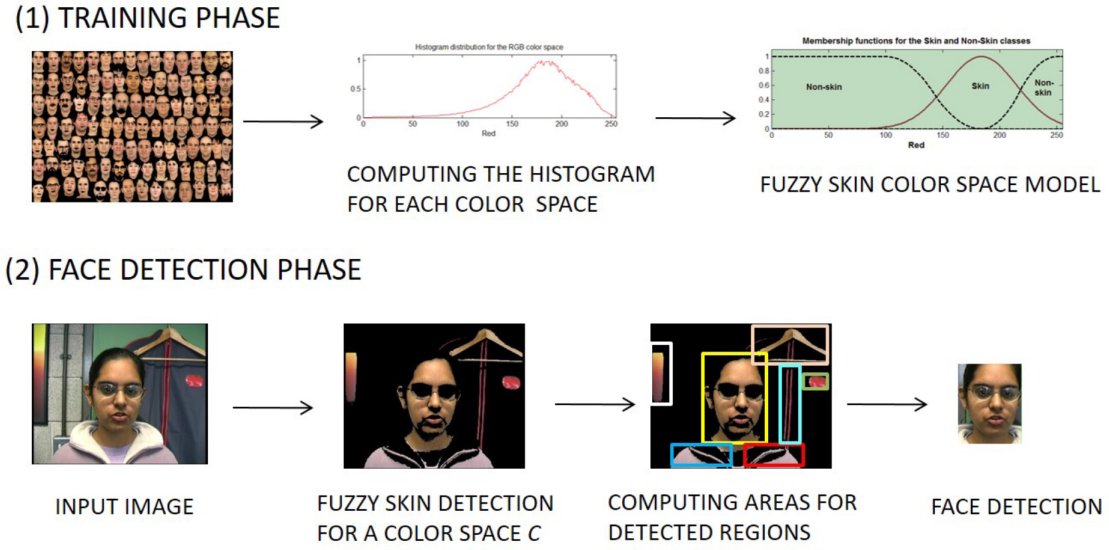


Figure 1: Face detection using color, system phases (Pujol et al., 2017)

Unfortunately, some drawbacks are identified:

- The lightning conditions can cause shadows on the face, therefore splitting the face region into various color. This can be mitigated using a specific space colors. For instance, in the YCrBr space color, a shadow would affect mostly the Luminance (Y) component leaving Chrominance (Cr and Br) components available for the analysis (Ning and Xutao, 2019).
- The varying skin color. The skin color can vary vastly depending on the subject ethnicity, age or even be affected by makeup. To mitigate this, special care must be taken on the first step of the method (Figure 1): the training must be done with an adequate dataset covering the skin color diversity. Also the presence of other skin areas can affect the algorithm. For example, hands or a belly can cause false positives.
- The need to process color, which translates into a higher computational cost. This cost can even be exacerbated if various color spaces are used simultaneously. Moreover, the need of color also affects the lightning and camera requirements since the illumination must be enough to differentiate colors.

In our particular IAAD setting, this approach was discarded because, to process the iris, an infrared lightning is employed. Therefore the camera should be able to pick up this infrared image and also does not need to detect and process color (alternative options using color were not valid at relatively high camera-eye distances (Borza et al., 2016)).

Edge detection based

Other methods are based on edge detection for face extraction. To detect edges in an image, the changes in grayscale levels of each pixel are compared with their neighbours using various operators such as Canny, Laplace, Sobel, or Robert (Chen and Cheng, 2015). Experimental results showed that the canny operator appears to be the most effective for human faces (Dong

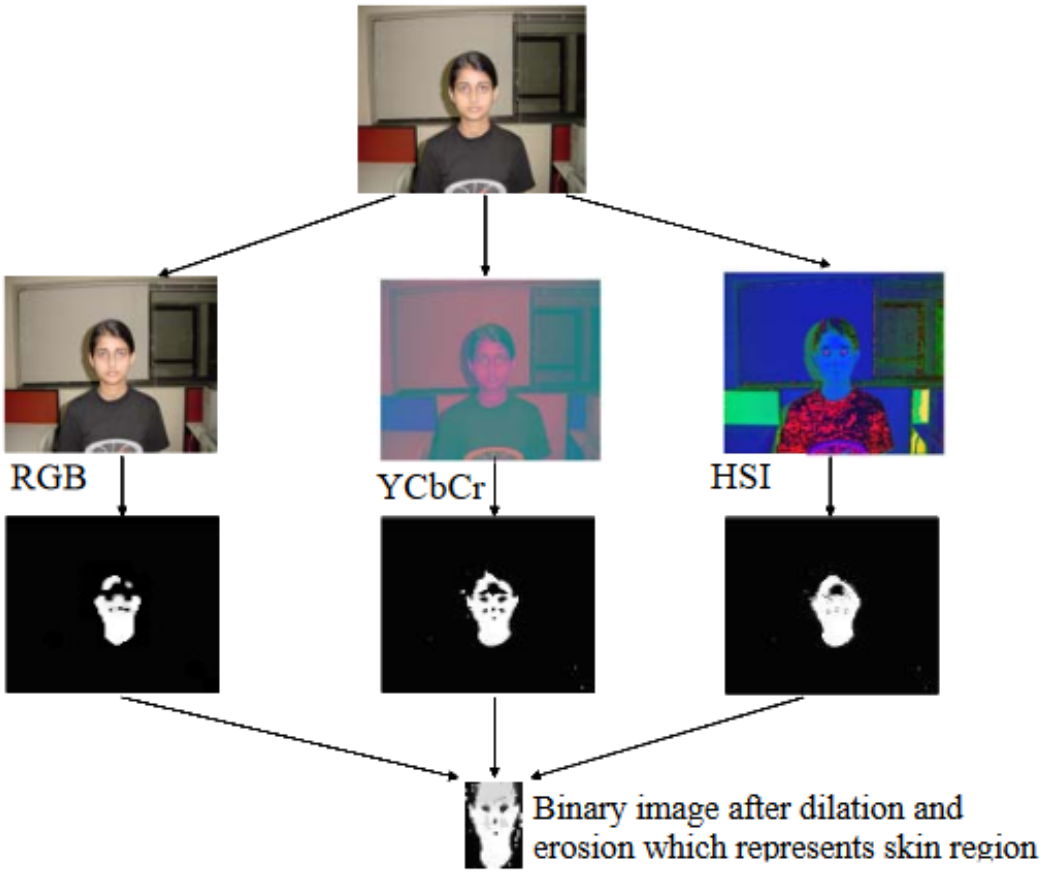


Figure 2: Face detection using different color spaces (Singh et al., 2003)

et al., 2021). In Figure 3, an example of the application of various of these operators in different faces can be seen.

Once the input image is processed and the edges are extracted, the next task is to locate the face and its landmarks within the image. Older approaches involved searching for a large oval-shaped figure that encloses smaller figures, that would be a face detection, and to find the eyes and mouth, it is assumed that the eyes will be located in the upper half, and the mouth in the lower half (Asteriadis et al., 2007). Furthermore, eyes are characterized by two concentric circles: the iris and the pupil. The pupil is typically fully visible, forming the inner circle, while the iris may be partially obscured by the eyelids so the irises can be found by searching concentric circles (Carneiro et al., 2009). Newer methods involve using classifiers such as Local Binary Patterns (LBP) (Dharma et al., 2022) or using neuronal networks (Putra et al., 2018).

These methods using edge detection have drawbacks:

- Edge detection is highly sensitive to noise, shadows, hair or even camera noise artifacts can create edges that can affect the face detection. This can be mitigated by either adjusting the edge detection sensibility and also by using Gaussian preprocessing to soften high frequency noise.

- The need to adjust the system if the camera location or the illumination is changed, if

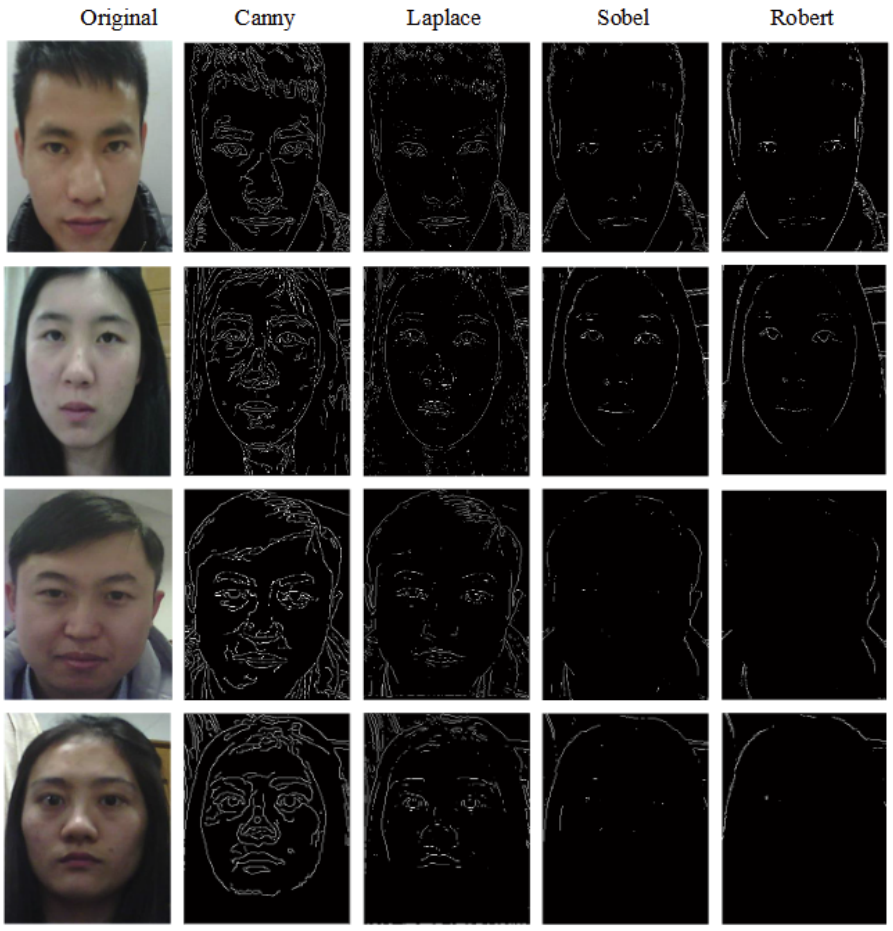


Figure 3: Face detection using edge operators (Chitra and Ponmuthuramalingam, 2015)

there is too much light, detecting edges becomes challenging, and if there is insufficient or non-uniform lighting, shadows can appear, distorting the detected edges. To mitigate this the implementation can use easy to change configurations by means of software defines or hardware jumpers. To accomplish the real time and power constraints hardware such as FPGA can be used, however varying a hardware design each time there is a change in the camera location or illumination is not viable.

- After the edge detection a classifier is still needed to find the ROI, in this approach the edge detection can be treated as a preprocessing and then some kind of classifier must be implemented.
- It has various steps that must be made sequentially, in this case the Gaussian preprocessing, then the edge operator (for example Canny) and then the detector.

This approach was discarded because the camera must be easy to reconfigure if the location or lightning is changed.

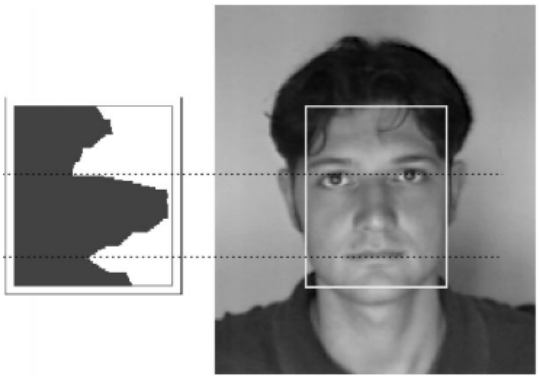


Figure 4: Face detection using histogram (Maio and Maltoni, 2000)

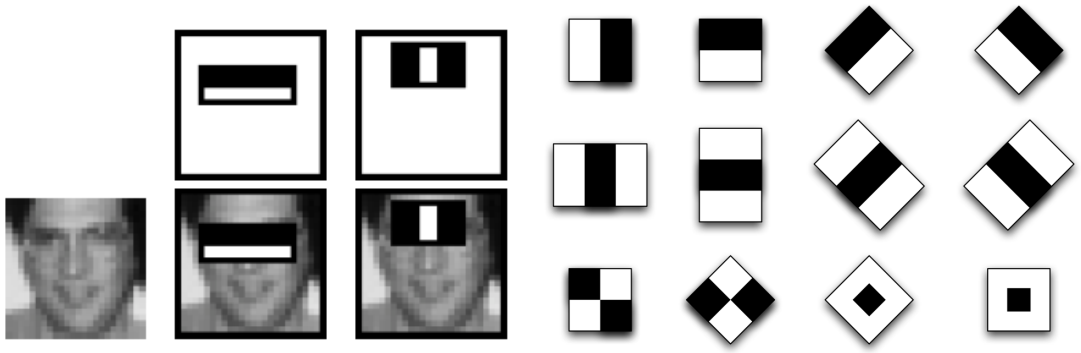
Feature extraction based

Another way to extract the face from an image is by analyzing the gray values of the pixels and by searching for certain features in the image. An early example of this is to use a face template as a model, where the peaks and valleys of the gray level of the image must fit, darker areas correspond to the eyes, eyebrows, or mouth. Similarly brighter areas correspond to the nose or cheeks. For this task, a histogram or a face model can be used, which can be seen in Figure 4 (Maio and Maltoni, 2000).

Similarly, a more refined method is to search for different features such as the Haar features. The detection of faces through these is based on the fact that the eyes are darker than the nose and the cheekbones. This difference between areas of the face can be extracted by integrating different areas of the image and then comparing the difference (yao Lu and Yang, 2019). An example of the features over a face image can be seen in Figure 5a. To simplify the algorithm, instead of using the original Haar features (Gaussian), the features are calculated using rectangles, whose internal sum of gray scale values is estimated using the concept of integral image. Based on this scheme, the popular Viola-Jones algorithm was created, which was the first algorithm for real-time face detection (Shamia and Chandy, 2017). This method uses Haar features like a convolutional kernel. Each feature is a unique value obtained by subtracting the sum of pixels in the white rectangle from the sum of pixels in the black rectangle, those features are calculated over a search window of 24 $\times$ 24 pixels, resulting in more than 160,000 possible features. Some of them are shown in Figure 5b. Out of all these features, many are irrelevant, therefore not calculated. To discriminate between relevant and irrelevant features, the AdaBoost classifier is introduced. It consists of a cascade of weak classifiers in which each stage classifier evaluates some Haar features. The cascade stages are executed sequentially, and if at any stage the evaluation is not passed, the sub-image is discarded, and the remaining stages are not executed, thus saving processing time.

This approach has drawbacks such as:

- It was designed to be executed on software, therefore it has various steps that must be made sequentially: first the preprocessing to get the integral and standard variation matrix, then the characteristics extraction, and then the classifier.
- The need to train the classifier to accurately detect eyes within an image.
- There are various calculation that require high resource utilization within the classifiers.



(a) Haar within a face image (b) Haar features

Figure 5: Haar characteristics (Viola and Jones, 2001)

To make this approach viable to work in our chosen hardware, changes must be made to enable it work directly with the video stream and to make the weak classifier calculations in parallel.

Deep Learning-based approaches

If image processing techniques combined with a specific set of features and a classifier were the working framework for decades, in recent years, the use of Deep Learning techniques displaced them in image processing, particularly, by the so-called Convolutional Neural Networks. Artificial neural networks were proposed years ago, inspired by the brain neurons, where the interconnections between them determine whether they activate or not based on input stimuli. In the artificial analogy, both neurons and their interconnections are mathematically modeled by using weights and activation functions. Through training, it is possible to find the set of weights for the interconnections that allow the network to generate an appropriate response for a specific input. Training was typically slow and complex, which has always hindered the practical application of this solution. With the development of new network schemes and technological advancements that allow the use of millions of patterns in relatively short training times, this problem has been addressed.

Specifically, for computer vision, CNNs have been proposed. In 1998, the first model for recognizing letters, whether handwritten or printed, was introduced, LeNet-5 (Lecun et al., 1998). Initially, the image is processed to reduce the resolution to a manageable size; then features are extracted through convolutions, and finally, the neural network is fed with sub-images for classification. A diagram of this process is shown in Figure 6. Although designed for character recognition, with appropriate training, this model is useful for detecting other objects and is very flexible in terms of what can be classified.

Therefore the training of the neural network can be done for face eye detection. For this purpose, it is possible to use publicly available databases such as the ORL face database, currently known as the Database of Faces (AT&T Laboratories Cambridge, 2001), the BioID Face Database (BioID GmbH, 2002), or the CASIA 3D Face Database (Center for Biometrics and Security Research, 2010), among others.

In this case, although high performance must be maintained, it is also mandatory that the video stream will be analyzed as fast as possible. As it is detailed in Ruiz-Beltrán et al. (2023b),

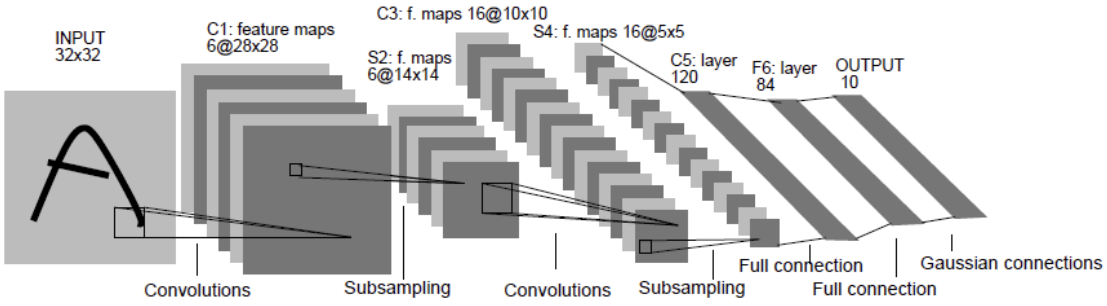


Figure 6: Lenet model architecture (Lecun et al., 1998)

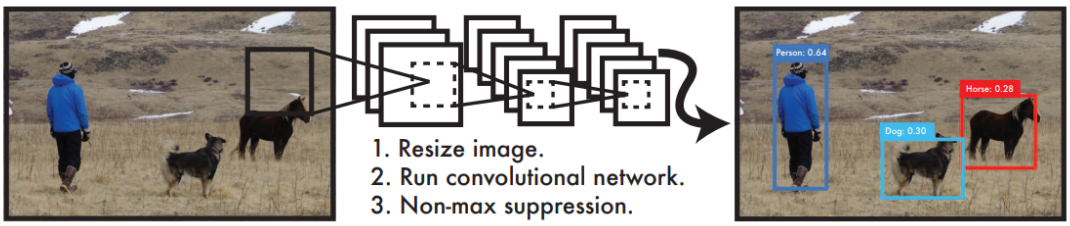


Figure 7: YOLO framework (Redmon et al., 2016)

various efforts have been made to improve the performance of an embedded CNN. We can identified two mainstream object detection strategies. In the two-stage strategies, a heuristic or regional suggestion generation method is firstly employed to obtain a set of candidate boxes. Then, these boxes are screened, classified and regressed. The typical example is the region based CNN (Ren et al., 2017). In the one-stage strategies the object detection problem is transformed into a global regression problem, giving results in an end-to-end manner. Global regression is not only capable of simultaneous assignment of place and category to the set of candidate boxes, but also of enabling models to get a clearer separation between object and background. The most popular one-stage proposal is YOLO (*You Only Look Once*) (Redmon et al., 2016). Approaches using the two-stage strategy perform a little better than those ones using the one-stage strategy in datasets such as MS COCO20217 (Liu et al., 2023). However, they are far from meeting real-time requirements on edge computing devices . One-stage proposals can keep a balance between real-time computing and performance. This is why the YOLO series is updated so quickly, being applied in real-time object detection such as video analysis, autonomous vehicles, and surveillance. An example of the framework is shown in Figure 7.

One of the most popular versions of the YOLO framework is YOLOv3 (Redmon and Farhadi, 2018). The compact version is known as Tiny YOLOv3. Intended for use on constrained systems, this version offers good performance while maintaining real-time processing speed. It is, therefore, a very suitable version for use in embedded systems, such as the MPSOC hardware used throughout the development of this Doctoral Thesis.

Summary of included papers

This Doctoral Thesis is presented as a compilation of publications, as such, the body of the thesis is organized into two distinct parts. In the first part (*Thesis Description*), a brief introduction to the Doctoral Thesis is provided in Spanish and then in English. The second part (*Included papers*) includes the three articles that form the work carried out in the Doctoral Thesis. An outline of the articles included in this second part of the Doctoral Thesis is provided, as well as the author's contributions to each of them.

Paper A: A parallelized version of the Viola-Jones algorithm synthesized to deploy it on an MPSoC

Description: This paper presents an embedded hardware-based solution for real-time eye detection. From an algorithmic standpoint, the proposal involves a redesign of the popular Viola-Jones algorithm, achieving a fully parallelized implementation of image processing (characterization and classification) executed in a single pass. Synthesized and implemented on an MPSoC, this approach processes more than 88 images per second, with the classifier taking less than 2 ms per image. Experimental validation was successfully carried out on a real iris recognition identification system, where people pass without stopping and walk at a normal pace. In this case, the prototype uses a CMOS digital image sensor that provides 16 MP images and outputs a stream of detected eyes as 640 x 480 images. The experiments to determine the accuracy of the proposed system employ the CASIA-Iris-distance V4 database, achieving a detection accuracy rate of 100%.

Author's contribution: There was an initial version of the algorithm that worked with a 5 MP sensor at a speed of around 25 frames per second. My work involved the redesign of the architecture and synthesize it to work on an AMD/Xilinx UltraScale+ MPSoC to achieve the values described above.

Paper B: Integrate a contrast evaluation core into the hardware architecture

Description: This work describes the integration of a functional block for evaluating the level of image blur into the logical part of the architecture already designed in the previous work. This way, the system can discard images that do not meet the required focus quality, this new core was implemented using Vitis High Level Synthesis (HLS), achieving a processing rate of over 57 images per second. For validation, an expanded version of the CASIA-Iris-distance V4 database was used. Experimental evaluation shows that the proposed framework can successfully discard out-of-focus eye images. Most notably, in a real implementation, this proposal allows discarding up to 97% of out-of-focus eye images, which will not need to be processed by the iris segmentation and normalized pattern extraction blocks.

Author's contribution: The technical work proposed in this paper has been entirely carried out by me, building upon the initial implementation of contrast estimation using convolution masks proposed in the literature by various authors.

Paper C: Deployment of a YOLOv3 Tiny CNN on the MPSoC

Description: In this paper, the deployment of an eye detection system based on the YOLOV3 Tiny CNN on an MPSoC Zynq XCZU4EV UltraScale+ is described. This eye detector is not only capable of processing high-resolution images captured at high speed but also discards those that are seriously affected by blur. To achieve this, the network is trained only with properly focused eye images. Additionally, taking advantage of the ability of neural networks to work with multi-channel inputs, the inputs to the CNN will be the grayscale image and a high-pass filtered version, typically used to determine if the iris is in focus. The complete system synthesizes other cores and implements the CNN using the so-called Deep Learning Processing Unit (DPU), the IP core proposed by AMD/Xilinx. Compared to previous designs that synthesize CNNs on FPGA, the DPU optimizes the typical functions used by deep learning algorithms, allowing it to accelerate neural network inference. As in previous cases, experimental validation has been successfully carried out in a real-world scenario, working correctly with moving people, and demonstrating that it is possible to detect only the eye images that are in focus. In the tests conducted, the prototype correctly discards up to 95% of the eyes in the input images for not being properly focused.

Author's contribution: In this proposal, the author has been responsible for all stages of the design, implementation, and evaluation process. Various CNN implementations were analyzed, ultimately deciding on YOLO v3 Tiny. The complete system architecture was designed, integrating the Xilinx/AMD DPU core.

Conclusions and future work

Conclusions

The two approaches implemented in this PhD Thesis, Viola Jones and Tiny YOLO v3 Tiny, allow the eye detection to operate at a higher speed than that provided by the image sensor used in the real implementation (the Teledyne e2v EMERALD 16MP), allowing the processing of all the frames provided by the sensor (47 frames per second). As a result, the real system, working with a very narrow depth of field, can accurately identify people walking in the walking zone.

The first implemented solution using the parallelisation of Viola and Jones' algorithm allows the classifier to operate at speeds that are hard to beat (over 700 frames per second), but the system in its evaluation in the real environment, while having a high rate of true positives, is prone to generate excessive false positives, especially with out of focus images. In an attempt to mitigate that a contrast evaluation core was implemented, although it reduces the throughput, it still fast enough to be above the 47fps provided by the sensor. Perhaps the major drawback of this implementation is that, once the system is implemented, any change in the detection stage would involve the redesign and re implementation of the system which is slow and not scalable.

The second implemented solution using the YOLOV3 Tiny-based is very robust, with a detection rate close to 100% in its evaluation in the real environment and almost no false positive detection. The major advantage here the flexibility, once the the hardware is implemented, it is possible to change the training, it is also possible to migrate to another CNN model or to change platform to add bigger DPU cores. The DPU cores are configurable, so if the model allows it, it can be configured to work within a smaller cores to free resources. If the fpga is big enough it can be also configured to work with up to 4 cores to accelerate the CNN faster. The drawback of this method is that, compared with the other implementation of this thesis, the classifier is slower.

The comparison between the Viola-Jones and Tiny YOLO v3 approaches reveals a trade-off between speed and flexibility/accuracy. The Viola-Jones algorithm, with its exceptional speed, is highly effective in scenarios where rapid processing is crucial. However, its propensity for false positives and the inflexibility of the system design limit its practical application, particularly in environments with varying image qualities.

On the other hand, Tiny YOLO v3, while slower, offers superior accuracy and flexibility. Its accurate detection rate and the ability to adapt the system through retraining or hardware reconfiguration make it a more robust solution for dynamic environments. The flexibility to upgrade or modify the CNN model without a complete system overhaul adds significant value, especially for applications requiring continual improvement and adaptation to new conditions. In the current system the clock DPU was implemented with 150MHz and 300MHz, it is a trade off, with higher clock a higher throughput can be achieved but active cooling is needed, so the chosen speed allows for a passive heatsink to handle the cooling effectively.

In conclusion, the choice between these two approaches depends largely on the specific require-

ments of the application. For scenarios demanding ultra-high-speed processing with a tolerance for false positives, the Viola-Jones algorithm is suitable. Conversely, for applications prioritizing accuracy and adaptability, the Tiny YOLO v3 approach is more appropriate despite its comparatively lower speed.

In conclusion, Tiny YOLO v3 is the preferred choice for eye detection in this implementation. Its robustness, adaptability, and high accuracy make it a more suitable solution for real-world applications, despite its comparatively lower speed, it is enough to process all the 47 FPS provided by the sensor.

Future work

Future work will focus on two main objectives: the extraction of the normalized iris pattern and the implementation of fraud detection algorithms within the same architecture in the MPSoC (Multiprocessor System on a Chip).

Normalization of the iris pattern: Currently, the normalization process of the iris pattern involves finding the radius of the iris edges and transforming it into a rectangular texture pattern suitable for biometric identification. This task is presently handled by a server after the reception of images from multiple embedded systems, creating a bottleneck due to the high volume of data being processed simultaneously. To address this, future work will integrate the iris normalization process directly within the embedded system. This integration will have benefits such as reduced bandwidth usage and server resource optimization. By processing the images locally, the amount of data transmitted to the server will be significantly reduced, as only the essential biometric patterns, rather than entire images, will be sent and Offloading the normalization process to the embedded system will free up server resources, mitigating the current bottleneck and enhancing overall system efficiency.

Implementation of fraud detection algorithms: Another key focus will be the implementation of fraud detection algorithms within the embedded system. These algorithms are designed to detect presentation attacks, such as the use of textured contact lenses or prosthetic eyes, thereby enhancing the security and reliability of the biometric system. Integrating fraud detection at the embedded system level offers similar benefits to those seen with iris normalization, reduced Server load and enhanced security

Leveraging MPSoC Resources: In the latest implementation, a DPU (Deep Processing Unit) is used to accelerate CNNs (Convolutional Neural Networks), and Petalinux is deployed on the software part of the MPSoC. This setup facilitates the utilization of the full range of resources provided by the Ultrascale+ ZU4EV MPSoC:

- Four ARM Cortex-A53 Cores: These high-performance cores can handle complex computations and multi-threaded processes efficiently.
- Two ARM Cortex-R5F Real-Time Processing Units (RPU): These units are ideal for time-critical tasks and can ensure that real-time processing requirements are met.
- Video Codec Unit (VCU): The VCU can be utilized for efficient video stream encoding and decoding, enabling the transmission of video streams from the camera over an Ethernet connection.
- ARM MALI 400MP GPU: This GPU can be leveraged for graphics processing and additional computational tasks, aiding in the acceleration of both iris normalization and fraud detection algorithms.

In summary, the future work on this project aims to integrate more processing tasks within the embedded system, thereby reducing dependency on server resources and enhancing the overall performance, security, and scalability of the identification system. By fully leveraging the capabilities of the MPSoC, the system will be able to meet the demands more effectively.

References

Adarsh, P., Rathi, P., and Kumar, M. (2020). Yolo v3-tiny: Object detection and recognition using one stage improved model. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 687–694.

Asteriadis, S., Nikolaidis, N., Pitas, I., and Pardas, M. (2007). Detection of facial characteristics based on edge information. In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - Volume 2: VISAPP,*, pages 247–252. INSTICC, SciTePress.

AT&T Laboratories Cambridge (2001). Database of faces. http://cam-orl.co.uk/facedatabase.html.

BioID GmbH (2002). The bioid face database. https://www.bioid.com/facedb/.

Borza, D., Darabant, A. S., and Danescu, R. (2016). Real-time detection and measurement of eye features from color images. *Sensors*, 16(7).

Carneiro, M., Veiga, A., Castro, F., Flôres, E., and Carrijo, G. (2009). Processing the segmentation stage of an iris recognition system through evolutionary algorithm. *Journal of Communication and Information Systems*, 24.

Center for Biometrics and Security Research (2010). Casia database. hhttps://hycasia.github.io/dataset/casia-irisv4/.

Chen, X. and Cheng, W. (2015). Facial expression recognition based on edge detection. *International Journal of Computer Science & Engineering Survey*, 6:1–9.

Chitra, A. and Ponmuthuramalingam, D. P. (2015). An approach for canny edge detection algorithm on face recognition. *International Journal of Science and Research (IJSR)*.

Daugman, J. (2004). How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30.

Dharma, A. S., Tambunan, N., and Sinaga, L. E. (2022). Face recognition with edge detection and lbp feature extraction. In *2022 IEEE International Conference of Computer Science and Information Technology (ICOSNIKOM)*, pages 1–7.

Dong, J., He, J., and Wang, H. (2021). Edge detection of human face. In *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, pages 596–601.

Esen, F., Degirmenci, A., and Karal, O. (2021). Implementation of the object detection algorithm (yolov3) on fpga. In *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 1–6.

Ji, Y., Wang, S., Lu, Y., Wei, J., and Zhao, Y. (2018). Eye and mouth state detection algorithm based on contour feature extraction. *Journal of Electronic Imaging*, 27:1.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Liu, Q. and Peng, G. (2010). A robust skin color based face detection algorithm. In *2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010)*, volume 2, pages 525–528.

Liu, S., Zha, J., Sun, J., Li, Z., and Wang, G. (2023). Edgeyolo: An edge-real-time object detector.

Maio, D. and Maltoni, D. (2000). Real-time face location on gray-scale static images. *Pattern Recognition*, 33:1525–1539.

Ning, Z. and Xutao, G. (2019). Face detection based on skin color extraction scheme. *IOP Conference Series: Materials Science and Engineering*, 569(3):032006.

Oh, S., You, J.-H., and Kim, Y.-K. (2020). Implementation of compressed yolov3-tiny on fpga-soc. In *2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*, pages 1–4.

Pujol, F. A., Pujol, M., Jimeno-Morenilla, A., and Pujol, M. J. (2017). Face detection based on skin color segmentation using fuzzy entropy. *Entropy*, 19(1).

Putra, T. W. A., Minardi, J., Gaffar, A. F. O., Suprapty, B., Malani, R., and Supriadi (2018). Comparison of canny and centroid on face recognition process using gray level cooccurrence matrix and probabilistic neural network. In *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, pages 351–356.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *CoRR*, abs/1804.02767.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.

Ruiz-Beltrán, C. A., Romero-Garcés, A., González, M., Pedraza, A. S., Rodríguez-Fernández, J. A., and Bandera, A. (2022). Real-time embedded eye detection system. *Expert Systems with Applications*, 194:116505.

Ruiz-Beltrán, C. A., Romero-Garcés, A., González-García, M., Marfil, R., and Bandera, A. (2023a). Fpga-based cnn for eye detection in an iris recognition at a distance system. *Electronics*, 12(22).

Ruiz-Beltrán, C. A., Romero-Garcés, A., González-García, M., Marfil, R., and Bandera, A. (2023b). Real-time embedded eye image defocus estimation for iris biometrics. *Sensors*, 23(17).

Shamia, D. and Chandy, D. A. (2017). Analyzing the performance of viola jones face detector on the ldhf database. In *2017 International Conference on Signal Processing and Communication (ICSPC)*, pages 312–315.

Sharifara, A., Rahim, M. S. M., Navabifar, F., Ebert, D., Ghaderi, A., and Papakostas, M. (2017). Enhanced facial recognition framework based on skin tone and false alarm rejection. *CoRR*, abs/1702.04377.

Singh, S. K., Chauhan, D., Vatsa, M., and Singh, R. (2003). A robust skin color based face detection algorithm. *Journal of Applied Science and Engineering*, 6:227–234.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I.

XILINX (2023a). Dpu for convolutional neural network.

XILINX (2023b). Petalinux tools documentation: Reference guide (ug1144).

XILINX (2023c). Vitis ai optimizer user guide.

yao Lu, W. and Yang, M. (2019). Face detection based on viola-jones algorithm applying composite features. *2019 International Conference on Robots & Intelligent System (ICRIS)*, pages 82–85.

Zhang, H., Jiang, J., Fu, Y., and Chang, Y. (2021). Yolov3-tiny object detection soc based on fpga platform. In *2021 6th International Conference on Integrated Circuits and Microsystems (ICICM)*, pages 291–294.

Zhu, J., Wang, L., Liu, H., Tian, S., Deng, Q., and Li, J. (2020). An efficient task assignment framework to accelerate dpu-based convolutional neural network inference on fpgas. *IEEE Access*, 8:83224–83237.

Part II

Included papers

Real-time embedded eye detection system

Camilo A. Ruiz-Beltrán, Adrián Romero-Garcés, Martín González, Antonio Sánchez Pedraza, Juan A. Rodríguez-Fernández, Antonio Bandera,
Real-time embedded eye detection system,
Expert Systems with Applications,
Volume 194,
2022,
116505,
ISSN 0957-4174,
https://doi.org/10.1016/j.eswa.2022.116505.
(https://www.sciencedirect.com/science/article/pii/S0957417422000070)
Abstract: The detection of a person's eyes is a basic task in applications as important as iris recognition in biometric identification or fatigue detection in driving assistance systems. Current commercial and research systems use software frameworks that require a dedicated computer, whose power consumption, size and price are significantly large. This paper presents a hardware-based embedded solution for eye detection in real-time. From an algorithmic point-of-view, the popular Viola–Jones approach has been redesigned to enable highly parallel, single-pass image-processing implementation. Synthesized and implemented in an All-Programmable System-on-Chip (AP SoC), this proposal allows us to process more than 88 frames per second (fps), taking the classifier less than 2 ms per image. Experimental validation has been successfully addressed in an iris recognition system that works with walking subjects. In this case, the prototype module includes a CMOS digital imaging sensor providing 16 Mpixels images, and it outputs a stream of detected eyes as 640 × 480 images. Experiments for determining the accuracy of the proposed system in terms of eye detection are performed in the CASIA-Iris-distance V4 database. Significantly, they show that the accuracy in terms of eye detection is 100%.

Real-time embedded eye image defocus estimation for iris biometric

Ruiz-Beltrán, Camilo A. and Romero-Garcés, Adrián and González-García, Martín and Marfil, Rebeca and Bandera, Antonio,
Real-Time Embedded Eye Image Defocus Estimation for Iris Biometrics,
Sensors,
Volume 23,
2023,
17,7491
ISSN 1424-8220,
https://doi.org/10.3390/s23177491
(https://www.mdpi.com/1424-8220/23/17/7491)

Abstract: One of the main challenges faced by iris recognition systems is to be able to work with people in motion and where the sensor is at an increasing distance (more than 1 metre) from the person. The ultimate goal is to make the system less and less intrusive and require less cooperation from the person. When this scenario is implemented using a single static sensor, it will be necessary for the sensor to have a wide field of view and for the system to process a large number of frames per second (fps). In such a scenario, many of the captured eye images will not have adequate quality (contrast or resolution). This paper describes the implementation in an MPSoC (Multiprocessor System-on-Chip) of an eye-image detection system that integrates, in the programmable logic (PL) part, a functional block to evaluate the level of defocus blur of the captured images. In this way, the system will be able to discard images that do not have the required focus quality in the subsequent processing steps. The proposals have been successfully designed using Vitis High Level Synthesis (VHLS) and integrated into an eye detection framework capable of processing over 57 fps working with a 16 Mpixel sensor. Using for validation an extended version of the CASIA-Iris-distance V4 database, the experimental evaluation shows that the proposed framework is able to successfully discard unfocused eye images. But what is more relevant is that, in a real implementation, this proposal allows discarding up to 97% of out-of-focus eye images, which will not have to be processed by the segmentation and normalised iris pattern extraction blocks.

FPGA-Based CNN for Eye Detection in an Iris Recognition at a Distance Systemc

Ruiz-Beltrán, Camilo A. and Romero-Garcés, Adrián and González-García, Martín and Marfil, Rebeca and Bandera, Antonio,
FPGA-Based CNN for Eye Detection in an Iris Recognition at a Distance System,
Electronics,
Volume 12,
2023,
22,4713,
ISSN 2079-9292,
https://doi.org/10.3390/electronics12224713
(https://www.mdpi.com/2079-9292/12/22/4713),
Abstract: Neural networks are the state-of-the-art solution to image-processing tasks. Some of these neural networks are relatively simple, but the popular convolutional neural networks (CNNs) can consist of hundreds of layers. Unfortunately, the excellent recognition accuracy of CNNs comes at the cost of very high computational complexity, and one of the current challenges is managing the power, delay and physical size limitations of hardware solutions dedicated to accelerating their inference process. In this paper, we describe the embedding of an eye detection system on a Zynq XCZU4EV UltraScale+ multiprocessor system-on-chip (MPSoC). This eye detector is used in the application framework of a remote iris recognition system, which requires high resolution images captured at high speed as input. Given the high rate of eye regions detected per second, it is also important that the detector only provides as output images eyes that are in focus, discarding all those seriously affected by defocus blur. In this proposal, the network will be trained only with correctly focused eye images to assess whether it can differentiate this pattern from that associated with the out-of-focus eye image. Exploiting the neural network's advantage of being able to work with multi-channel input, the inputs to the CNN will be the grey level image and a high-pass filtered version, typically used to determine whether the iris is in focus or not. The complete system synthetises other cores and implements CNN using the so-called Deep Learning Processor Unit (DPU), the intellectual property (IP) block released by AMD/Xilinx. Compared to previous hardware designs for implementing FPGA-based CNNs, the DPU IP supports extensive deep learning core functions, and developers can leverage DPUs to conveniently accelerate CNN inference. Experimental validation has been successfully addressed in a real-world scenario working with walking subjects, demonstrating that it is possible to detect only eye images that are in focus. This prototype module includes a CMOS digital image sensor that provides 16 Mpixel images, and outputs a stream of detected eyes as 640 $\times$ 480 images. The module correctly discards up to 95% of the eyes present in the input images as not being correctly focused.