TESIS DOCTORAL

Visión artificial aplicada en escenarios reales. Una aproximación práctica

Autor:

Mª Araceli Sánchez Sánchez

Director/es:

Enrique Cabello Pardos
Cristina Conde Vilda

Programa de Doctorado en Tecnologías de la Información y las Comunicaciones. Escuela Internacional de Doctorado

2021

A mis hijas: María y Alejandra

AGRADECIMIENTOS

Hace mucho tiempo, como los cuentos, en la Universidad de Salamanca empezó a gestarse esta tesis y, como en los cuentos también, el transcurrir de la vida hizo que surgieran muchas aventuras que hicieron que ese cuento no acabase. Hoy parece que es cuento se ha hecho realidad y ha llegado el día que podamos decir lo de "colorín colorado…."

Llegados al final sólo me queda agradecer a todas aquellas personas que me han acompañado a lo largo de esta larga aventura. En primer lugar a mis directores Enrique y Cristina, a los dos mis más profundas gracias, por su paciencia, por su dedicación y por todo. En especial a Quique, lleva en mi vida tanto como mi marido, después de tanto tiempo no puedo considerarlo otra cosa que no sea mi amigo, un buen amigo al que le debo no sólo esta tesis sino mi profesión, sin su ayuda nunca me habría dedicado a la docencia.

Quiero agradecer también a todas esas personas que siempre me han animado a continuar y acabar este proyecto, seguro que me dejo gente, pero no me quiero olvidar de mis compañeros de Béjar, Paco, Alejandro, Miguel Angel y muy especialmente a Carmen, mi conciencia, sin conciencia no se puede vivir. No puedo olvidar tampoco a Quino, que yo creo que ya perdió la esperanza, a Ana de Luis y Ana Belén Gil que siempre me han animado a continuar.

A mi familia, mi madre, mis hermanos Juan Carlos y Miguel Angel, mi hermana Ana, mi cuñado Víctor, mis cuñadas Petri, Erika y Pili y mis sobrinas Emma y Maia, que siempre me han apoyado y sé que siempre lo harán.

Y cómo no, a César, María y Alejandra, mi vida, sin ellos, nada tiene sentido ni es motivo para hacer nada. Cuando se empieza la tesis uno piensa en agradecimiento a sus padres porque han hecho posible llegar hasta ese punto, en mi caso, con lo dilatado en el tiempo que ha sido este parto, el pensamiento no es quién te ha llevado hasta aquí sino a quién dejas lo que has conseguido y no puedo estar más orgullosa de dejarle esta humilde contribución a mis dos maravillosas hijas,

María y Alejandra. Y de César no puedo más que darle las gracias, no por su apoyo incondicional, no por su ayuda, no por su paciencia, sino por una vida. Dos piezas que encajan a la perfección como en el mejor Ravensburguer para componer una imagen tan bonita como un Falcon.

A todos mil gracias.

Y colorín colorado este cuento se ha acabado

RESUMEN

La presente tesis recoge la aplicación de un conjunto de técnicas de visión artificial a una serie de situaciones reales muy dispares entre sí, demostrando que para todas ellas existe una metodología adecuada para la resolución del problema en cuestión.

Los campos en los que se ha trabajado han sido los siguientes:

Identificación de personas por medio de su cara en situaciones controladas empleando redes neuronales y métodos para la reducir el tamaño de las imágenes sin pérdida de información útil para el reconocimiento.

En la determinación del tamaño de rocas en una situación de trabajo real en una mina a cielo abierto se han empleado métodos similares, con las dificultades que conlleva el trabajar en unas condiciones no favorables como puede ser la deficiencia en la iluminación o el exceso de polvo.

También se ha trabajado en otras dos situaciones reales como son un aeropuerto, en el que se realiza la detección de ataques a la identificación automática de los individuos en un desplazamiento natural por el paso fronterizo del mismo.

Por último se ha realizado un estudio de dos pasos de peatones, uno con semáforo y otro sin semáforo, donde el objeto de estudio era detectar situaciones de conflicto entre coches y peatones con el objetivo de mejorar la seguridad vial.

ABSTRACT

The present thesis collect the usage of artificial vision techniques to several real-life problems that are very different, showing that in any case there are methodologies able to solve the problem.

The fields in which the work has been done are the following:

Person identification by neural network face recognition combined with methods able to reduce the size of the images without loss of relevant information.

Similar methodologies have been applied to the determination of rock size in a real environment of an open pit mining, with the inconvenient of working in an environment of low illumination.

The other situations are an airport in which attacks to people identification systems are detected in a non-static situation as the natural walking through a border crossing.

Finally a study in two crosswalks, one of them was controlled by a traffic light while the other one was non-controlled. The aim of this study is to detect the conflicts between pedestrians and automobiles in order to improve the security of the crossings.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE ECUACIONES

ÍNDICE DE TABLAS

CAPITULO 1

INTRODUCCIÓN

1.1. Introducción y motivación

Los sistemas de visión artificial suponen un gran avance en el desarrollo de estudios de multitud de situaciones y contextos que ayudan al hombre a conseguir mejores resultados y más precisos, en algunos casos, y facilitar la realización de tareas en muchos otros.

La aplicación de la inteligencia artificial a sistemas de visión artificial nos ofrece en la actualidad logros que se suponían ciencia ficción hasta hace unos pocos años y que en la actualidad son una realidad como puede ser el caso de la conducción automática.

Existe una amplísima gama de técnicas aplicables a visión artificial que han hecho posible su utilización en campos muy dispares como pueden ser medicina, identificación de individuos, desplazamientos autónomos, etc. Casi todos ellos implican extraer información de una imagen, ya sea tomada en tiempo real, ya sea almacenada, identificando los elementos que aparecen en ella y obtener una conclusión o realizar una acción en función de lo que muestra la imagen.

Trabajar con imágenes implica que una gran cantidad de información debe ser procesada, lo que dificulta la tarea por los requisitos de memoria y tiempo de computación que necesitan. En particular, cuando se realiza en tiempo real las dificultades se incrementan, añadiendo una serie de imponderables que pueden surgir y que el sistema desarrollado debe solucionar para ser útil.

Los sistemas de visión artificial deben ser diseñados de acuerdo a aquellas variables que sean relevantes para obtener la respuesta adecuada a la necesidad que se trata de solucionar. En especial el control en la toma de imágenes, ya sea porque se pueden establecer unas condiciones ideales o no, o porque exista una voluntad de alterar las mismas implicarán decisiones críticas en cuanto al sistema que de debe emplear. Igualmente la necesidad de la inmediatez de los resultados también condiciona de manera fundamental los sistemas, ya que no es lo mismo realizar un estudio estadístico de determinados eventos que se tratan que detectar, que detectar un peligro inminente sobre el que el tiempo de actuación resulta fundamental.

1.2. Objetivo de la tesis

El objetivo de este trabajo de tesis es estudiar las distintas técnicas empleadas en visión artificial, lo que supone conocer y emplear herramientas para el procesado de las imágenes, reducir su tamaño sin pérdida de información, mejorar la calidad de la escena, obtener solo aquella información útil para el problema en cuestión despreciando lo que podríamos considerar “escenario”. También implica estudiar las técnicas necesarias para identificar y reconocer objetos dentro de la escena. Si la imagen es en vivo, como sucede en algunos de los ejemplos que se van a presentar, se hace necesario el empleo de métodos de seguimiento de objetos en movimiento.

Todo ello con el objeto de desarrollar sistemas automáticos que sea capaz de, empleando los métodos más adecuados para cada situación, trabajar de forma autónoma facilitando la realización del trabajo posterior del usuario.

1.3. Planteamiento del problema y solución propuesta

Cuando empecé a trabajar en esta tesis lo hice motivada por el atractivo que suponía trabajar en visión artificial y con redes neuronales con la de posibilidades que eso implicaba. Empecé estudiando las distintas técnicas y métodos. Al estar integrada en la Universidad como parte del profesorado, surgen proyectos a lo largo del tiempo que hacen que esos conocimientos que uno tiene se puedan extrapolar a situaciones reales donde implantarlos, pasando de ser mera teoría a aplicaciones reales con todas sus implicaciones.

Podría decir que esa fue la motivación de este trabajo de tesis, la participación en proyectos donde unos conocimientos teóricos se han visto plasmados en la resolución de problemas de diversa índole. En concreto presento cuatro contextos muy diferentes.

Identificación de personas

Este trabajo podría considerarse, al menos yo así lo hago, el más teórico de todos, lo que no implica que no se pudiera extrapolar a situaciones reales de identificación de individuos, de hecho sirvió de base para otro de los trabajos que se van a presentar.

En este caso se disponía de imágenes almacenadas en una base de datos pública que contenían caras de individuos, todas ellas tomadas en condiciones muy controladas tanto de iluminación como de posición del individuo delante de la cámara.

Se trató de implementar técnicas para la extracción de información de la imagen y modificación de la misma sin pérdida de información útil con el fin de mejorarlas y reducir su tamaño para ver si el posterior empleo de una red neuronal permitía la identificación de los sujetos.

Identificación de rocas en minería

Es un problema que se planteaba en una mina a cielo abierto donde se realizan voladuras de las que se recogen rocas para extraer el uranio que contienen, esas rocas deben pasar por unas cintas transportadoras para ser conducidas a la zona de extracción. El problema puede surgir cuando el camión que recoge los restos de la voladura vuelca las rocas en unas tolvas que dirigen a las rocas a las cintas. Cuando el tamaño de la roca es excesivo, la tolva se obstruye impidiendo que la descarga de la misma se realice de forma normal y al saturarse supone un punto de peligro para el operario que allí trabaja.

La solución que se plantea es instalar un sistema de visión compuesto por una cámara y un ordenador de forma que se van obteniendo imágenes en las que se identifican las rocas, en el momento que alguna de ellas excede el tamaño de salida de la tolva, se detiene el proceso y se avisa al operario para que pueda ser dividida en condiciones de seguridad antes de que se produzca el atasco de la tolva por una roca sepultada entre otros restos de la voladura.

El sistema propuesto empleaba redes neuronales para determinar si se producía situaciones de posible atasco en la tolva.

Detección de suplantación de personalidad en una frontera

La experiencia en el trabajo previo nos permitió llevar a cabo un nuevo escenario en una frontera en un aeropuerto real, el de Madrid, para identificar y detectar situaciones de accesos no permitidos de personas. En este caso de la imagen se extraen marcadores biométricos para la identificación del usuario. Detectando situaciones de sabotaje como podrían ser máscaras que dificultasen su identificación.

El empleo de redes neurales, convolucionales en este caso, permitieron obtener buenos resultados teniendo en cuenta la dificultad de trabajar en tiempo real en un

sistema con un mínimo control en la adquisición de las imágenes a diferencia del primer caso donde el usurario colaboraba en la captura más beneficiosa de la imagen.

Conflictos peatón-vehículo

En este caso el objetivo es mejorar la seguridad vial en las ciudades cuando los peatones cruzan por un paso de cebra, pese a ser lugares seguros para ellos, se dan circunstancias donde se producen accidentes en estos lugares.

Nuevamente con un sistema de cámaras se obtienen imágenes en movimiento, en este caso video donde se ve discurrir tanto a peatones como a vehículos, el sistema automático debe ser capaz de prever que se va a producir un conflicto salvo que algo de los dos, o ambos, agentes implicados modifiquen su conducta.

En este caso son necesarias técnicas de identificación de objetos y de seguimiento de los mismos en una secuencia para que una red neuronal sea la encargada de determinar si se producirá o no el encuentro traumático entre ambos o si se ha generado una situación de peligro con la suficiente frecuencia como para que recomiende la alteración de las condiciones del paso.

1.4. Estructura de la memoria

La memoria de la presente tesis se ha dividido en capítulos, este primero corresponde a la introducción, los objetivos que se persiguen, cómo se plantea abordar la tesis y las soluciones que se proponen y el presente apartado de estructura de la memoria de tesis.

Los siguientes capítulos se corresponden con los distintos problemas que se han utilizado para aplicar las técnicas de visión artificial.

El primero de ellos aborda el problema de identificar personas mediante una imagen estática de la cara empleando una red neuronal, el segundo lo que persigue identificar en este caso son rocas en una mina trabajando en tiempo real.

El tercer ejemplo es similar al primero en cuanto a objetivo, identificar personas, pero en este caso se trata de un sistema en tiempo real en un aeropuerto.

El último capítulo de ejemplos se presenta un sistema cuyo objetivo es predecir conflictos entre peatones y vehículos en un sistema en tiempo real localizado en 2 pasos de cebra.

La memoria concluye con un último capítulo que incluye las conclusiones globales de la tesis después trabajar con escenarios tan diversos.

La memoria se complementa con un apartado relativo a la bibliografía consultada para la realización de este trabajo.

CAPITULO 2

IDENTIFICACIÓN DE CARAS MEDIANTE REDES NEURONALES

2.1. Introducción

El rostro humano está lleno de estímulos visuales, detalles y matices que hacen que seamos diferentes unos de otros. Incluso en el caso de los gemelos se puede encontrar algún matiz que los diferencia, esto está en oposición al "mundo de los bloques", utilizado a veces en Visión Artificial donde todo presenta una estructura más regular que facilita su identificación, donde lo único que nos interesa, por ejemplo es localizar las estructuras para evitar la colisión de un robot, sin preocuparnos de si se trata de un objeto u otro.

La identificación de personas por medio de su cara es un problema complejo de resolver debido a la gran cantidad de posibles matices que puede tener el rostro humano, sin embargo el cerebro humano realiza esta función de forma rápida y fiable. Este problema ha despertado un gran interés desde hace mucho tiempo, sobre todo en entornos policiales y de seguridad, donde tener un sistema que de forma automática reconociese a un determinado individuo entre muchos otros basándose para ello en una imagen de su cara, permitía ganar tiempo y evitar tener complicados sistemas de reconocimiento. De ahí que ya en 1.910 Galton [1] empleaba fórmulas numéricas para resolver el problema de la identificación de caras.

Una consideración a tener en cuenta es que cuanta más información esté disponible para realizar la identificación, más fácil resultará el reconocimiento. Por ejemplo, el contar con información acerca de la voz, la altura de la persona, la forma de andar, de vestir, etc. sería de gran utilidad. Al utilizar solamente una imagen con la cara del individuo estamos limitando la información de que disponemos, lo que hace que situaciones ambiguas que se producen y sobre las que no podemos tomar una decisión, se resolverían fácilmente con información de otro tipo. El caso de los forenses ilustra de manera clara este problema, ellos intentan emplear todo aquello que pueda ser útil para el reconocimiento, ya sean características físicas o genéticas, llegando a conseguir grados de aciertos muy elevados.

El problema no es trivial, debido a su complejidad hace que incluso los seres humanos se equivoquen. Resulta, por tanto, poco realista el suponer que un sistema automático va a conseguir un porcentaje del 100% de éxito en unas condiciones equiparables a las del mundo real, porcentaje que sí se podría lograr en un estado entorno controlado y con condiciones totalmente idóneas.

Existen dos posibles contextos en los que resulta interesante el reconocimiento de sujetos a través de su cara:

* Localizar una persona en una base de datos de caras. El resultado proporcionado por este tipo de sistemas son todos aquellos individuos que más se parecen a la persona buscada. En estos sistemas sólo se utilizan dos o tres imágenes por persona y no se requiere una respuesta en tiempo real. Son bases de datos que cuentan con un elevado número de imágenes teniendo todas ellas características similares en cuanto a iluminación, posición del individuo, etc., con el fin de facilitar el proceso de búsqueda. Un ejemplo de este tipo de situación son los retratos utilizados por la policía.

* Identificación de personas en tiempo real. En este caso se almacenan múltiples imágenes de cada persona que difieren entre ellas de alguna forma, como puede ser en la inclinación de la cabeza. El sistema

dará como resultado sólo una de ellas, la que más parecido tenga con el individuo buscado. Al disponer de varias imágenes por persona el espacio de almacenamiento necesario aumenta, pasando a ser considerable este aumento al ir añadiendo personas a la base de datos. Pese a este inconveniente es una ventaja dar una respuesta en tiempo real así como una mayor fiabilidad en el resultado. El hecho de poder reconocer individuos a partir de imágenes tomadas en condiciones variables, lo hacen muy útil para sistemas de seguridad, sobre todo en el control de accesos.

2.2. Estado del arte

Desde el principio de la Historia ha existido una gran preocupación por el problema de la identificación, muchos han sido los trabajos realizados en este el tema. Brevemente citaremos a continuación algunas de las aproximaciones más recientes empleadas para resolver el problema.

Existen dos excelentes revisiones bibliográficas [2, 3] que recogen los trabajos realizados. Hemos de tener en cuenta, no obstante, que en la mayoría de los trabajos se han asumido una serie de limitaciones:

Utilización de bases de datos de reducido tamaño.

Individuos en posiciones frontales o casi frontales.

Los individuos empleados, en muchos casos tienen características muy claras e identificativas.

Estas consideraciones lo que hacían es facilitar el reconocimiento y, en ningún caso, son equiparables con una situación del mundo real.

En bibliografía se han utilizado principalmente dos soluciones para el reconocimiento de caras:

- Características Geométricas.
- Plantillas Flexibles.
- Algoritmos de Autocaras

2.2.1.1. Características Geométricas

La solución geométrica consiste en medir las características de los rasgos del individuo (diámetro de los ojos, longitud de los labios, de la nariz...) para posteriormente utilizar estas medidas para identificar a una persona en concreto. La idea subyacente es el emparejamiento de las medidas del individuo a identificar y las almacenadas en una base de datos. Pese a que este sistema ofrece buenos resultados, tiene el inconveniente de que no existe un algoritmo preciso de localización automática de puntos en el rostro humano.

- ***Kanade*** [4] elaboró un método automático para la extracción de características basado en medir distancias en el rostro del individuo, utilizó para ello una base de datos compuesta por 20 personas consiguiendo un porcentaje de reconocimiento que oscilaba entre el 45 y el 75%.

- ***Brunelli y Poggio*** [5] obtuvieron mejores resultados, llegando a conseguir hasta un 90% de aciertos, analizando características del individuo como pueden ser la longitud y profundidad de la nariz, posición de la boca, forma de la barbilla, etc. La base de datos empleada en este caso era de 47 personas.

- ***Cox et al.*** [6] lograron un 95% de identificaciones empleando una técnica de mezcla de distancias. El sistema que utilizaron consistía en

extraer manualmente de cada cara 30 distancias que la representaban, como puede ser la distancia entre los ojos o el tamaño de la nariz. La base de datos empleada estaba formada por 685 individuos de los que se extraían 95 características.

2.2.1.2. Plantillas Flexibles

En el caso de las plantillas flexibles se elabora una especie de plantilla del rostro de la persona que se emparejaría con las almacenadas en la base de datos hasta encontrar aquella con la que más concuerda. Este método es efectivo siempre que las imágenes tengan igual escala, orientación, iluminación, etc. que las imágenes empleadas para el entrenamiento [6]. Presentando problemas a la hora del emparejamiento si en la imagen del sujeto aparece, por ejemplo, con los ojos cerrados o si cambian las condiciones de iluminación.

- ***Brunelli y Poggio*** [5] demostraron que con la misma base de datos que había sido empleada previamente para el método de las características geométricas se podía conseguir un 100% de reconocimiento utilizando un sistema de emparejamiento de plantillas.

- ***Burt et al.*** [7] emplearon una aproximación basada en plantillas flexibles de multiresolución.

- ***Yuille et al.*** [8] desarrolló un método para detectar y describir rasgos del rostro utilizando plantillas flexibles. Los rasgos interesantes para el reconocimiento, por ejemplo los ojos, se representan mediante plantillas parametrizadas. Se utiliza una función para definir qué ejes de enlace, picos y valles en la imagen corresponden con los de la plantilla, la plantilla interactúa dinámicamente con la imagen modificando sus parámetros para minimizar los valores de la función con el objeto de conseguir un ajuste mejor.

2.2.1.3. Autocaras

Se trata de otra aproximación para solucionar el problema de la identificación de caras, se basa en que las imágenes de las caras se proyectan en un " espacio de cara " que es el que mejor recoge la cantidad de variación que se produce de unas imágenes a otras, conociendo las imágenes. El espacio de cara es definido mediante *autocaras* (eigenfaces) las cuales son autovectores del conjunto de casos. Estos autovectores no corresponden necesariamente a características aisladas como los ojos, nariz y orejas, por ejemplo. El sistema permite la posibilidad de aprender a reconocer nuevas caras empleando un método de aprendizaje no supervisado.

- ***Turk y Pentland*** [9, 10] emplearon este sistema para la identificación de caras humanas. Utilizaron un método en el que las caras eran proyectadas dentro de las componentes principales del conjunto de imágenes de entrenamiento original. Las autocaras resultantes eran clasificadas por comparación con individuos conocidos. La técnica de componentes principales lineales asume que las caras caen dentro de un espacio dimensional pequeño y que de la suma o media de dos caras se obtiene una cara. Claramente esto no es cierto cuando los componentes principales se aplican a toda la cara [11]. La base de datos que utilizaban estaba formada por imágenes de 16 sujetos con variaciones en la orientación, escala e iluminación. Los porcentajes alcanzados en el reconocimiento fueron del 96%, 85% y 64% para variaciones de iluminación, orientación, y escala. En este caso se empleaba una red de neuronas para el reconocimiento.

- ***Pentland et al.*** [12, 13] obtuvieron muy buenos resultados (95%) pero las características de las imágenes empleadas eran muy poco realistas, ya que las imágenes de un mismo individuo no presentaban apenas variaciones.

- ***Moghaddam y Pentland*** [14] consiguieron igualmente buenos resultados empleando la base de datos FERET. El sistema emplea un

excesivo preprocesamiento para la localización de la cabeza, detección de características y normalización para la geometría de la cara, traslación, iluminación, contraste, rotación y escala.

2.2.2 Niveles de Gris de las Imágenes

Recientemente ha habido una explosión de trabajos en este tema que han expandido otra posible solución puesta en marcha para el reconocimiento de caras, consistente en utilizar los niveles de gris de la imagen para luego emparejarlos con los de imágenes almacenadas en una base de datos. Esta alternativa presenta la ventaja, frente a los casos anteriores, que no se pierde ninguna información de la imagen ya que estamos utilizando todos los rasgos que proporciona la cara del sujeto mediante el empleo de los niveles de gris de la imagen. Contamos, así, con información de todo el rostro y no sólo de algunas partes como sucedía en los casos de las características geométricas o de las plantillas flexibles expuestas anteriormente, el inconveniente que tiene es que las necesidades de almacenamiento se incrementan de forma considerable ya que disponemos de un elevado volumen de información.

2.2.3. Preprocesamiento de las Imágenes

Otra cuestión a tener en cuenta a la hora del reconocimiento por medio de imágenes es si éstas se utilizan directamente o si es necesario incluir una etapa previa de preprocesamiento de las imágenes. Considerando los posibles problemas de almacenamiento que se pueden presentar, parece razonable el reducir el tamaño de las imágenes de forma que, sin perder información útil, podamos reducir el espacio de almacenamiento que requieren.

En cualquier caso el preprocesamiento al que se sometan las imágenes debe ser el mínimo posible para conseguir tiempos de reconocimientos menores.

- ***Turk y Pentlland*** [9], mediante sus trabajos sugieren la existencia de un reconocimiento basado en un procesamiento de imágenes de bajo nivel, es decir, utilizando información bidimensional. Esta teoría del procesamiento de bajo nivel no contradice la expuesta por Marr [15] sobre el procesamiento de alto nivel. Su argumento se basa en el rápido desarrollo y velocidad del reconocimiento de caras en los seres humanos y en los experimentos fisiológicos realizados con monos [16], en los cuales parece haberse aislado un conjunto de neuronas que responden selectivamente ante rostros humanos.

2.2.4. Redes Neuronales

Teniendo en cuenta los apartados anteriores, parece lógico pensar que una buena solución sería aproximarse al método empleado por el cerebro para resolver el problema, se trataría por lo tanto de construir un sistema que simule el comportamiento del cerebro humano. La tecnología actual nos presenta a las Redes de Neuronas como lo más parecido al cerebro.

Una red de neuronas es un programa que es un modelo simplificado del cerebro. Contiene nodos que equivalen a las neuronas y conexiones entre nodos que se asemejarían con las sinapsis.

Las redes de neuronas solucionan problemas difíciles de emparejamiento de patrones. Cada nodo suma las entradas que recibe desde otros nodos y, mediante una función de trasferencia, transforma esa suma en su salida. La salida de un nodo se propaga a través de las conexiones para convertirse en la entrada de otros nodos. A cada conexión se le asocia un peso, que es un valor que se emplea para ponderar esa conexión.

Las redes neuronales requieren un proceso de aprendizaje basado en la modificación, en función del error cometido en el último patrón, de los valores de los pesos de las conexiones durante la fase de entrenamiento. Para entrenar a la red basta

con proporcionarle una serie de entradas. Durante el entrenamiento, los pesos de las conexiones son reajustados sucesivamente hasta producir la salida deseada, lo cual supone la existencia de un aprendizaje real.

Las redes neuronales se presentan, por lo tanto, como una herramienta idónea ya que se adecúa perfectamente al problema del reconocimiento de caras. Una vez entrenada la red nos permite identificar individuos utilizando imágenes distintas a las empleadas durante el periodo de entrenamiento.

- ***Turk y Pentland*** [9, 10] emplearon un esquema de reconocimiento basado en redes de neuronas que utilizaban una proyección de las imágenes para el entrenamiento de la red. La base de datos de la que se servían para hacer las pruebas contaba con 16 sujetos con imágenes con diferente orientación, escala e iluminación así como ligeras variaciones en la expresión de la cara. Los resultados que consiguieron con estas condiciones fueron de un 96, 85 y 64% frente a variaciones de la luz, orientación y escala respectivamente.

- ***Lee et al***. [17] Implementaron un algoritmo de autocaras y obtuvieron un 10% de error con una base de datos de entrenamiento formada por 200 individuos y empleando otros 200 para probar los resultados. Para el entrenamiento transformaban las imágenes en vectores. El método de aprendizaje empleado es no supervisado basado en mapas autoorganizados (SOM). Utilizaban también el método de Karhunen-Loéve para reducir la redundancia existente en la base de datos.

- ***Fleming and Cottrell*** [18] emplearon unidades no lineales para entrenar una red de neuronas del tipo de las *Back Propagation* para clasificar imágenes de caras humanas. Se basaron en los trabajos realizados sobre el tema por Kohonen y Lahtio [19].

- ***Stonham*** [20] empleo un sistema WISARD con algo de éxito a imágenes binarias de caras.

De los resultados mostrados hay que destacar que en muchos de los casos en los que se obtienen buenos resultados se debe a que la base de datos empleada es muy pequeña y las imágenes no presentan grandes variaciones de luz ni rotaciones.

2.2.5. LVQ

Otro método apropiado para el reconocimiento es el **LVQ** (**L**earning **V**ector **Q**uantization) que fue intencionadamente desarrollado para el reconocimiento de patrones estadísticos. Su principal ventaja es la reducción de las operaciones computacionales comparando con otros métodos estadísticos tradicionales, al tiempo que se puede conseguir una exactitud de reconocimiento óptima.

Su base es la clasificación de los datos de entrada en regiones de clases empleando para ello uno o más vectores de código para cada región de cada clase. El LVQ utiliza un esquema supervisado de aprendizaje a base de patrones de vectores que le son proporcionados en la fase de entrenamiento.

La exactitud que se puede lograr en cada tarea de clasificación para la que se emplea el LVQ y el tiempo necesario para el aprendizaje depende de los siguientes factores:

- ✓ Número óptimo de vectores de código asignados a cada clase y sus valores iniciales.
- ✓ El algoritmo concreto, una tasa de aprendizaje adecuada y un criterio correcto para detener el aprendizaje.

El presente trabajo aborda el problema de la identificación de caras comparando diferentes métodos de procesamiento de imágenes de niveles de gris de rostros humanos utilizando posteriormente una red neuronal para simular el comportamiento del cerebro humano en el reconocimiento.

El sistema empleado nos proporciona como salida aquella que más se aproxima al patrón buscado.

2.3. Base de datos

La base de datos empleada para el desarrollo del presente trabajo está formada por imágenes que provienen de una base de datos pública de la Universidad de Berna (Suiza) [21] y corresponden a 30 personas todas ellas de sexo masculino y de aproximadamente la misma edad. Todos ellos son de raza blanca algunos con gafas y varios de ellos presentan rasgos asiáticos. Con respecto al pelo se observan diferentes formas, estilos de peinado y tamaños.

De cada individuo contamos con 10 imágenes en las que aparece sólo la cabeza con la siguiente distribución:

2 corresponden al sujeto mirando de frente a la cámara.

2 corresponden al sujeto mirando a la derecha de la cámara.

2 corresponden al sujeto mirando a la izquierda de la cámara.

2 corresponden al sujeto mirando hacia arriba.

2 corresponden al sujeto mirando hacia abajo.

En la siguiente figura aparecen las imágenes de que disponemos de uno de los individuos empleados en este trabajo tras una etapa de preprocesamiento consistente en la aplicación del operador de Gauss para reducir el tamaño de la imagen original.

Figura 2.1. *Imágenes pertenecientes a uno de los individuos de la base de datos (ampliación 300% sobre el tamaño empleado en el trabajo).*

En las imágenes se aprecia que el grado de giro de la cabeza es grande, así como la iluminación, que fue controlada en todo momento evita, en lo posible, cambios bruscos.

El tamaño original de las imágenes es de 512x342 pixeles con 256 niveles de gris, tamaño que fue reducido posteriormente mediante la etapa de preprocesamiento a 32x22. La figura 2.2 nos muestra una de las imágenes en tamaño original y la misma imagen tras la reducción.

Figura 2.2. *Imagen en tamaño natural y tras el procesamiento.*

Los sujetos fueron fotografiados sobre un fondo blanco y la cara ocupa casi toda la imagen. La figura 2.3 nos muestra una imagen de cada uno de los individuos de la base de datos, las imágenes han sido reducidas en un 80 % de su tamaño:

Figura 2.3. *Imagen frontal de cada individuo de la base de datos.*

2.4. Objetivos

Los objetivos más importantes que se persiguen en el presente trabajo son:

* **Diseñar un sistema de reconocimiento de caras humanas**. Se emplea para ello una red de neuronas con diferentes algoritmos de entrenamiento con el fin de poder hacer estudios comparativos con respecto al porcentaje de aciertos obtenidos en función del algoritmo empleado.

  Los algoritmos empleados son LVQ y una red neuronal de retropropagación [22, 23].

* **Estudiar cómo se comporta la red construida** (expresado como la modificación en el porcentaje de aciertos) **frente a variaciones en el ángulo de giro que presentan las caras**. Se trata de ver qué ocurre cuando cambia la posición de la cara (se observará que se producen variaciones significativas en función de la posición).

* **Utilizar toda la información disponible en las imágenes**. Para ello la técnica empleada consiste en trabajar con los niveles de gris de la imagen, que evitan que se pierda información. El inconveniente este método es que precisa gran cantidad de espacio de almacenamiento dado el gran volumen de información que se maneja.

* **Realizar un estudio de la calidad del Software desarrollado.** Se pretenden evaluar si las herramientas empleadas han proporcionado las prestaciones que de ellas se esperaban y en qué grado.

2.5. Planteamiento teórico

El presente trabajo se centra en intentar solucionar el problema de la identificación de caras, intentando aproximar la solución a la utilizada por el cerebro humano.

Se ha descrito en bibliografía la similitud del funcionamiento de la red neuronal con el cerebro lo que podría convertir a esta herramienta en una de las más indicadas para solucionar el problema de la identificación de caras humanas. De acuerdo con esta hipótesis se decidió construir una red de neuronas cuyo objetivo sería llevar a cabo el reconocimiento de dichas caras.

Los trabajos antes mencionados de Marr [15] y Poggio [5], así como los de Turk y Pentland [9, 10], demostraron que en el reconocimiento de caras debe existir una etapa de procesamiento rápido en el cerebro. Por lo que las imágenes deben haber pasado una fase previa que se haya encargado del preprocesamiento de las mismas antes de que estas entren en la red neuronal para llevar a cabo la etapa de entrenamiento.

Siguiendo la línea de C. Lee Giles [17], se disminuyen lo máximo posible las etapas previas de procesamiento con el fin de:

- evitar la pérdida de información útil, que podría traducirse en una disminución de la eficiencia en el reconocimiento de la red neuronal.
- evitar la aparición de información redundante, que aumentaría los problemas de almacenamiento de los datos de la imagen, incrementándose también los tiempos de entrenamiento de la red.
- disminuir los tiempos de procesamiento de las imágenes con el fin de hacer más rápido el proceso de la identificación.

La situación a la que esto da lugar la muestra la figura 2.4.

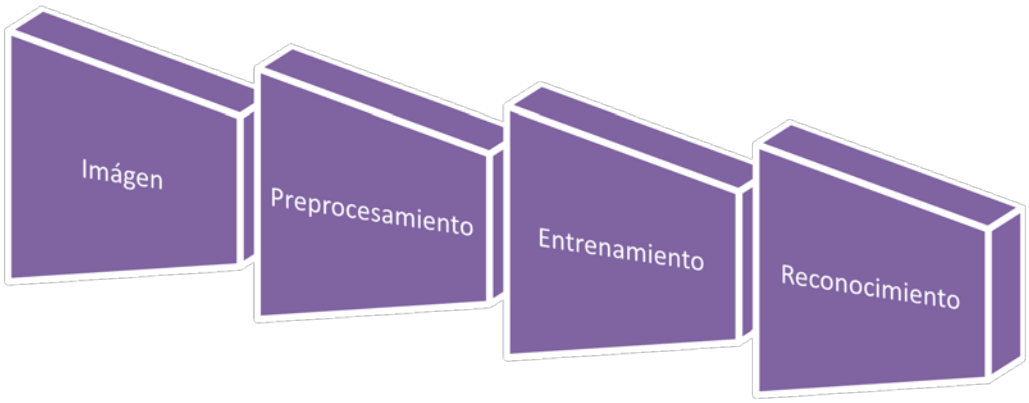

Figura 2.4. *Esquema general del sistema.*

2.6. Preprocesamiento

Trabajar con imágenes proporciona gran cantidad de información visual ya que contienen información de todos los aspectos del objeto que estemos considerando. Las imágenes empleadas en este trabajo tienen un tamaño de 512x342 píxeles con una resolución de 256 niveles de gris. Esto supone un total de 175.104 bytes por imagen, son por lo tanto 175.104 bytes de información que se puede emplear. Para una imagen de una cara supones el considerar todos los rasgos que caracterizan a ese individuo a través de su cara.

El objetivo que habíamos planteado era utilizar esta información en una red de neuronas para resolver el problema del reconocimiento facial. La realización de una red neuronal que procesase estas imágenes necesitaría tantas neuronas en su primera capa como pixeles tuviese la imagen; serían necesarias por lo tanto, 175.104 neuronas en la capa de entrada de la red, dicha cantidad es demasiado elevada ya que daría lugar a tiempos de entrenamiento muy grandes, así como un volumen de memoria considerable para almacenar tanto la información de la imagen como la de las neuronas y los pesos.

En vista de ello se debe localizar y analizar solamente la información esencial para el trabajo que nos ocupa y eliminar la gran cantidad de información irrelevante que contienen las imágenes, si queremos obtener unos tiempos de procesamiento razonables.

Teniendo en cuenta las consideraciones previas relativas a minimizar el preprocesamiento de las imágenes, parece lógico que la técnica empleada no contradiga los objetivos planteados para este trabajo, en concreto el referente a la utilización de toda la información que posee la imagen. De ahí que las técnicas que se han empleado estén encaminadas principalmente a reducir el elevado espacio que se necesita al utilizar los niveles de gris y al trabajar con imágenes con tanta resolución, sin que ello suponga perder información.

Los tipos de preprocesamiento a los que se someten las imágenes son:

- Operador de Gauss.
- Segmentación de las imágenes.
- Normalización.

El operador de Gauss es la única de las técnicas empleadas que conduce a la reducción del tamaño de las imágenes. Esto lo convierte en una etapa necesaria independientemente de cuál sea el proceso que se siga a continuación. De esta manera se originan tres tipos de preprocesamiento diferentes:

1º Emplear solamente el operador de Gauss.

2º Emplear el operador de Gauss y a continuación una fase de normalización.

3º Emplear el operador de Gauss y a continuación una fase de segmentación.

Los casos anteriores se analizan por separado a continuación.

Caso 1: Operador de Gauss solamente.

El proceso que se sigue en este caso lo refleja la figura 2.5.

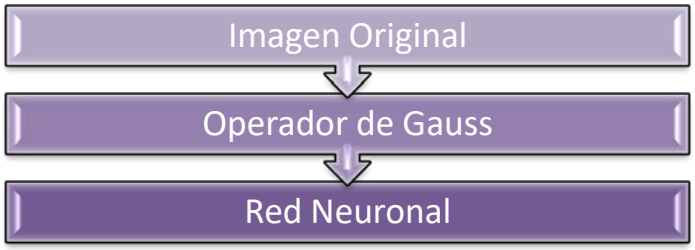

Figura 2.5. *Etapas de procesamiento mediante la utilización exclusiva de una pirámide Gaussiana.*

Las imágenes, como se ha indicado anteriormente, tienen un tamaño de 512x342 píxeles siendo este excesivo para ser procesado directamente por una red de neuronas, ya que la primera capa de dicha red debería disponer de 175.104 neuronas, número que claramente se aprecia que es excesivo. Es por ello que se optó por reducir el tamaño de las imágenes hasta 32x22 pixeles, con lo que se da lugar a una primera capa que precisa de 704 neuronas, muchas menos que las necesitábamos utilizando el tamaño original de las imágenes. Pese a que este es un número elevado aún, su simulación resulta factible tanto en volumen como en el tiempo que posteriormente necesitará para su entrenamiento.

Para realizar la reducción del tamaño de las imágenes se utiliza una pirámide jerárquica o pirámide de Gauss de cinco niveles.

Una pirámide jerárquica está formada por un conjunto de imágenes del mismo objeto pero con diferentes resoluciones. Al ser un pirámide de cinco niveles obtendremos cinco imágenes iguales pero con diferentes resoluciones siendo estas de

512x342, 256x171, 128x85, 64x42 y 32x22 píxeles. Como se aprecia lo que se consigue con cada nivel es reducir a la mitad la resolución de la imagen anterior.

Para la construcción de la pirámide se utiliza el operador de Gauss cuya definición es la que presenta la ecuación 2.1:

$$\mathbf{G} = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Ecuación 2.1. *Operador de Gauss.*

El operador de Gauss hace que cada pixel de la imagen englobe información de los que le rodean, esto origina la aparición de información redundante, por lo que tras la aplicación de dicho operador es necesaria la eliminación de filas y columnas de la imagen para eliminarlas. Mediante esta supresión de filas y columnas se consigue reducir la resolución inicial de la imagen a la mitad en ambos ejes. Se suprimen las filas y columnas situadas en posiciones impares de las imágenes. Esto nos permite tras cinco aplicaciones del operador conseguir disminuir la resolución inicial de 512x342 hasta 32x22 pixeles.

Otro efecto no deseado que se consigue tras la sucesiva aplicación del operador es suavizar los contrastes de las imágenes originales, de manera que se dificulta el proceso de identificación de individuos. A pesar de todo, la ventaja obtenida con la reducción a la cuarta parte del número de bytes necesarios para cada imagen compensa la pequeña perdida de información sufrida.

Un ejemplo que ilustra la aplicación de la pirámide gaussiana sobre una de las imágenes de la base de datos empleada lo muestra la siguiente figura:

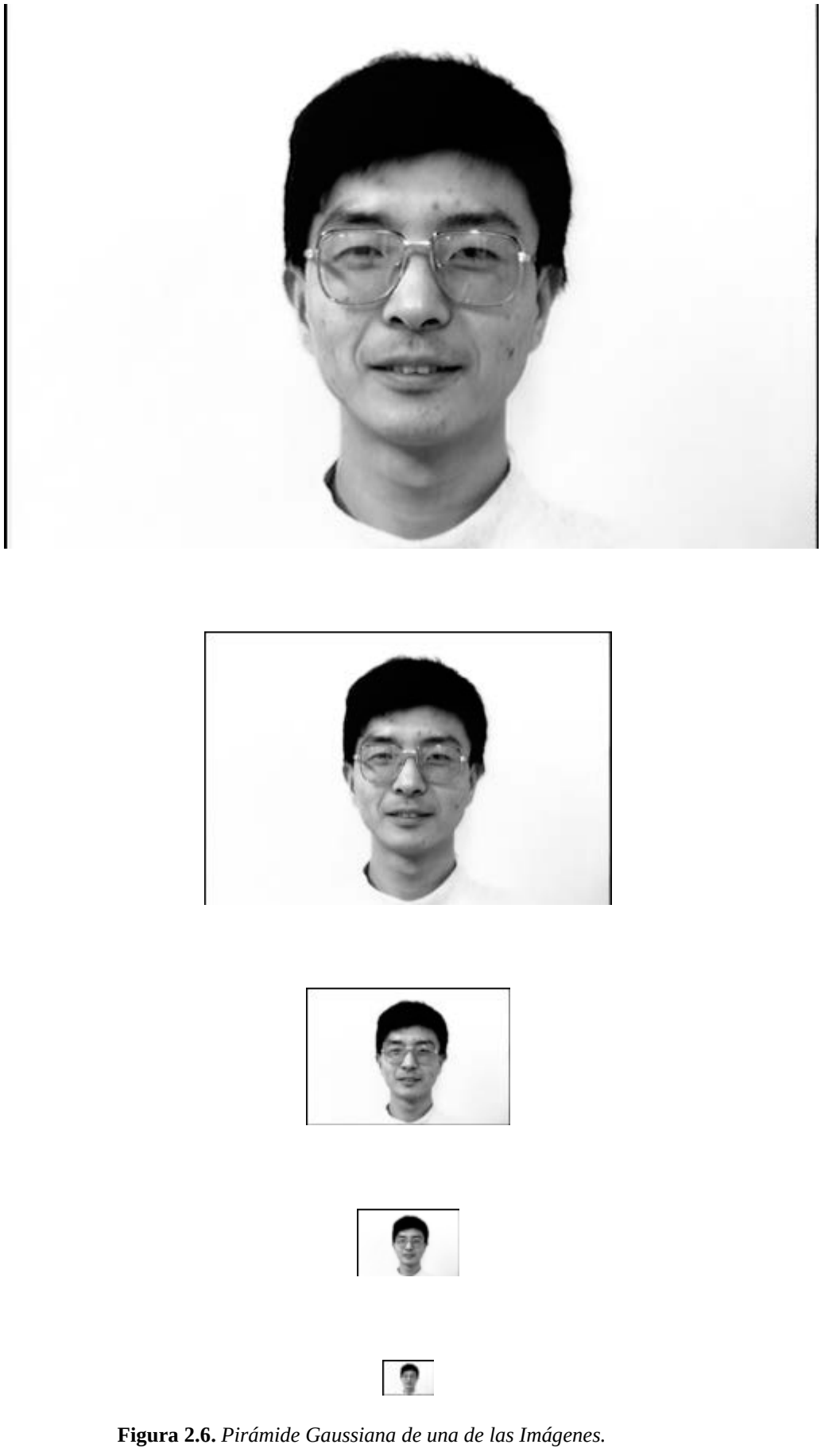

Figura 2.6. *Pirámide Gaussiana de una de las Imágenes.*

La aplicación del citado operador nos permite reducir el tamaño de las imágenes sin pérdidas importantes de información visual. Esto nos permite solucionar el problema del almacenamiento y a la vez no perder información que pudiera ser útil para el reconocimiento.

Como se aprecia en la figura anterior, la aplicación del operador de Gauss ha eliminado detalles secundarios de la imagen, manteniendo los rasgos más importantes de esta. El efecto global producido en la imagen es equivalente a la foveacion:

- **Fijar la atención en los datos más importantes.** Los objetos de poco interés visual van desapareciendo con la aplicación del operador de Gauss, lo que implica que los elementos que aparezcan en el nivel de mínima resolución son los de máximo interés visual.

- **Enfocar / desenfocar la imagen.** El conjunto de imágenes generadas por medio de una pirámide jerárquica es similar al mecanismo atencional empleado por la visión humana. Si observamos un objeto que se encuentra alejado, solo podremos identificar su silueta, sin embargo, a medida que nos acercamos al objeto vamos distinguiendo más detalles y características del mismo hasta que llegamos a estar tan cerca que podemos ver todos los detalles.

- **Actuar como integrador.** La aplicación del operador de Gauss, como ya se ha comentado anteriormente, hace que cada pixel de la imagen contenga información de los que le rodean, estamos por lo tanto integrando la información de la imagen en los distintos píxeles que la forman.

Una pirámide Gaussiana actúa de forma similar a como lo haría la retina humana, ya que integra la información que contienen los niveles de gris, de tal forma que reduce el gran volumen de datos de las imágenes con grandes resoluciones, manteniendo la información esencial para el reconocimiento al igual que operaria la retina del ojo humano.

Las imágenes, una vez disminuido su tamaño mediante el operador de Gauss, fueron utilizadas como entrada a una red de neuronas. Previa eliminación de 30 bytes de cabecera con que contaban las imágenes, siendo el tamaño definitivo de las mismas de 674 bytes.

Caso 2: Operador de Gauss y Normalización.

El proceso empleado en este caso lo refleja la figura 2.7.

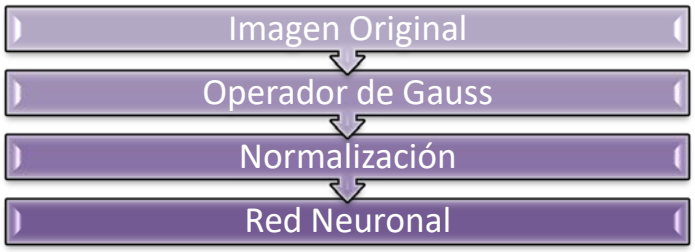

Figura 2.7. *Etapas de procesamiento mediante la utilización de una pirámide Gaussiana y una fase de Normalización.*

Este nuevo procesamiento se basa en la aplicación de una etapa de normalización a las imágenes obtenidas tras la aplicación del operador de Gauss a las imágenes originales.

Las imágenes se someten a una reducción de su tamaño mediante una pirámide gaussiana y posteriormente se normalizan los niveles de gris de las mismas con el fin de eliminar cambios bruscos en la iluminación de las imágenes, ya que estos cambios se

dan pese a que la iluminación fue controlada durante el proceso de adquisición de las imágenes.

Cada una de las imágenes se normaliza por separado en función del valor medio de los niveles de gris de dicha imagen y de la desviación estándar de los niveles con respecto a la media obtenida.

La fórmula empleada para la normalización es la indicada en la ecuación 3.2.

$$nivel\_gris_{\ normalizado} = \frac{nivel\_gris_{actual} - nivel\_gris\_medio}{desviación}$$

Ecuación 2.2. *Fórmula empleada en la normalización de las imágenes.*

donde

nivel_gris$_{actual}$ es el nivel de gris que tiene el punto considerado de la imagen.

nivel_gris$_{normalizado}$ es el nivel de gris normal del punto considerado.

nivel_gris_medio es el nivel de gris medio de la imagen.

desviación es la desviación estándar de los niveles de gris de la imagen.

A cada nivel de gris de la imagen se le aplica la ecuación anterior para obtener los niveles de gris normalizados. El resultado es que todas las imágenes tienen por nivel de gris medio 0 y por desviación estándar 1, estamos haciendo que los niveles de gris de la imagen sigan una distribución Normal. La consecuencia que esto provoca es que no van a existir imágenes más claras o más oscuras, es decir, se minimiza la influencia de la iluminación existente en la etapa de adquisición de los datos.

Las imágenes una vez normalizadas se emplean como entrada a la red de neuronas empleada.

Caso 3: Operador de Gauss y Segmentación.

El proceso que se sigue en este caso lo refleja la figura 2.8.

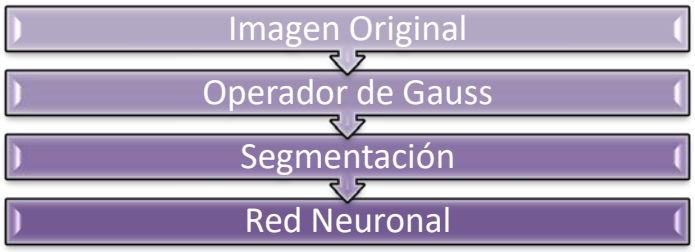

Figura 2.8. *Etapas de procesamiento mediante la utilización de una pirámide Gaussiana y una fase de Segmentación.*

En este caso además de reducir el tamaño de las imágenes por medio de la pirámide de Gauss, se incluyó una etapa adicional de segmentación de imágenes.

La segmentación realiza una partición de la imagen en regiones significativas. Segmentar una imagen es, por tanto, descomponerla en regiones homogéneas separadas por bordes. Esta particularidad de la técnica la hace muy interesante para el tratamiento de imágenes.

Las imágenes, tras la aplicación del operador de Gauss y su consiguiente reducción a un tamaño de 32x22 píxeles, se someten a un proceso de segmentación utilizando un método que proviene de la morfología matemática [24, 25]. Este método ofrece una segmentación diferente en función de un único parámetro que debe ser ajustado por el usuario dependiendo del tipo de imagen de que se trate, ya que lo que determina es la existencia de un mayor o menor número de regiones. Este parámetro se

denomina *parámetro de escala* y lo que hace es fijar, de forma indirecta, el tamaño más pequeño de las regiones, lo que establece el número máximo de regiones.

Valores grandes de este parámetro de escala provocan la disminución del número de regiones, en cambio valores pequeños del parámetro aumentan dicho número. También influye este factor de escala en la forma de los bordes de las distintas regiones, de forma que para valores pequeños los bordes de las regiones resultan más abruptos que para valores grandes, que dan lugar a suavidad en las fronteras entre regiones.

Para segmentar las imágenes se utilizó un programa que realiza esta función [24 – 26, 34 - 35], asignándole un valor al parámetro de escala de 10.

La siguiente figura muestra las imágenes que resultan tras pasar por las etapas del operador de Gauss y de segmentación, así como la imagen de partida esta última ha sido reducida y las dos primeras ampliadas para poder apreciar mejor los efectos de cada una de las etapas.

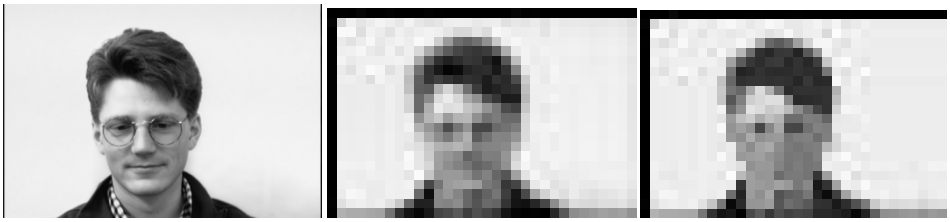

Figura 2.9. *Imágenes original y reducida mediante Gauss y Segmentada*.

Como se aprecia en la figura, a efectos visuales, el individuo queda prácticamente irreconocible tras la segmentación debido a que las regiones unifican toda la zona de imagen que incluyen a un mismo nivel de gris, perdiendo, por tanto, información que permite la posterior identificación del sujeto. Hecho que quedara constatado posteriormente con los resultados.

2.7. Red de Neuronas

Uno de los objetivos planteados era la solución del problema de la identificación utilizando para ello una red de neuronas artificiales, ya que la bibliografía demuestra que esta herramienta tiene un comportamiento próximo a la forma de actuar del cerebro [27].

2.7.1. Redes Empleadas

Una vez seleccionado el algoritmo que se desea emplear para entrenar la red caben dos posibilidades:

* Utilizar una herramienta, ya existente, que capaz de construir, entrenar y evaluar una red de neuronas.
* Implementar un programa que simule una red neuronal como la que se precisa para la solución del problema de identificación de caras planteado.

El presente trabajo hace uso de ambas opciones.

Como herramienta ya existente, se ha utilizado un paquete de software llamado **SNNS** (**S**ttutgart **N**eural **N**etwork **S**imulator) [28], elaborado por la Universidad de Sttutgart (Alemania) en el que están implementadas distintos tipos de redes neuronales, incluyendo las ya mencionadas redes de retropropagación.

Se construyó, utilizando este software, una red con 4 capas. La capa de entrada formada por 674 neuronas, por requisitos de las imágenes, la primera capa intermedia consta de 75 neuronas, la segunda de 15 neuronas y la capa de salida cuenta con 30 neuronas, tantas como el número de individuos que se trata de diferenciar.

La elección del número de neuronas de las capas intermedias se ha tomado al azar ya que, como se comentó anteriormente, no existe un criterio claro para determinar este número. El número total de neuronas de la red es, por lo tanto, de 794 neuronas, este número como se aprecia es elevado y más si consideramos el número de conexiones a las que darán lugar, sin embargo resulta abordable.

Como criterio de finalización se optó por establecer un número de iteraciones tras el cual se detiene el proceso de entrenamiento, a medida que se van ejecutando las iteraciones el programa va presentando una serie de valores que nos dan una idea de cómo va evolucionando el sistema. Por lo general muestra información relativa al error que se está cometiendo.

Una vez finalizado el entrenamiento se realiza un test para ver que el porcentaje de aciertos en el reconocimiento que tiene el sistema construido. Para realizar este test se utiliza una imagen de cada individuo, la imagen empleada es diferente de las utilizadas durante la fase de entrenamiento.

Este software nos permite almacenar los pesos de las conexiones una vez concluido el entrenamiento, de forma que si este hubiera sido insuficiente, podríamos reiniciarlo empleando los valores de los pesos almacenados, aprovechando de esta forma los entrenamientos anteriores.

La otra posibilidad, construir un programa que simule el funcionamiento de una red de neuronas, también se aborda en el presente trabajo.

A la hora de construir un programa lo primero que hay que decidir es el lenguaje en el que se va a implementar. La decisión no resulta trivial ya que cada lenguaje ofrece distintas prestaciones que lo hacen más o menos adecuado para resolver el problema en cuestión.

En este caso se decidió utilizar el lenguaje Ada en su versión orientada a objetos, ***Ada95*** [29 - 31].

El resultado es un programa modular formado por varios paquetes que contienen cada uno las funciones y procedimientos correspondientes a cada una de las tareas necesarias para la implementación de la red. Así contamos con los siguientes paquetes:

* Paquete Ada para manejar el menú de opciones (entrenar o test) : Menu.ad(b/s).
* Paquete Ada para realizar el entrenamiento: Entrenar.ad(b/s).
* Paquete Ada para realizar las pruebas: Test.ad(b/s).
* Paquete Ada con las variables requeridas por los otros paquetes: Variables.ads.
* Paquete Ada con el procedimiento para iniciar la ejecución: Rn.adb

El simulador se ha construido para implementar una red neuronal con una única capa intermedia, de forma que el esquema de esta red es:

- Capa de entrada de 674 neuronas
- Capa oculta de 100 neuronas
- Capa de salida de 30 neuronas

El algoritmo empleado para el entrenamiento y prueba de la red es el mismo que el utilizado con el paquete SNNS tomando, en este caso, un valor de 0.2 para la tasa de aprendizaje y como criterio de finalización utiliza el consistente en indicarle un número de iteraciones. Durante la fase de entrenamiento muestra el valor del error cuadrático medio que se produce tras procesar todos los patrones de entrada en cada

iteración. Para la fase de prueba el programa indica cuál es la neurona activa para cada patrón de test, presentando al final de la fase el porcentaje de aciertos conseguido.

El programa comienza mostrando un menú con tres opciones:

* Entrenar la red.
* Probar la red.
* Salir.

Para entrenar la red el programa en primer lugar solicita el nombre del archivo que contiene a los patrones que se van a emplear en esta fase, permite la posibilidad de generar aleatoriamente los valores de los pesos de las conexiones, esto es necesario si se trata de la primera vez que se entrenan esos patrones y no se dispone de valores para esos pesos, o bien utilizar los pesos almacenados en algún archivo, en cuyo caso será necesario indicarle el nombre del archivo. Tras esto el sistema solicita el número de iteraciones de que va a constar el entrenamiento, comenzando a continuación con el mismo.

Al finalizar el entrenamiento permite almacenar los pesos conseguidos en un archivo, previa indicación del nombre del archivo. Una vez hecho esto el programa vuelve al menú inicial.

En la fase de prueba se le suministra al programa el archivo que contiene las imágenes de los individuos a reconocer así como el archivo donde están guardados los pesos de las conexiones, el sistema propaga las imágenes a través de la red indicando para cada una, la neurona de la capa de salida con el mayor valor de activación.

Al finalizar muestra el porcentaje de aciertos que se ha producido, volviendo al menú inicial.

Las imágenes, tanto de entrenamiento como de test, están almacenadas en el archivo por orden alfabético de los nombres de los individuos, tomando como criterio que la primera neurona de la capa de salida corresponde al primer individuo según el orden alfabético y la última al último. Este criterio para la construcción de archivos de entrenamiento y pruebas se emplea también para el programa SNNS. De hecho, salvo la cabecera requerida por el SNNS para los archivos de patrones, indicando el número de neuronas de cada capa, los archivos son iguales.

2.7.2. Ada

Ada es un lenguaje extenso, ya que engloba muchos aspectos importantes relacionados con la programación de sistemas prácticos en el mundo real. Algunos puntos claves en Ada son:

***Legibilidad**: Se considera que los programas se leen muchas más veces de las que se escriben. Es importante conseguir evitar una notación demasiado concisa, que aunque permita escribir un programa rápidamente hace casi imposible leerlo.

***Tipado fuerte**: Esto asegura que todo objeto tenga un conjunto de valores que este claramente definido, e impide la confusión entre conceptos lógicamente distintos. Como consecuencia el compilador detecta mucho errores que en otros lenguajes darían como resultado programas ejecutables, pero incorrectos.

***Construcción de grandes programas**: Se necesitan mecanismos de encapsulado, para la compilación separada y para la gestión de bibliotecas, con vistas a poder escribir programas transportables y mantenibles de cualquier tamaño.

***Manejo de excepciones**: Es un hecho que los programas reales raramente son correctos. Es necesario proporcionar medios para que un programa pueda construirse en capas y por partes, de tal forma que se puedan limitar las consecuencias de los errores que se presenten en cualquiera de las partes.

***Abstracción de datos**: Tal como mencionamos anteriormente, se puede obtener mayor transportabilidad y mejor mantenibilidad si se pueden separar los detalles de la representación de los datos, de las especificaciones de las operaciones lógicas sobre los mismos.

***Procesamiento paralelo**: Para muchas aplicaciones, es importante que el programa se conciba como una serie de actividades paralelas, en vez de como una simple secuencia de acciones. Dotando al lenguaje de mecanismos al respecto se evita tener que añadirlos por medio de llamadas al sistema operativo, y con ello se consigue mayor transportabilidad y fiabilidad.

***Unidades genéricas**: En muchos casos la lógica de parte de un programa es independiente de los tipos de valores que están siendo manipulados. Por tanto, se necesita un mecanismo que permita la creación de piezas de programa similares a partir de un solo original. Esto es especialmente útil para la creación de bibliotecas.

Como se aprecia el lenguaje escogido proporciona múltiples ventajas, todas ellas encaminadas al desarrollo de programas de gran calidad software, así como de grandes programas que resuelvan problemas de muy diferente naturaleza, ya que el lenguaje tiene prestaciones capaces de abordar cualquier tipo de problema sin estar indicado para ninguno en concreto.

2.8 LVQ

El programa software empleado para implementar el método LVQ es el LVQ_PAK (Learning Vector Quantization Program Package) en su versión 3.1 de Abril de 1995 [32, 33]. Se trata de un programa desarrollado por el equipo de Programación en LVQ de la Universidad de Helsinki (Finlandia).

Este paquete está formado por una serie de programas que deben ejecutarse en el orden adecuado para el correcto funcionamiento del mismo. Existen dos posibilidades a la hora de ejecutar el conjunto de programas:

- Llamar a un programa ejecutor que se encargaría de pedir los parámetros precisos para la ejecución de forma interactiva.

- El usuario es el encargado de ir ejecutando los programas en el orden adecuado:

    - Programas de asignación de valores iniciales.
    - Programas de entrenamiento.
    - Programas de pruebas.

En la segunda opción, al principio de la ejecución de cada programa es necesario suministrarle al programa los parámetros y archivos que posteriormente va a emplear, esto se haría en la misma línea de comando al efectuar la llamada al programa en cuestión.

Algunos de los parámetros más importantes que utiliza el programa son:

noc : Número de vectores de código.

rlen : Número de iteraciones para el entrenamiento.

alpha : Tasa de aprendizaje inicial.

2.9. Resultados y discusión

En los apartados anteriores se comenzó poniendo de manifiesto un problema, como es el del reconocimiento de individuos de forma automática, se siguió indicando las posibles soluciones que se podían emplear para resolverlo y se seleccionaron algunas de ellas para aplicarlas al problema en cuestión. En este apartado se recogen los resultados obtenidos al aplicar esas técnicas empleando los datos presentados y los algoritmos descritos anteriormente. Tal y como se ha indicado se han utilizado tres tipos de procesamiento:

- Red Neuronal con una capa oculta
- Red Neuronal con dos capas ocultas
- LVQ

Los resultados que se obtuvieron se muestran a continuación.

Durante la fase de entrenamiento de las redes se fueron recogiendo datos relativos a los errores que se iban cometiendo para las iteraciones realizadas. Una vez entrenada la red se tomaron datos acerca del porcentaje de aciertos que se obtenía, lo que nos va a permitir contrastar porcentajes frente a número de iteraciones requeridas para tales porcentajes.

2.9.1. Resultados obtenidos mediante una red de neuronas con dos capas ocultas

2.9.1.1 Estudio del error cuadrático medio

Las siguientes tablas nos muestran el error cuadrático medio por unidad de la capa de salida en función del número de iteraciones.

IMAGENES ROTADAS CON GAUSS

ITERACIONES	SSE/UNIDAD DE SALIDA
1	12.5777
3001	5.5133
6001	5.4815
9001	5.4684
12001	5.4277
15001	5.4256
18001	5.3873

Tabla 2.1: *Error Cuadrático Medio por unidad de la capa de salida en función del número de iteraciones para imágenes rotadas reducidas mediante el operador de Gauss.*

IMAGENES FRONTALES CON GAUSS

ITERACIONES	SSE/UNIDAD DE SALIDA
1	12.2485
3001	5.9291
6001	6.0917
9001	6.0796
12001	6.0774
15001	6.0749
18001	6.0701
21001	6.0666

Tabla 2.2: *Error Cuadrático Medio por unidad de la capa de salida en función del número de iteraciones para imágenes frontales reducidas mediante el operador de Gauss.*

IMAGENES FRONTALES CON GAUSS Y NORMALIZADAS

ITERACIONES	SSE/UNIDAD DE SALIDA
1	12.9272
3001	0.2405
6001	0.2388
9001	0.2383
12001	0.2380
15001	0.2379
18001	0.2377
21001	0.2375
24001	0.2372
27001	0.2371
29701	0.2370

Tabla 2.3: *Error Cuadrático Medio por unidad de la capa de salida en función del número de iteraciones para imágenes frontales reducidas mediante el operador de Gauss y Normalizadas.*

IMAGENES ROTADAS CON GAUSS Y NORMALIZADAS

ITERACIONES	SSE/UNIDAD DE SALIDA
1	13.4489
3001	0.0026
6001	0.0012
9001	0.0008
12001	0.0006
15001	0.0005
18001	0.0004
21001	0.0003
24001	0.0003
27001	0.0002
29701	0.0002

Tabla 2.4: *Error Cuadrático Medio por unidad de la capa de salida en función del número de iteraciones para imágenes rotadas reducidas mediante el operador de Gauss y Normalizadas.*

La representación gráfica de la información suministrada por las tablas 2.1, 2.2 y 2.3 aparece en la siguiente figura:

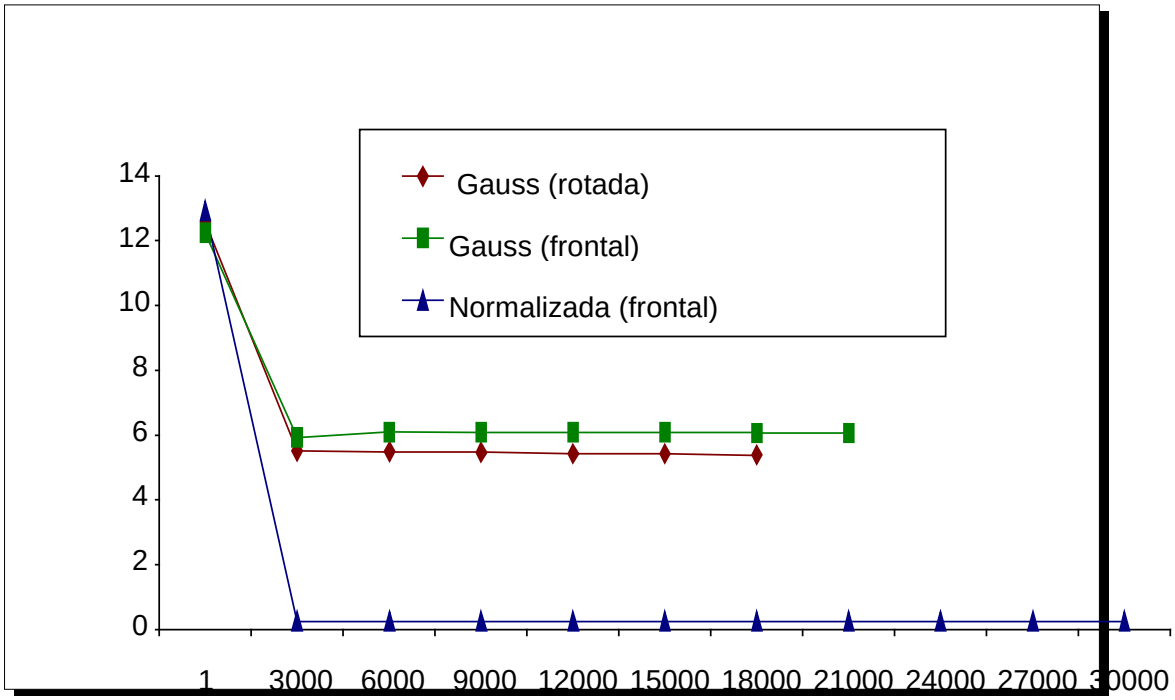

Figura 2.10*: Error Cuadrático medio por unidad de la capa de salida frente al número de iteraciones de entrenamiento para las tres redes neuronales utilizadas.*

La tabla correspondiente a imágenes rotadas a las que se le aplica el operador de Gauss y una etapa posterior de normalización no se incluye debido a los valores tan pequeños que se obtienen. Hay que destacar el hecho de que estas imágenes presentan el error cuadrático medio más pequeño de todas las empleadas. La forma de la curva sería muy similar a la originada por la tabla 2.3.

Las tablas anteriores nos permiten observar varias cuestiones:

1º.- Los errores iniciales son elevados, con anterioridad se comentó que esta situación se produciría debido a que inicialmente se toman valores aleatorios para los pesos.

2º.- Las Redes Neuronales dependen del tipo de procesamiento al que se someta a los datos, la propia experiencia, así como la bibliografía, nos permite afirmarlo, cuando se normalizan las imágenes el error disminuye.

2.9.1.2 Estudio del porcentaje de aciertos

No solamente se han recogido datos sobre los errores, sino también sobre el porcentaje de aciertos que se produce en la salida de la red de neuronas. Estos son presentados en las siguientes figuras:

TIPO DE PROCESAMIENTO	**PORCENTAJE DE ACIERTOS**
Gauss (frontal)	43.33
Gauss (rotada)	26.66
Normalizado (frontal)	83.33
Normalizado (rotada)	80.00

Tabla 2.5*: Resultados alcanzados (en % de aciertos) para cada tipo de procesamiento.*

El diagrama de barras correspondiente a la tabla anterior sería:

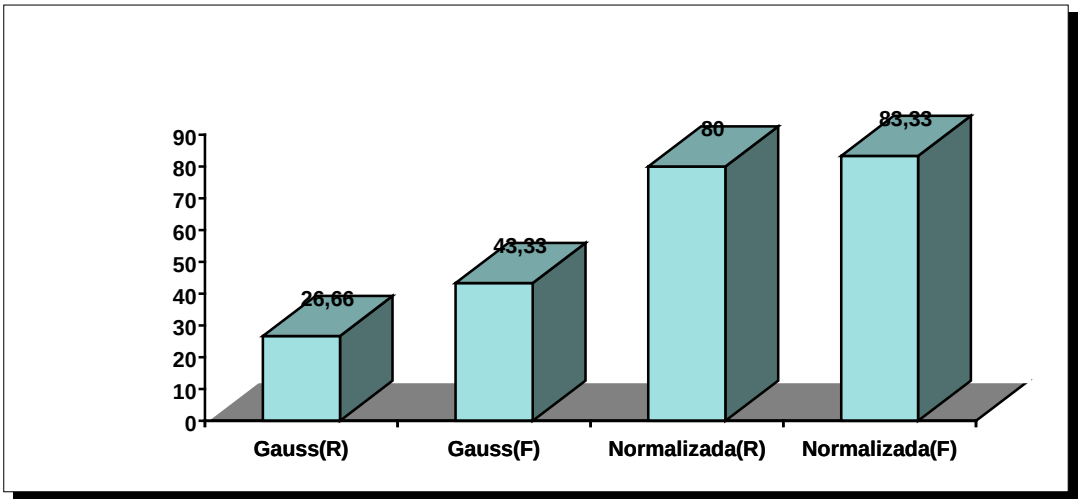

Figura 2.11: *Diagrama de barras del porcentaje de aciertos de las distintas técnicas empleadas.*

El mayor porcentaje de aciertos se produce en las imágenes normalizadas. Esto es debido a que la normalización uniformiza los niveles de gris de las imágenes, haciendo que influyan menos los contrastes que puedan existir en las mismas, lo que lógicamente origina un aumento en el porcentaje de aciertos.

2.9.2 Resultados obtenidos mediante la técnica LVQ

Para aplicar el algoritmo LVQ se emplearon uno, dos y tres vectores de código por clase para las imágenes, los porcentajes de aciertos que se consiguieron en cada caso los muestra la tabla siguiente:

Tipo de procesamiento	LVQ con un vector de código por clase	LVQ con dos vectores de código por clase	LVQ con tres vectores de código por clase
Gauss (frontal)	96.67	96.67	96.67
Gauss(rotada)	83.33	96.67	96.67
Normalizado (frontal)	93.33	96.67	96.67
Normalizado (rotada)	90.00	93.33	93.33
Segmentación	40.00	40.00	40.00

Tabla 2.6: *Porcentaje de acierto empleando distinto número de vectores de código por clase.*

La tabla anterior se puede expresar de forma gráfica mediante el diagrama de barras que se presenta a continuación:

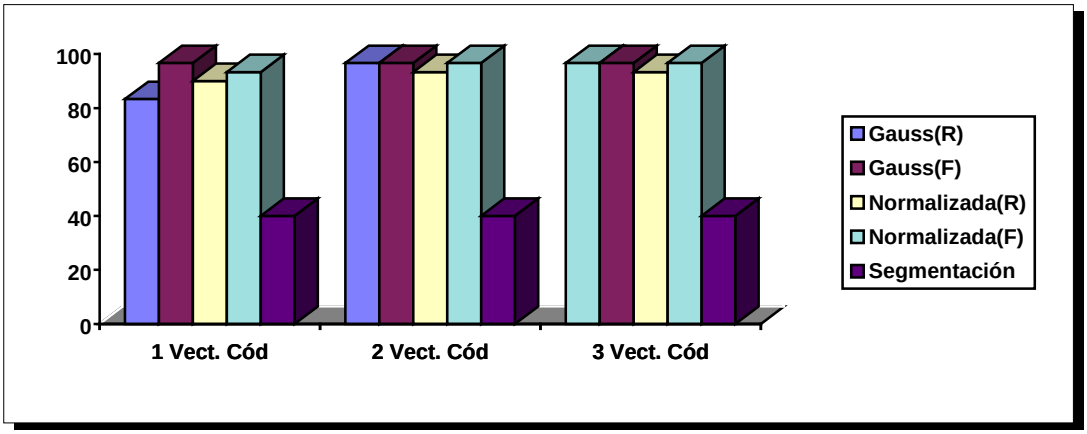

Figura 2.12. *Diagrama de barras correspondiente a la tabla 6.6.*

La técnica LVQ ofrece resultados mucho mejores tanto en el porcentaje de aciertos como en el tiempo necesario para el reconocimiento. En una Red Neuronal se precisan varios días o incluso semanas para entrenar la red, mientras que el LVQ sólo emplea minutos o en el peor de los casos horas.

En la gráfica también se pone de manifiesto que el LVQ es insensible al preprocesamiento al que se sometan las imágenes (normalizado) pero no a la segmentación. Lo mismo ocurre para el número de vectores de código que empleemos, los resultados no se ven afectados en gran medida.

2.9.3 Resultados obtenidos con una red neuronal de una capa oculta

El programa en Ada que se construyó se empleó para tener un criterio de comparación con el programa SNNS. La comparación se hizo empleando aquella técnica que había proporcionado mejores resultados para la red con dos capas ocultas. Al igual que el programa SNNS nos permite recoger información referente al error que

se comete en cada una de las etapas además del porcentaje de aciertos que se obtiene. Los resultados del programa construido se muestran a continuación.

2.9.3.1 Estudio del error cuadrático medio

La tabla siguiente nos muestra el error cuadrático medio que se produce en función del número de iteraciones que se han empleado en el entrenamiento de la red.

Numero de Iteraciones	**Error Cuadrático Medio**
1	0.0619
100	0.0130
200	0.0126
300	0.0107
400	0.0104
500	0.0093
600	0.0092
700	0.0082
800	0.0081
900	0.0072
1000	0.0071

Tabla 2.7. *Error cuadrático medio en función del número de iteraciones.*

El diagrama de barras correspondiente sería:

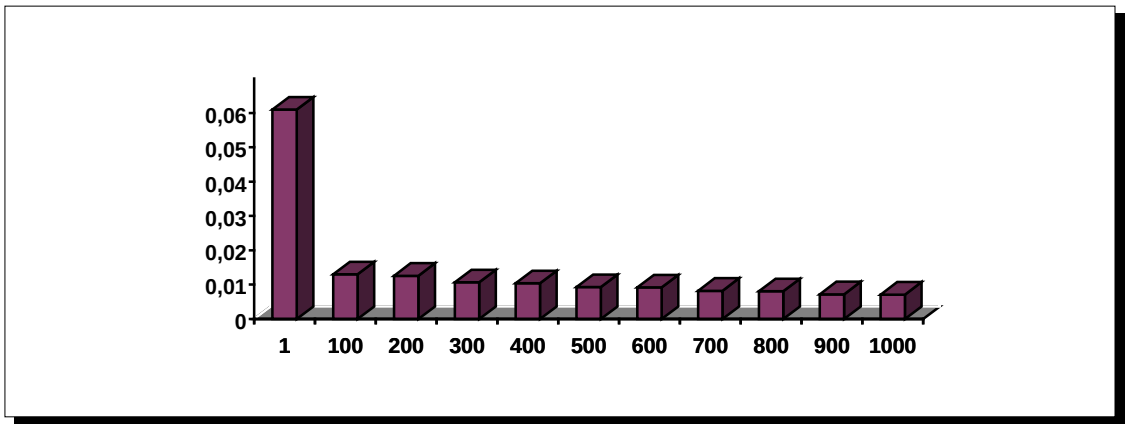

Figura 2.13. *Error Cuadrático frente al número de iteración*

Observamos a partir de la tabla que en este caso el error cuadrático medio es menor que en los casos anteriores, siendo su valor inferior tanto al comenzar el entrenamiento como al finalizarlo. Se observa también que se mantiene la circunstancia de que el primer valor es considerablemente mayor que el segundo valor (considerando los órdenes de magnitud en los que nos movemos) esto nuevamente es debido a la aleatoriedad de los valores de los pesos de las conexiones con los que comienza el entrenamiento.

En este caso sólo se han empleado 1000 iteraciones para entrenar la red frente a las 30.000 empleadas para la red de dos capas ocultas. Obteniéndose en este caso una constante reducción del error sin que se produzcan mínimos locales, también hay que indicar que es más lenta que la red de dos capas.

2.9.3.2 Estudio del porcentaje de aciertos

En este caso se tomaron resultados relativos al porcentaje de aciertos tras ser entrenada la red con diferentes números de iteraciones, siendo tales números 3, 23, 30, 200 y 1000. Los resultados obtenidos se muestran en la siguiente tabla.

Numero de Iteraciones	% de Aciertos
3	23.33
23	50
30	53.33
200	63.33
1000	73.33

Tabla 2.8. *Porcentaje de aciertos para distintos números de iteraciones empleados en el entrenamiento.*

La gráfica que refleja la información de la tabla anterior es:

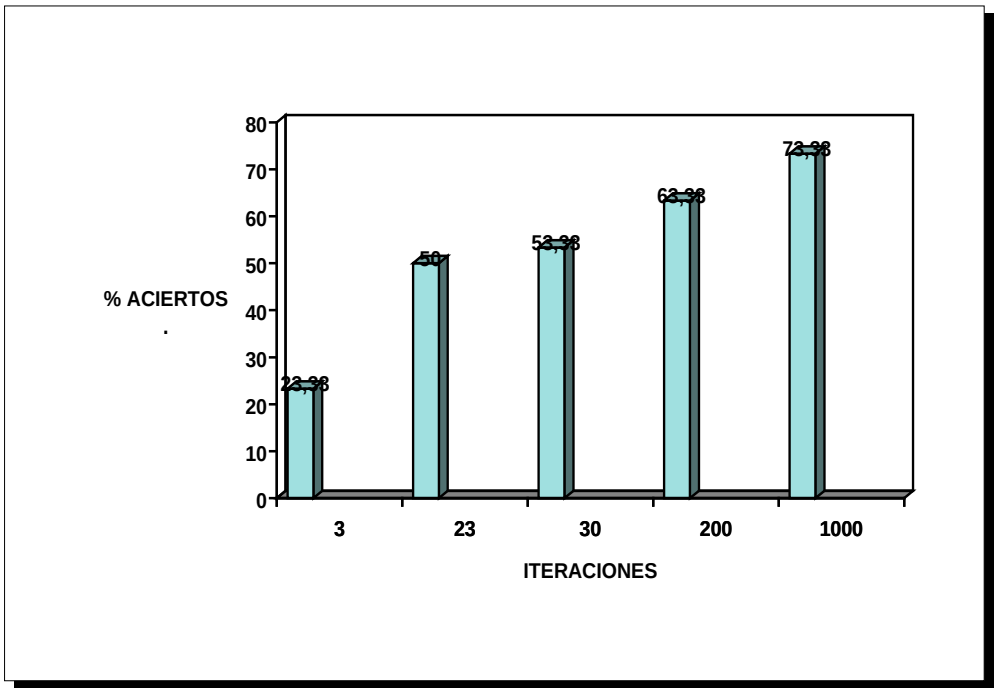

Figura 2.14. *Porcentaje de aciertos en función del número de iteraciones.*

Obviamente, a medida que aumenta el número de iteraciones aumenta el porcentaje de aciertos, pero también se incrementa el tiempo necesario para entrenar la red, señalando el hecho de que el tiempo necesario para entrenar la red con una sola capa ha resultado mayor que el necesario para la red con dos capas ocultas.

2.9.4. Comparación de las tres técnicas empleadas

Uno de los objetivos consistía en realizar un estudio comparativo entre los tres tipos de procesamiento para saber cual era mejor. Los resultados mostrados en los apartados anteriores de este capítulo nos permiten construir una tabla que nos muestre los resultados ofrecidos por los distintos métodos para los mismos datos de entrada.

	Red Neuronal 1 capa oculta	**Red Neuronal 2 capas ocultas**	**LVQ**
Gauss (frontal)	73.33 %	43.00 %	96.00 %

Tabla 2.9. *Porcentaje de aciertos para imágenes frontales y Gauss empleando las distintas técnicas.*

Gráficamente la tabla anterior nos la muestra la siguiente figura.

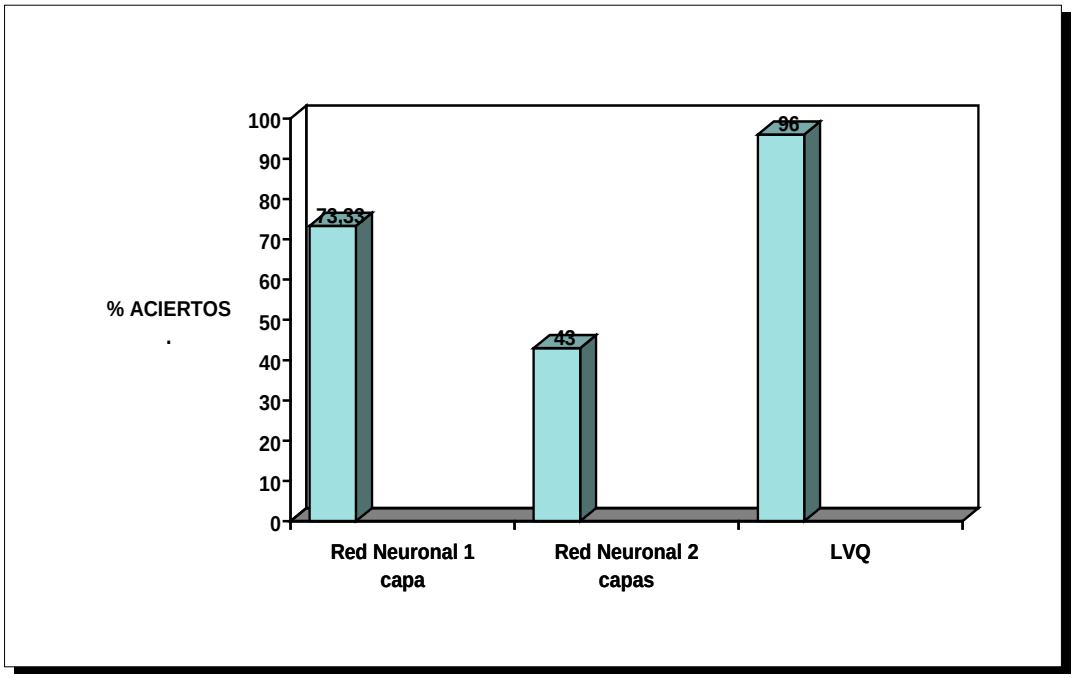

Figura 2.15. *Porcentaje de aciertos para imágenes frontales y Gauss empleando las distintas técnicas.*

Los resultados conseguidos reflejan que la mejor de las técnicas empleadas, considerando el porcentaje de aciertos que ha logrado, ha resultado ser LVQ (96 %), mientras que la peor es la Red Neuronal con dos capas (43 %). La Red Neuronal con una capa oculta se mantiene en una posición intermedia entre ambas con un porcentaje del 73.33 %.

En el caso de las Redes Neuronales destacar que para la red de dos capas ocultas se han empleado 18.000 iteraciones en imágenes tratadas simplemente con el operador de Gauss y cerca de 30.000 iteraciones empleando las técnicas de segmentación y normalización. Sin embargo, para la red de una capa oculta han sido necesarias tan sólo 1.000 iteraciones para casi duplicar los resultados obtenidos con la red de dos capas ocultas.

2.10. Conclusiones

El trabajo presentado muestra un método efectivo de reconocimiento de bajo nivel para caras humanas basado directamente en niveles de gris. Con el adecuado procesamiento se puede minimizar la influencia de las condiciones de adquisición de imágenes en el proceso de reconocimiento. Este procesamiento de bajo nivel puede verse complementado con posteriores etapas de reconocimiento de alto nivel.

Las conclusiones a las que se llega tras analizar los resultados han sido apuntadas en el apartado anterior a medida que se presentaban dichos resultados. Podemos, no obstante, agruparlas y generalizarlas en los puntos siguientes:

1.- Se ha comprobado que es posible el diseño de un sistema de reconocimiento de caras humanas basado en imágenes utilizando una reducción Gaussiana.

2.- Los resultados muestran que tanto las Redes Neuronales como el método LVQ son adecuados para resolver el problema que dio lugar a este trabajo.

3.- El método LVQ proporciona mejores resultados que las Redes Neuronales, siendo también más robusto frente a cambios introducidos en las imágenes durante la adquisición de las mismas.

4.- Otra ventaja que presenta el LVQ frente a las Redes de Neuronas y que no se aprecia en las tablas y gráficas expuestas en el capítulo anterior, aunque sí se apunta en algún momento, es que el tiempo que requiere para el entrenamiento es mucho menor, diferenciándose en varios órdenes de magnitud.

5.- Tener en cuenta los pobres resultados que se obtienen segmentando las imágenes.

6.- Destacar que los sistemas empleados consiguen buenos resultados con imágenes que presentan ángulos de inclinación altos, lo que demuestra la robustez de los mismos frente a cambios de este tipo.

7.- El software diseñado se ha realizado dentro de una normativa adaptándola al problema que se estaba considerando.

Para finalizar, comentar que los resultados obtenidos con este trabajo han dado lugar a una serie de artículos, capítulos de libros y ponencias a congresos que se citan a continuación.

Libros y revistas:

Conde, C.; Sanchez, A.; Cabello, E. “Influence of location over several classifiers in 2D and 3D face verification”. 3161, pp. 153 - 158. Springer Lecture Notes Computer Science GuidelinesSpringer Verlag, 05/2005 .

Tipo de producción: Capítulos de libros

Tipo de soporte: Libro

Enrique Cabello; M. Araceli Sánchez; Luis Pastor.” Some experiments on face recognition with neural networks”. pp. 589 - 599.Springer Verlag, 1998.

Tipo de producción: Capítulos de libros

Tipo de soporte: Libro

Enrique Cabello; M. Araceli Sánchez; Luis Pastor.”Reconocimiento de caras humanas mediante una red neuronal con Ada95”. Revista Ada Spain. 35, pp. 29 - 38.1998.

Tipo de producción: Artículo

Tipo de soporte: Revista

Congresos:

Título: Adaptación de modelos geométricos a imágenes: aplicación a caras humanas

Nombre del congreso: XX Jornadas de Automática.

Ciudad de realización: Salamanca, España

Fecha de realización: 09/1999

Javier Gómez; M. Araceli Sánchez; Enrique Cabello.09/1999.

Título: Supervised methods for face recognition using geometric characteristics.

Nombre del congreso: IASTED International Conference on Signal Processing and Communications Sponsored by IASTED, ULPGC, IAC, IEEE.

Ciudad de realización: Canarias, España

Fecha de realización: 02/1998

Ciudad: España

Enrique Cabello; M. Araceli Sánchez; Angel Luis Labajo; Luis Pastor; Juan Alonso.02/1998.

Título: Modelos conexionistas para el reconocimiento de caras.

Nombre del congreso: IX Congreso de la sociedad española de psicología comparada.

Ciudad de realización: Salamanca, España

Fecha de realización: 09/1997

Enrique Cabello; Araceli Sánchez; Luis Pastor.09/1997.

Título: Reconocimiento de caras humanas: una aproximación por medio de redes neuronales.

Nombre del congreso: VII RPIC (Reunión de Trabajo en Procesamiento de la Información y Control). **Fecha de realización:** 09/1997

Araceli Sánchez; Enrique Cabello; Luis Pastor. Lugar: San Juan; Argentina.09/1997.

Título: Automatic face recognition using neural networks: gray level images versus geometric characteristics.

Nombre del congreso: 15 IMACS WORLD CONGRESS on Scientific Computation, Modelling and Applied Mathematics. Sponsored by IMACS, DFG, IEEE, IFAC, IFIP, IFORS, IMEKO.

Ciudad de realización: Berlín, Alemania

Fecha de realización: 08/1997

Enrique Cabello; M. Araceli Sánchez; Luis Pastor; Juan Alonso.08/1997.

Título: Automatic face recognition using neural networks: gray level images versus geometric characteristics.

Nombre del congreso: FACE RECOGNITION. **Ciudad de realización:** Sirling, Reino Unido

Fecha de realización: 07/1997

Entidad organizadora: OTAN

Enrique Cabello; M. Araceli Sánchez; Luis Pastor; Juan Alonso.07/1997.

Título: Una red neuronal con Ada95, aplicación al reconocimiento de caras humanas.

Nombre del congreso: Jornadas Técnicas de Ada Spain.

Ciudad de realización: Madrid, España

Fecha de realización: 02/1997

Araceli Sánchez; Enrique Cabello; Luis Pastor.02/1997.

Título: Procesamiento Digital de Imágenes: aplicación de redes neuronales al reconocimiento de caras humanas.

Nombre del congreso: XIV Congreso de la Sociedad Española de Ingeniería Biomédica

Ciudad de realización: Navarra, España

Fecha de realización: 09/1996

Enrique Cabello; Araceli Sánchez; Luis Pastor.09/1996.

CAPITULO 3

UNA NUEVA APROXIMACIÓN A LA IDENTIFICACIÓN DE ROCAS GRANDES CON APLICACIÓN EN LA INDUSTRIA MINERA

3.1.- Introducción

Una tarea importante en la industria minera es el machaqueo de la roca que se obtiene en la mina, donde el exceso de tamaño de algunos bloques puede dar lugar a importantes atascos en la alimentación de la maquinas de machaque lo que supone un elevado coste. Este capítulo presenta un sistema de visión por computador para estimar el tamaño de las rocas a medida que éstas van avanzando por el alimentador de una machacadora de impactos.

ENUSA (Compañía Nacional de Uranio de España), en la Planta Quercus, posee una mina a cielo abierto cerca de Salamanca (España). Las rocas provienen directamente de la mina y son depositadas en una gran tolva. Mediante un alimentador vibrante y un precribador se alimenta una machacadora de impactos de gran tamaño.

Figura 3.1. *Tolva.*

El problema surge cuando una roca tiene un tamaño mayor que la boca de la machacadora pudiendo bloquearla. La solución empleada hasta el momento consistía en que un trabajador debía estimar visualmente el tamaño de las rocas y detener el alimentador y precribador si detectase una que, a su juicio, pudiera obstruir la entrada en la máquina y proceder a su reducción mediante un martillo hidráulico montado sobre un brazo articulado.

Figura 3.2. *Empleado en el interior de la tolva.*

La existencia de dos líneas en paralelo con las características descritas hacía difícil que un único trabajador pudiera mantener la atención suficiente como para detectar de forma eficiente todas las rocas de gran tamaño. Una solución más eficiente es la inspección automática usando procesamiento digital de la imagen.

Un sistema de visión por computador puede ayudar al trabajador a detectar las rocas de gran tamaño, aumentando la seguridad y ahorrando dinero a la compañía. La gran reducción de coste de los incidentes compensa el pequeño desembolso que supone la adquisición de un ordenador y una cámara y la implementación de un sistema de visión.

El proceso ha sido automatizado por medio de un sistema de visión artificial que estima el tamaño de las rocas, controla los alimentadores y precribadores y detecta cualquier roca con tamaño mayor que la boca de las machacadoras. El sistema de visión satisface los requisitos de tiempo impuestos por la industria para trabajar en tiempo real, así como los debidos a la extrema dureza ambiental como son:

- Polvo en el aire, que puede ocultar la visibilidad de la escena total o parcialmente, especialmente cuando un camión descarga las rocas en la tolva.
- El área cubierta por la cámara debe ser iluminada constantemente para impedir la perdida de información. En caso contrario, en nuestro sistema el sol podría iluminar las tolvas modificando la cantidad de luz.
- Las rocas tienen colores no uniformes. En nuestro caso las rocas en ocasiones están mojadas y son por lo tanto casi negras, lo que hace muy difícil la identificación de su forma.

El sistema, basado en una mezcla de técnicas de procesamiento de imágenes con redes neuronales, funciona de la siguiente manera: una vez que la imagen es obtenida, se realiza una etapa de preprocesado, filtrando la imagen y extrayendo un conjunto de rocas candidatas. A continuación una red neuronal procesa las rocas

candidatas para asegurar una correcta detección. Se aplica un algoritmo de seguimiento para impedir falsas detecciones debidas al agrupamiento de rocas. Utilizando información geométrica, es posible estimar las dimensiones reales de las rocas. Nuestro sistema de visión por computador satisface las necesidades de tiempo impuestas por la industria en tiempo real y ha sido puesto en funcionamiento en la mina. El algoritmo presentado es independiente de la forma de la roca. Se presentan los resultados obtenidos durante nueve meses de trabajo sin supervisión, mostrando que nuestro sistema es apto para trabajar bajo diferentes condiciones de luz, y es suficientemente fiable en condiciones de trabajo reales.

3.2. Estado del arte

Las industrias mineras son un campo de interés para el uso de sistemas de visión artificial en una gran cantidad de actividades [1]. En particular, las técnicas de visión artificial podrían ser empleadas para mejorar la calidad del mineral y para evitar operaciones de riesgo. La estimación del tamaño de una roca en una cinta transportadora es útil para evitar situaciones de bloqueo o para determinar la distribución granulométrica de las rocas. Diferentes trabajos han considerado varios algoritmos para detectar el tamaño de una roca en una cinta transportadora.

Un algoritmo de clasificación de imágenes ha sido presentado por Wang y Stephansson [2] para estimar las características de fragmentos de roca (distribución de tamaño y forma), pero sus resultados dependen en gran medida de la calidad de la imagen. Una técnica basada en las sombras que rodean una roca fue presentado por Wu *et al*. [3] pero es bastante sensible a las condiciones de la luz y a la textura de la roca. Un sistema multirresolución presentado por Crida *et al*.[4, 5] está basado en el análisis de 12 características y en el conocimiento de las características de las rocas estudiadas. En este caso se deben elegir umbrales experimentales para cada caso particular.

Crida y De Jager [6] y Fernandez *et al*.[7] han utilizado redes neuronales artificiales en este tipo de estudios. Crida y De Jager [6] consideraron una red neuronal

entrenada con un algoritmo de retropropagación para detectar rocas. Se emplea un paso de preprocesamiento para mejorar el contraste y la detección de los bordes. Este trabajo no proporciona resultados numéricos, pero los autores declararon que las redes neuronales artificiales por sí solas no fueron suficientes para reconocer una roca.

Fernandez *et al*. [7] utilizó una técnica de codificación para permitir que una red neuronal artificial estime la distribución granulométrica de una imagen.

En este caso, la distribución métrica de rocas en una cinta transportadora se calculó con excelentes resultados bajo condiciones de luz controlada.

3.3. Equipamiento y entorno

En la mina de uranio objeto del trabajo hay dos tolvas con dos cintas transportadoras situadas en paralelo.

El sistema se aplica de forma independiente a cada una de las 2 tolvas, también independientes entre sí, dada la disposición funcional simétrica del entorno. De esta forma cualquier problema o mal funcionamiento de una de las tolvas no impide el funcionamiento del sistema en la otra, de manera que cuando se produce una incidencia no es necesario detener completamente la producción, sino que solamente se reduce el ritmo de producción.

El sistema de visión que se ha instalado está constituido por:

2 Cámaras CCD en B/N. Permiten tomar imágenes de cada una de las tolvas.

- **1** Ordenador Personal (PC) basado en un procesador Pentium 166 MMX ™ y Pentium II ™.

- **2** Tarjetas de adquisición y digitalización de imágenes Matrox Meteor™. Estas tarjetas permiten la comunicación entre cada cámara y el ordenador.
- **2** Tarjetas digitales M1701. Utilizadas para la comunicación a través del puerto serie.

Cada cámara se conecta a una tarjeta de adquisición y enfoca a un alimentador. La resolución de la imagen está fijada en 640x480 pixeles con 256 niveles de gris.

El sistema se ha ubicado en la burbuja de control situada en la planta trituradora. Desde este lugar se controla el funcionamiento de las tolvas y del martillo neumático por parte de un operario, que es también el encargado de controlar el sistema. La acción del operario sobre el sistema se limita a su puesta en funcionamiento y su parada, y la confirmación de las posibles situaciones de bloqueo detectadas por el software.

Una visión esquemática del sistema se muestra en la figura 3.3. Únicamente se instaló un monitor ya que así el trabajador puede ver las dos imágenes de un único golpe de vista.

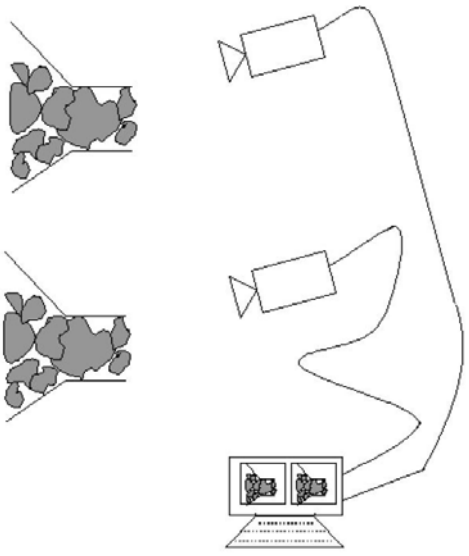

Figura 3.3. *Sistema de visión.*

La figura 3.4 muestra el alimentador en funcionamiento. Podemos observar en el centro derecha de la imagen dos cadenas metálicas cuyo efecto es evitar avalanchas de rocas y que se superpongan unas sobre otras. Además, como las cadenas barren las rocas, obligan a que las dimensiones más grandes de las rocas se encuentren en el plano de visión. Es casi imposible que la longitud mayor de la roca permanezca perpendicular al plano de visión ya que en ese caso las socas no se mantendrían en equilibrio.

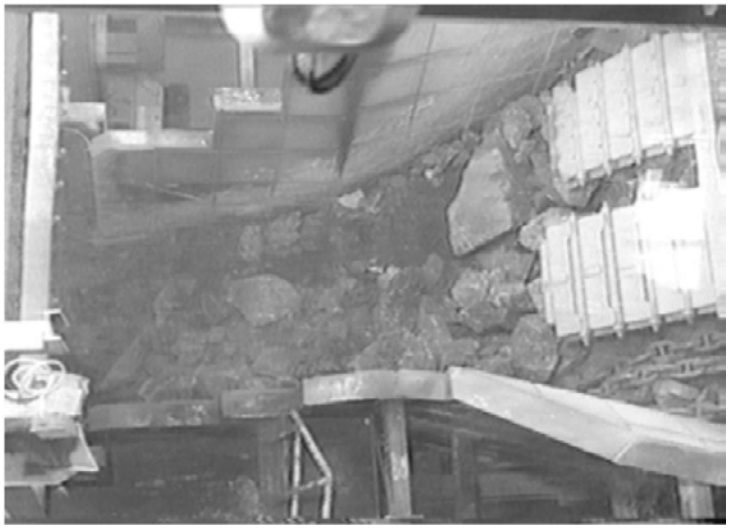

Figura 3.4. *Un alimentador en funcionamiento.*

Hay que tener en cuenta que el tamaño de los alimentadores es de 2.5 m de ancho y 6.5 m de largo. La velocidad de las rocas en los alimentadores es difícil de medir y de hecho este dato se desconocía al principio del proyecto. El sistema realiza una estimación del dicha velocidad entre 0.5 y 0.8 m/s, dependiendo de tamaño y número de rocas.

El sistema logra una velocidad de procesamiento de 8 imágenes por segundo (4 imágenes/s por cada tolva) suficientes para controlar los dos alimentadores, ya que en el peor de los supuestos las rocas habrían avanzado 25 centímetros entre toma y toma.

Debido al tamaño y localización de los alimentadores no es posible controlar eficientemente las fuentes de luz. Los alimentadores están situados en un edificio que el

sol ilumina de forma diferente a lo largo del día. La solución a este problema fue la instalación de fuentes de luz sobre los alimentadores. Aunque se colocaron cuatro focos para iluminar la escena, al final de la tarde el sol ilumina directamente los alimentadores y precribadores, saturando las CCDs. Ha resultado ser un obstáculo insalvable en el sistema, la ventaja es que la mina funciona las 24 horas del día y como la energía eléctrica es más barata por la noche, la mayor parte del trabajo se realiza a esas horas. En este caso, los focos son fuentes de luz adecuadas.

3.4. Algoritmos utilizados

El programa desarrollado ha sido estructurado en tres etapas dependiendo del tratamiento aplicado a las imágenes: adquisición, preprocesamiento (para obtener un conjunto de ventanas con "posibles rocas" en la imagen) y procesamiento (utilizando una red neuronal artificial y un algoritmo de seguimiento para asegurar una correcta detección). El esquema de la figura 2.5 muestra la descomposición del algoritmo. Cada fase se describe brevemente a continuación.

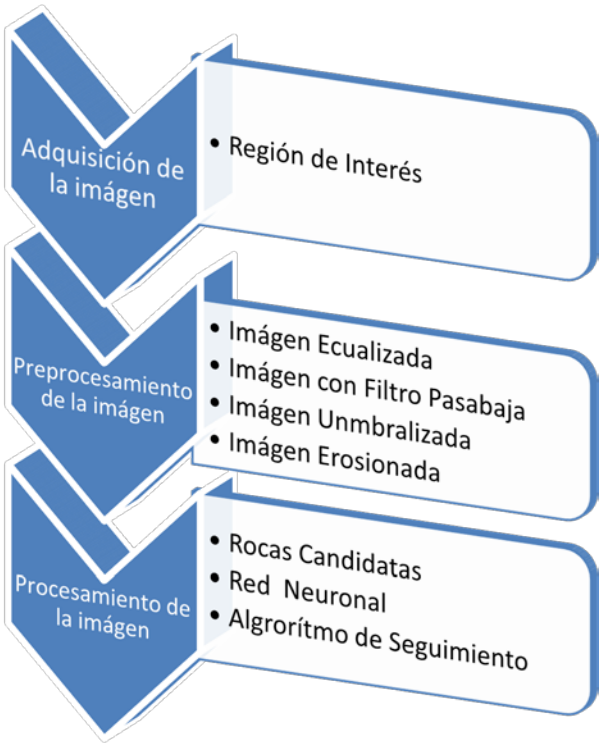

Figura 3.5. *Esquema del proceso*

Una vez que el preprocesamiento está finalizado se dispone de un conjunto de regiones candidatas a ser consideradas rocas grandes en la imagen. La fase de procesamiento se inicia a partir de ese conjunto de regiones candidatas y la imagen inicial en niveles de gris. El resultado debe ser la correcta detección de rocas grandes en el alimentador.

3.4.1. Adquisición de imágenes

El ordenador adquiere y procesa alternativamente una imagen de cada cámara. Como las cámaras están firmemente sujetas en la pared, el alimentador se ve siempre en la misma posición. El sistema solo procesa la denominada *región de interés*. Se puede definir una región de interés conteniendo solo el alimentador de tal modo que la zona de la imagen inicial fuera de la región de interés es eliminada, y por lo tanto el tiempo de procesamiento se reduce sin pérdida de información. La forma de la región de interés se almacena en un archivo de configuración, que se carga al iniciarse el programa. Un ejemplo puede verse en la figura 3.6.

Figura 3.6. *Imagen inicial con los límites de la región de interés.*

La velocidad de procesamiento es independiente del número de rocas presentes en la imagen pero depende del número de pixeles de la región de interés, de ahí la importancia de eliminar las partes que no aportan información.

3.4.2. Preprocesamiento de imágenes

En esta fase se aplican algoritmos rápidos y se obtiene información sobre las regiones en la imagen que pueden contener rocas. Algunos de los algoritmos utilizados son comunes en visión artificial (ecualización del histograma y filtro pasa baja [8]) pero otros han tenido que ser adaptados a nuestro problema (binarización con dos umbrales y algoritmo de separación entre regiones) [1, 2, 3].

El preprocesamiento es una etapa crítica en la cual la entrada es una imagen de niveles de gris y la salida final es un conjunto de ventanas de la imagen que pueden contener una roca (un conjunto de “posibles rocas”). La fase de preprocesamiento se enfrenta con los problemas relacionados con los niveles de gris de la imagen y por lo tanto se debe diseñar un algoritmo robusto. Como se ve en la figura 3.5, esta etapa se ha dividido en varios procesos, cada uno de los cuales representa la aplicación de un algoritmo para separar las distintas rocas que pueden aparecer en la imagen.

Se consideraron diferentes opciones (varios detectores de bordes y algoritmos de segmentación entre otros) pero los resultados no fueron adecuados para este proyecto. Por ejemplo, los detectores de bordes ofrecen un conjunto de múltiples segmentos, pero muy pequeños y sin ningún orden. Los algoritmos de segmentación o bien ofrecen un gran número de regiones pequeñas o bien requieren mucho tiempo de procesamiento, lo que los descartó para la aplicación en un sistema de tiempo real.

Además, las rocas mojadas son casi negras y los algoritmos de segmentación tienden a agrupar las rocas con el fondo en la misma región.

Los procedimientos que mejores resultados proporcionaron fueron los ya mencionados en la figura 3.5 y que se desarrollan a continuación.

Ecualización del histograma

El histograma se ecualiza [8, 9] con el fin de obtener una distribución uniforme de niveles gris a lo largo de los pixeles de la imagen, esto implica un incremento en el contraste que mejora la diferenciación de las rocas. La figura 3.7 muestra el resultado de este proceso aplicado a la misma imagen de la figura 3.6. Con este paso se persigue mitigar los efectos en los cambios en las condiciones de luz que se producen durante el periodo de funcionamiento de la tolva. En las imágenes correspondientes a rocas mojadas se consigue aumentar de forma general el brillo de la imagen.

Sin embargo, la ecualización no funciona bien cuando hay una zona con incidencia de luz solar y otra parte de la imagen en sombra, en nuestro caso este efecto no es muy importante ya que prácticamente todo el tiempo tenemos la imagen completa con luz solar o en sombra, por lo que este algoritmo proporciona resultados aceptables.

Figura 3.7. *Imagen ecualizada.*

Filtro Pasabaja

Se trata de una función encargada de suavizar la imagen, eliminando zonas de ruido asociadas a componentes de alta frecuencia de la imagen tal y como se puede observar en la figura 3.8. La información de las frecuencias altas representa rápidas variaciones pixel a pixel asociadas al ruido o a la reflexión de la luz. Por el contrario, la información de las bajas frecuencias representa la forma de las rocas y se conseserva. Este filtro se implementa mediante la convolución de la imagen con un filtro cuyo núcleo es

$$\frac{1}{10}\begin{bmatrix}1 & 1 & 1\\ 1 & 2 & 1\\ 1 & 1 & 1\end{bmatrix}$$

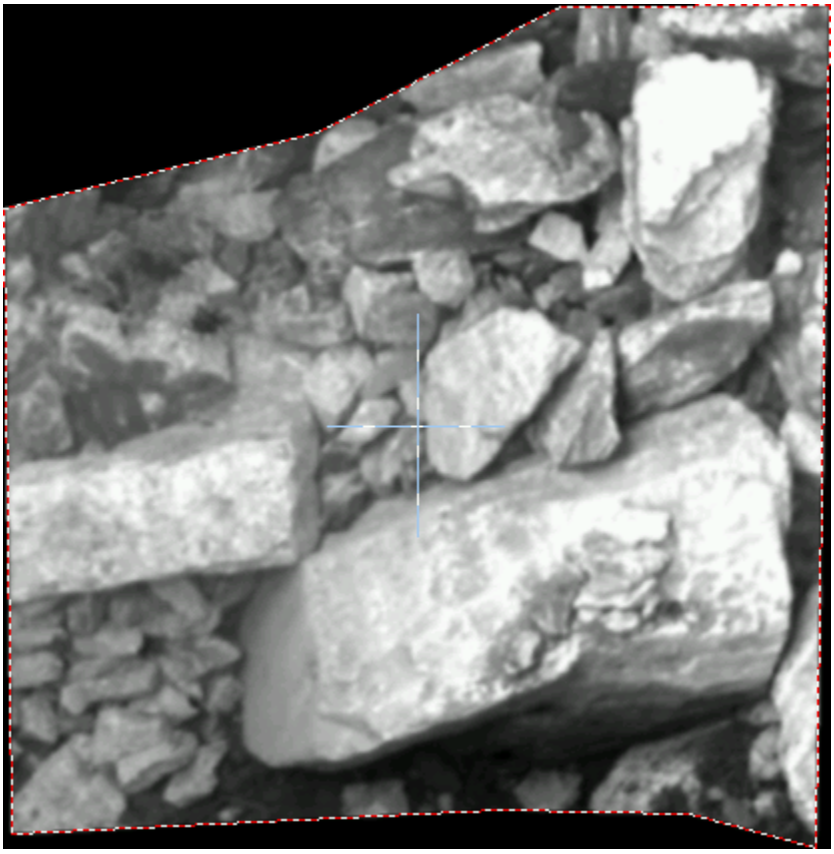

Figura 3.8. *Imagen tras aplicar el filtro pasabaja*

Se consideraron otras posibilidades para en núcleo del filtro que fueron las siguientes

$$\frac{1}{16}\begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix} \qquad \frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix} \quad \text{y} \quad \frac{1}{100}\begin{bmatrix}1 & 2 & 4 & 2 & 1\\ 2 & 4 & 6 & 4 & 2\\ 4 & 6 & 8 & 6 & 4\\ 2 & 4 & 6 & 4 & 2\\ 1 & 2 & 4 & 2 & 1\end{bmatrix}$$

Sin embargo fueron desechadas por los resultados que proporcionaban.

Bilevel theresholded

Tras los pasos anteriores se aplica un umbral de dos niveles cuyo resultado lo muestra la figura 3.9. El objeto de esta etapa es el de aislar las rocas presentes en la imagen. Las rocas aparecen destacadas como regiones blancas sobre un fondo negro.

Trabajos previos de otros autores [10, 11] para automatizar el cálculo del umbral, así como los desarrollados empleando la media y la mediana [12] fueron considerados. Pero ninguno de ellos ofrecía buenos resultados para las características de nuestro problema particular.

Varios autores [13 - 15] han realizado comparaciones de los distintos métodos de umbral, aunque la dificultad común a todos ellos ha sido la de encontrar una métrica adecuada para determinar la efectividad de cada uno. Sahoo *et al*. [13] consideró la uniformidad y la forma del umbral de los objetos para determinar la eficiencia de varios métodos y concluyó que el algoritmo de Otsu era mejor (aunque no se consideró el método de Tsai). Tsai comparó su algoritmo con el de Otsu y concluyó que ambos producían similares resultados, pero el método de Otsu funciona peor en presencia de ruido y sombras. Otra ventaja es que el método de Tsai es significativamente más rápido.

En nuestro caso, la aplicación del umbral directo sobre la imagen no permite una representación correcta de la roca por diferentes motivos como son las sombras, rocas que se superponen o se tocan entre ellas, que hacen que el proceso de identificación de las rocas sea poco fiable. Nuestro proyecto utiliza un algoritmo basado en el método de Tsai [10].

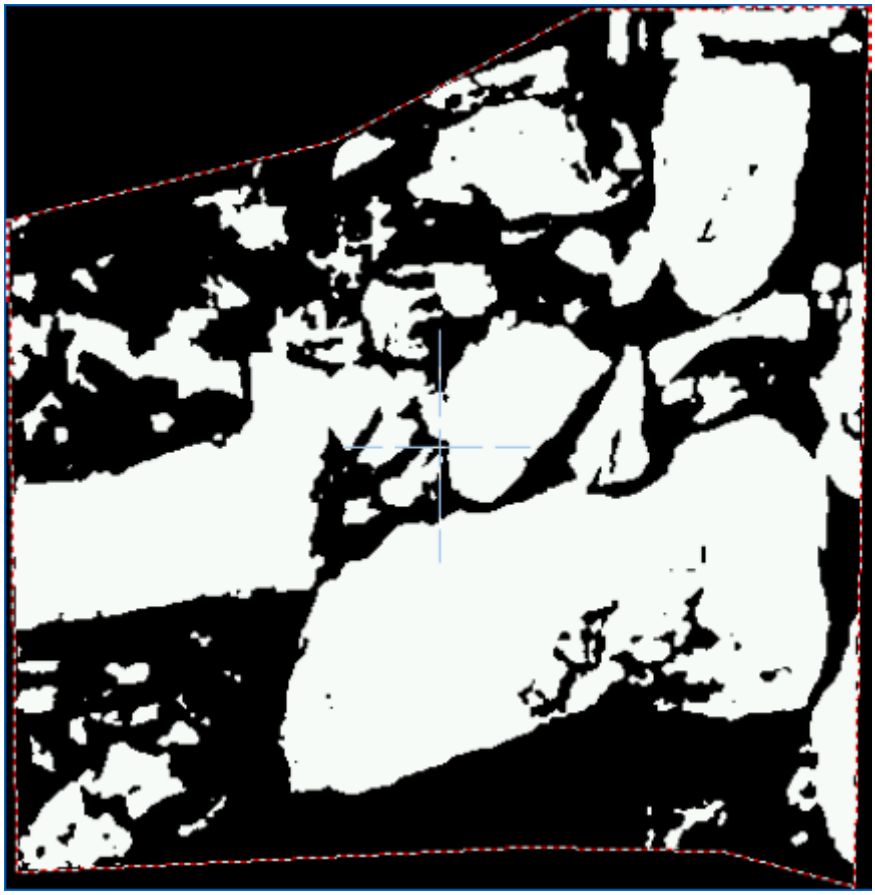

Figura 3.9. *Imagen tras aplicar el umbral*

Empleando el histograma de la región de interés, un primer umbral óptimo se calcula usando los tres primeros momentos, para después eliminar todos los niveles de gris por encima del umbral. Con el histograma resultante se obtiene un segundo umbral utilizando el mismo algoritmo. El umbral final es la media entre estos dos umbrales previos. Este algoritmo de dos niveles es capaz de tolerar el ruido aun presente en la imagen y funciona bien con imágenes de rocas mojadas. Cuenta con la ventaja adicional que el tiempo de cálculo es menor que otros métodos considerados.

Algoritmo de erosión

Para obtener una separación completa entre las zonas blancas de la imagen correspondientes a rocas es necesaria la aplicación sobre la última imagen obtenida de una serie de transformaciones adicionales.

Se implementa un algoritmo basado en la erosión [16]. La aplicación de este paso basado en la morfología ofrece mejores resultados en el aislamiento de rocas, eliminado pequeñas regiones blancas que no suponen ninguna pérdida ya que nuestro problema viene provocado por las rocas de gran tamaño con capacidad para bloquear la salida de la tolva y no por las rocas más pequeñas.

El algoritmo funciona de la siguiente manera: si 1 de los 8 vecinos de un pixel es negro, ese pixel será negro. Este proceso se aplica 8 veces para darnos regiones blancas más pequeñas y mejor separadas. Este número se establece de acuerdo con el tamaño de la imagen y considerando el mínimo tamaño que debe ser detectado. De esta manera se eliminan las pequeñas uniones entre regiones blancas. El proceso se finaliza con un algoritmo separador que borra las regiones blancas con pocos píxeles tanto en vertical como en horizontal.

Figura 3.10. *Imagen obtenida del proceso de erosión*

La figura anterior nos muestra la imagen obtenida al finalizar esta etapa.

La erosión se debe aplicar junto a un operador de dilatación que convierte en negro el contorno inicial de las rocas con un mínimo error. El proceso de dilatación se lleva a cabo durante la siguiente etapa por lo que se mantiene el tamaño inicial de las regiones.

3.4.3 Procesamiento de imágenes

Una vez que la imagen ha sido preprocesada, el sistema asume que cada región blanca es una roca por lo que cada posible roca aparece inscrita en un rectángulo cuyo tamaño lo proporciona el operador de dilatación. La figura 3.11 muestra dichos rectángulos superpuestos a la imagen original. Esta zona de la imagen se etiqueta como “posible roca” y sólo los rectángulos con tamaño mayor que un umbral son considerados por la red neuronal. El tamaño del rectángulo es una estimación del tamaño real de la roca. Como la distancia entre la cámara y el alimentador se conoce y el foco es constante, una simple transformación geométrica nos proporciona una estimación del tamaño real de la roca, basándonos en la longitud del rectángulo.

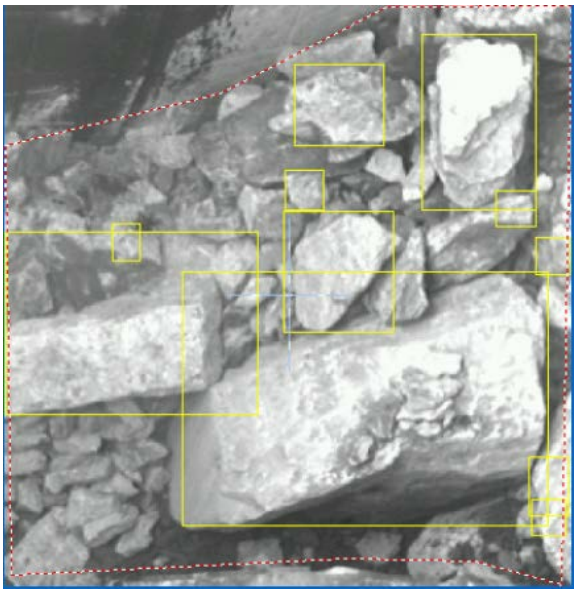

Figura 3.11. *Imagen con las rocas candidatas*

Las redes neuronales artificiales son una técnica de clasificación muy conocida y han mostrado su enorme potencial como herramientas de clasificación [17-19]. En nuestro caso, una red neuronal decide si una “posible roca” va a ser clasificada como roca o no: la entrada es una imagen de niveles de gris y la salida es su clasificación como roca o no. Las redes neuronales son muy útiles para detectar si dos o más rocas próximas forman una roca candidata.

La red neuronal sólo analiza las regiones que superan un umbral de tamaño determinado. Los resultados de la red para la imagen que se viene considerando como muestra son los que muestra la figura 3.12 que presenta las rocas identificadas por la red neuronal.

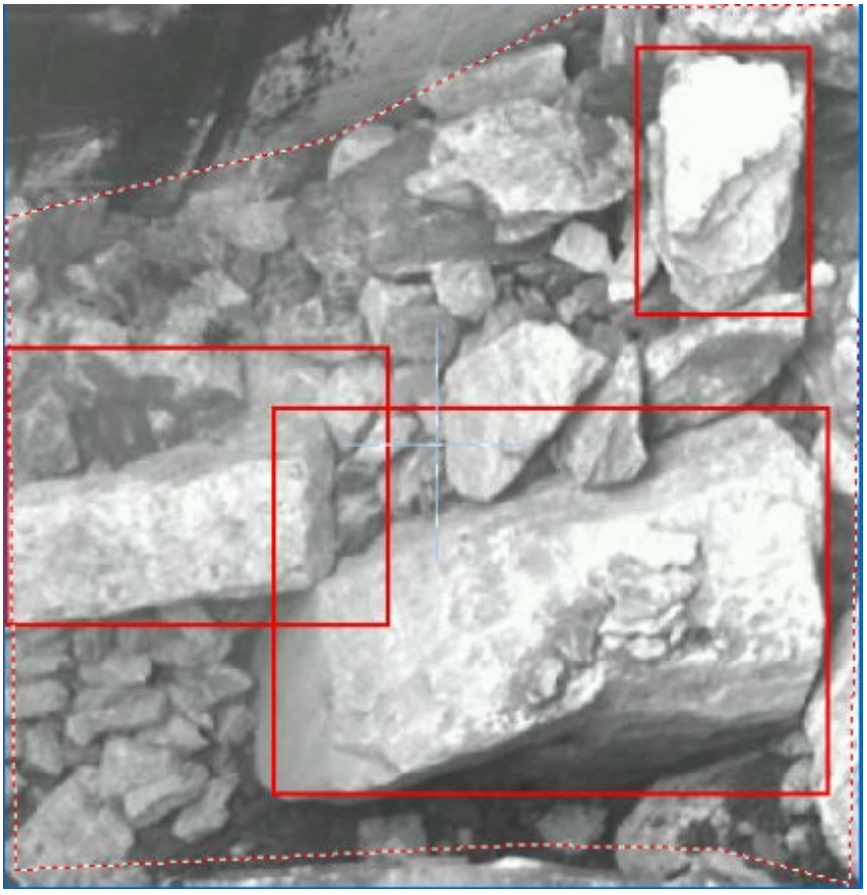

Figura 3.12. *Resultados de la red neuronal*

Para mejorar la precisión del reconocimiento se añadió un mecanismo de seguimiento: cuando una roca es detectada, se sigue a lo largo de una secuencia de 10 imágenes (aproximadamente equivale a 2.5 s.), y si se detecta en al menos la mitad de las imágenes, se considera una roca bloqueante. Para considerar que una roca grande es la misma en 2 imágenes el tamaño de la región debe ser similar y el desplazamiento debe ser acorde con la dirección de caída de la tolva. Este método elimina falsas detecciones debido al agrupamiento de rocas, ya que no persisten a lo largo de una secuencia temporal. Cuando se considera que una roca puede bloquear la machacadora, el operador es alertado y la línea puede ser detenida para la destrucción de la misma.

3.5. Resultados

En este trabajo se compararon dos tipos de redes neuronales: una red neuronal de tres capas alimentada hacia delante y entrenada con el algoritmo de retro-propagación (FANN) y una red neuronal de funciones de base radial (RBF). La topología utilizada fue la misma en los dos casos: una capa de entrada de 400 (20x20) neuronas, una capa oculta de 20 neuronas y una capa de salida con 2 neuronas. La figura 2.13 nos muestra la red utilizada.

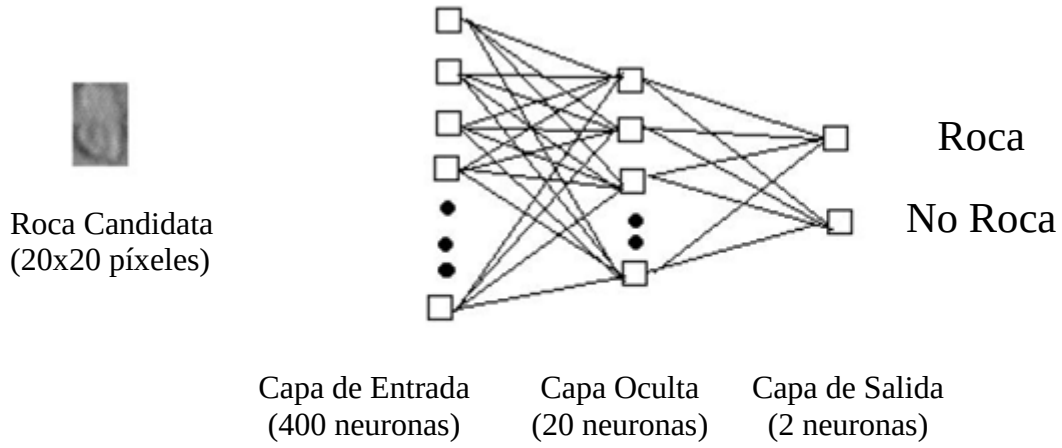

Figura 3.13. *Esquema de la red neuronal utilizada*

La entrada de la red neuronal es una “posible roca”, reducida a un tamaño de 20x20 píxeles y preprocesada. La capa oculta se fijó en 20 neuronas después de realizar varias pruebas. La activación de una de las neuronas en la salida significa que el patrón (la imagen) es una roca y la activación de la otra neurona de salida significa que no es una roca.

La bibliografía nos proporciona multitud de técnicas de reducción de la imagen de entrada [20-22]. Sin embargo ninguna de ellas resultó ser útil para este proyecto por el elevado tiempo de procesamiento que requerían y en nuestro caso el tiempo es un factor importante a tener en cuenta dado que ha de trabajar en tiempo real.

Para tomar la decisión final de las técnicas que se deben implementar en el sistema se tuvieron en cuenta los resultados de los distintos algoritmos propuestos. La tabla 1 muestra los resultados tras aplicar a la imagen una normalización de los niveles de gris, una etapa de ecualización previa a la normalización y un filtro pasabaja previo a la normalización.

	Imágenes Normalizadas	**Imágenes Ecualizadas**	**Imágenes Filtro Pasabaja**
FANN	93.2	86.6	93.2
RBF	83.3	80.0	96.6

Tabla 3.1. *Porcentaje de identificaciones correctas con el conjunto de prueba*

En un primer momento solo se disponía de 200 imágenes lo que hizo que se emplease una técnica de validación cruzada para entrenar la red neuronal. Se utilizó un conjunto formado por 170 imágenes para el entrenamiento y 30 para la evaluación del entrenamiento. La mitad de ellas correspondían con rocas correctas y la otra mitad con “no rocas”, polvo, varias rocas pequeñas juntas, partes de rocas…

En todos los casos las imágenes se escalaron a un tamaño 20x20 pixeles que suponían la entrada a la red neuronal, los mejores resultados se obtuvieron, como muestra la tabla, para las imágenes suavizadas con un filtro pasa baja.

Estas pruebas con las distintas técnicas de la etapa de preprocesamiento nos sirvieron para determinar cuál de ellas era la más adecuadas para aplicar a las imágenes para facilitar la posterior identificación de las rocas quedando patente que la solución más adecuada es la del empleo del filtro pasabaja.

Posteriormente se creó una base de datos de 1600 imágenes, la mitad de rocas y la otra mitad de no-rocas (polvo, trozos de roca, etc.). De ellas, 1360 se emplearon en el entrenamiento y 240 en la fase de prueba. La Tabla 2 muestra los resultados de las redes neuronales para este nuevo conjunto de imágenes y para imágenes reducidas a un tamaño de 20x20 pixeles y con filtro pasabaja.

	% identificación correcta
FANN	92.0
RBF	94.1

Tabla 3.2. *Porcentaje de éxito con el conjunto de prueba de imágenes extendido*

De las dos redes neuronales consideradas se optó por implantar en el sistema una red neuronal RBF debido a sus mejores resultados tanto en el modelo inicial como en el ampliado.

Inicialmente se realizaron pruebas in-situ para aumentar nuestro conocimiento sobre el funcionamiento del sistema en condiciones reales y para familiarizar al operario que se encargaría de manejarlo del final del sistema. Esto nos permitió estudiar el

comportamiento del sistema en condiciones reales y emplear esa información para intentar comprender por qué se daban situaciones de incorrectas detecciones.

La figura 3.14 muestra los resultados a lo largo de esos 9 meses, de enero a septiembre de 1999, trabajando de forma no supervisada en condiciones reales y con el programa definitivo. Donde “Detección correcta” significa que una roca que bloquearía el sistema ha sido detectada. “Detección incorrecta” representa el hecho de que se ha lanzado una alarma al operario por la presencia de una roca capaz de bloquear, pero que la roca que la ha provocado no bloquearía la tolva. Y “No detección” significa que hay una roca capaz de bloquear y el sistema no la ha detectado.

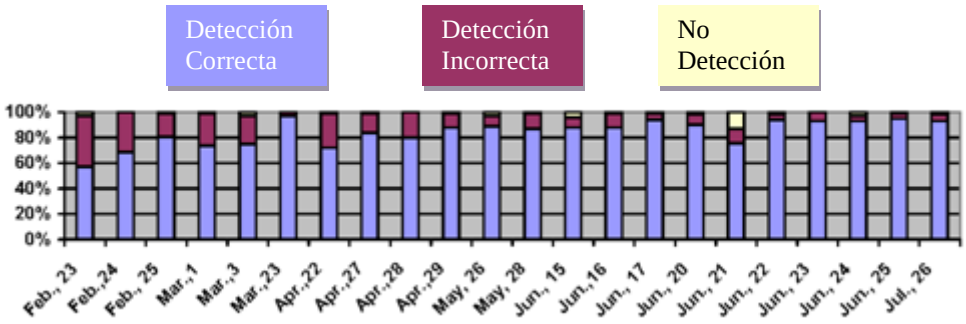

Figura 3.14. *Algunos resultados del funcionamiento del sistema.*

Cada columna vertical de la figura 3.14 corresponde a una sesión de trabajo del sistema en condiciones reales. Sólo se han mostrado aquellas sesiones con un número significativo de rocas. Las sesiones de trabajo son de 8 horas y el número de rocas que se analizan cada día. Hay que considerar también que los resultados son subjetivos ya que el operario encargado de supervisar el sistema etiqueta las situaciones como correcta/incorrecta/no detectada, pero no detiene el sistema para realizar la medición exacta de la roca. Dados los años de experiencia del operario podemos concluir que sus apreciaciones son correctas.

Dos cuestiones prioritarias del trabajo en la tolva bajo supervisión del sistema son evitar las situaciones de riesgo y que la cinta transportadora que alimenta la tolva sólo se detiene en caso de bloqueo.

La mina posee dos tolvas que están funcionando a la vez por lo que el único operario encargado de supervisarlas, en caso de atasco en una de ellas, se ve obligado a detener las dos hasta que la situación se ha resuelto, ya que su intervención le obliga a desatender la otra tolva. La mejora que supone nuestro sistema hace que la tolva pueda seguir funcionando con independencia de lo que suceda en la otra, al menos hasta la detección de otro positivo. Por lo que el sistema de visión supone una gran mejora en el funcionamiento del sistema de tolvas.

No es posible hacer una comparativa sobre si se realizan más o menos paradas o se obstaculiza la salida de la tolva más o menos veces con el sistema que sin él ya no existen datos registrados de funcionamiento previo al sistema informático. Sin embargo la compañía ha estimado que el número de paradas se ha reducido a la mitad.

Los resultados muestran una tasa de reconocimientos correctos superior al 70% que era la solicitada por ENUSA para considerar que la inversión era rentable. El sistema alcanza una velocidad de procesamiento de 8 imágenes/s, 4 imágenes/s para cada tolva, suficientes para controlarlas de forma correcta.

3.6. Conclusiones

La industria minera trabaja en unas condiciones de entorno muy hostiles tanto para humanos como para ordenadores. Los sistemas informáticos pueden ayudar a obtener mejores resultados y evitar situaciones de riesgo para los trabajadores. Ambos objetivos se han logrado con nuestro sistema.

El sistema de visión que se ha presentado a lo largo de este capítulo realiza la tarea de estimar el tamaño de las rocas para evitar bloqueos homogéneamente y de forma reiterada. Se han obtenido resultados significativos, en primer lugar se evitan situaciones de riesgo para el trabajador y colabora con el empleado facilitando su trabajo.

El sistema de visión desarrollado para entornos mineros debe ser robusto y se debe poner especial atención en los algoritmos que se emplean.

Se han propuesto diferentes técnicas en la literatura para abordar el problema de reconocimiento de rocas. Este trabajo presenta algunos resultados obtenidos utilizando una mezcla de algoritmos de preprocesamiento de imágenes y de redes neuronales. En la fase de preprocesamiento se han testado diferentes algoritmos y el resultado final elaborado es una sucesión de operadores, alguno de los cuales ha sido adaptado a nuestro problema particular. La salida de la fase de preprocesado es un conjunto de ventanas en la imagen que pueden contener una roca. A continuación, una red neuronal decide si una roca está presente o no. También se han considerado mecanismos para evitar falsas alarmas debidas a agrupaciones de rocas. Los requisitos de tiempo se han alcanzado y el sistema es capaz de supervisar dos líneas en tiempo real.

Nuestro sistema se ha desarrollado de cara a operar en una situación concreta con las restricciones que eso supone, pero creemos que con pequeñas variaciones podría ser utilizado en otros contextos similares de mecanismos que alimentan cintas transportadoras.

Una vez que el sistema se puso en funcionamiento de forma regular, el trabajador se siente ayudado por el mismo. Sólo se debe detener una línea cuando se produce una incidencia, lo que supone una gran mejora con respecto a la forma de trabajar previa al sistema de visión. En la pantalla del ordenador aparecen los dos

alimentadores y precribadores en tiempo real, y ambas líneas pueden ser supervisadas con una sola ojeada.

Las redes neuronales y los métodos de procesamiento de imágenes, trabajando juntos, mejoran el rendimiento del sistema y el número de falsas alarmas se mantiene bajo. Los problemas derivados de variaciones de luz y polvo en suspensión que se producen en el sistema se han tenido en cuenta para construir un algoritmo robusto. El sistema es fácil de operar y trabaja en condiciones reales por lo que los objetivos que se marcaron al inicio han sido cumplidos.

Para finalizar, comentar que los resultados obtenidos con este trabajo han dado lugar a una serie de artículos, capítulos de libros y ponencias a congresos que se citan a continuación.

Capítulos de libros y revistas:

Enrique Cabello; M. Araceli Sánchez; Javier Delgado. “A New Approach to Identify Big Rocks with Applications to the Mining Industry”. Real Time Imaging.8, pp. 1 - 9.02/2002.

Tipo de producción: Artículo

Tipo de soporte: Revista

Enrique Cabello; M. Araceli Sánchez; Julian Nieto; Jesús M. Berrocal; Guido Castro and Javier Delgado. “A Computer Vision Application in Real-Time to Identifying Big Rocks with Applications to the Minning Industry”. pp. 1 - 12.04/2000.

Tipo de producción: Artículo

Tipo de soporte: Revista

Enrique Cabello; M. Araceli Sánchez; Julian Nieto; Jesús M. Berrocal; Guido Castro and Javier Delgado. “A real-time vision system for on-line rock size estimation”. pp. 86 - 91. International Institute for Advanced Studies,1999.

Tipo de producción: Capítulos de libros

Tipo de soporte: Libro

A. Sánchez; J. Nieto; J. M. Berrocal; G. Castro; E. Cabello; J. Delgado. “Una aplicación de visión artificial para medir el tamaño de rocas en tiempo real”. Canteras y Explotaciones. 380, pp. 52 - 57.1999.

Tipo de producción: Artículo

Tipo de soporte: Revista

J. Nieto; J. M. Berrocal; A. Sánchez; G. Castro; and Enrique Cabello.A real time computer vision system for rock size estimation”. La Lettre de l’IA.134 - 135-136, pp. 282 - 284.EC2, 1998.

Tipo de producción: Artículo

Tipo de soporte: Revista

Congresos:

Título: Un sistema de visión para detectar y estimar el tamaño de rocas.

Nombre del congreso: XX Jornadas de Automática.

Ciudad de realización: Salamanca, España

Fecha de realización: 09/1999

M. Araceli Sánchez; Julian Nieto; Jesús M. Berrocal; Guido Castro; Enrique Cabello; Javier Delgado.09/1999.

Título: A real-time vision system for on-line rock size estimation.

Nombre del congreso: Special session on Advanced Concepts for Intelligent Vision Systems. (XI International Conference on Systems Research, Informatics and Cybernetics).

Ciudad de realización: Baden-Baden, Alemania

Fecha de realización: 08/1999

Enrique Cabello; M. Araceli Sánchez; Julián Nieto; Jesús M. Berrocal; Guido Castro and Javier Delgado.08/1999.

Título: A real time computer vision system for rock size estimation.

Nombre del congreso: EC2 and Developpement.

Ciudad de realización: Nimes, Francia

Fecha de realización: 05/1998

J. Nieto; J. M. Berrocal; A. Sánchez; Guido Castro; Enrique Cabello; Francia.05/1998.

CAPITULO 4

ENFOQUE DE UNA RED NEURONAL CONVOLUCIONAL PARA LA DETECCIÓN DE ATAQUES DE PRESENTACIÓN FACIAL MULTIESPECTRAL EN SISTEMAS AUTOMATIZADOS DE CONTROL DE FRONTERAS

4.1. Introducción

Los sistemas automatizados de control fronterizo son una infraestructura crítica al cruzar la frontera de un país. Que personas no autorizadas atraviesen una frontera es un riesgo de seguridad para cualquier país. Este capítulo presenta un análisis multiespectral de la detección del ataque de la presentación para la biometría facial usando las características aprendidas de una red de neuronas convolucional. Se consideran tres sensores para diseñar y desarrollar una nueva base de datos que se compone de imágenes visibles (VIS), infrarrojo cercano (NIR) y térmicas. La mayoría de los estudios se basan en ambientes de laboratorio o ideales con condiciones controladas. Sin embargo, en un escenario real, la situación de un sujeto se modifica completamente debido a diversas condiciones fisiológicas, como el estrés, los cambios de temperatura, la sudoración y el aumento de la presión arterial. Por esta razón, el valor añadido de este estudio es que esta base de datos fue adquirida in situ. Los ataques considerados fueron imágenes impresas, enmascaradas y exhibidas. Además, se utilizaron cinco clasificadores para detectar el ataque de presentación. Los resultados presentan mejores salidas cuando todos los sensores se utilizan juntos, independientemente de si se considera la fusión a nivel de clasificador o de nivel de característica. Por último, los clasificadores como KNN o SVM muestran un alto rendimiento y un bajo nivel de cálculo.

Una aplicación actual de los sistemas de verificación facial en entornos reales con elevadas limitaciones de seguridad son los sistemas ABC (Automated Border Control) [1, 2]. Estos se utilizan para verificar las identidades de los pasajeros automáticamente a través de la biometría [3, 4]. Gracias a esta tecnología, los sistemas ABC ayudan a los guardias de fronteras a llevar a cabo tareas rutinarias como el control de pasaportes y la verificación de rostros en los pasos fronterizos internacionales (BCP) de varios países en tiempos más cortos y con resultados estándar y homogéneos. Además, los retrasos en el tiempo de cola se mejoran utilizando estos sistemas, que se dedican principalmente a acelerar los cruces fronterizos para los viajeros "de buena fe" sin perturbar la comodidad de los mismos [5]. Una de las tareas más relevantes que realizan los sistemas ABC es la identificación biométrica del viajero. Esta función se realiza a menudo comparando la cara del sujeto capturado (denotada como una imagen in situ) en una posición estática frontal con la imagen de la cara almacenada en el chip del pasaporte (también conocido como eMRTD). En esta configuración, el sujeto puede permanecer delante de la cámara.

En Europa, la agencia Frontex ha publicado varias directrices técnicas y recomendaciones para los procedimientos de verificación facial ABC [6, 7]. Estas directrices están estrechamente relacionadas con los procedimientos de otros organismos internacionales de fronteras. Los grupos de normalización, tanto europeos como internacionales, mantienen niveles altamente armonizados y definen el procedimiento de verificación facial basado en la coincidencia in situ de la imagen con la imagen eMRTD.

Actualmente, la precisión de los sistemas de verificación facial ha logrado resultados notables, pero otros aspectos, como la robustez frente a los ataques, rara vez se consideran o se tratan como relevantes. Un ABC no es un sistema aislado y, por lo general, un guardia de fronteras supervisará entre seis y doce sistemas. En el caso de los sistemas ABC, se deben tener en cuenta los aspectos de seguridad y las vulnerabilidades del sistema. Por lo tanto, es obligatorio eliminarlos o al menos minimizarlos. En este sentido, los ataques de presentación facial son un tema cuya relevancia es cada vez mayor en el área de la biometría facial, y el número de referencias dedicadas a la detección de ataques faciales ha crecido en los últimos años. Este interés comienza en

laboratorios de investigación en los que las condiciones ambientales son diferentes a los escenarios reales, y todas las referencias encontradas se prueban en estas situaciones. El enfoque actual se basa en proyectos [8, 9] en los que se logran situaciones cercanas a la realidad. Probar los sistemas en un escenario fronterizo real es una situación más compleja que la del laboratorio. Teniendo en cuenta que los pasajeros llegan en condiciones de estrés, pueden cambiar de diferentes temperaturas ambientales (exterior e interior).

Los ataques al sistema, a nivel de sensor, se etiquetan comúnmente como ataques de presentación [10,11], y la prevención o detección de esos ataques se la denomina PAD [12]. Las capacidades PAD son relevantes para el diseño de sistemas ABC. La verificación de rostros se realiza en un escenario casi sin supervisión. Los sistemas BCP son entornos de alta seguridad en los que los guardias fronterizos hacen grandes esfuerzos para garantizar que sólo los titulares de pasaportes cruzan la frontera. La suplantación de identidad es utilizada en el cruce de líneas por organizaciones criminales. Por lo tanto, los sistemas automáticos deben estar listos para procesar, no sólo las identidades de los viajeros de buena fe, sino también para detectar ataques de presentación que se produzcan a los sistemas ABC.

En cuanto a las recomendaciones de Frontex, el tiempo máximo que un pasajero debe pasar en el proceso de cruce de fronteras no debe superar los 12 s [13]. En ese tiempo, el sistema ABC debe realizar tres tareas específicas principales. En primer lugar, se lee y valida la identificación del pasajero (normalmente se trata de un pasaporte o documento oficial nacional) con datos personales almacenados en un chip. En segundo lugar, se verifica si el viajero es el titular del documento presentado. Para llevar a cabo esta acción, es obligatorio tomar y analizar la imagen facial del pasajero en tiempo real. A continuación, esta imagen se compara con la imagen almacenada en el chip de documento. En tercer lugar, se envía una consulta a las bases de datos de control de fronteras para comprobar si el viajero puede o no puede entrar en el país. En un ABC real, el momento que se dispone para detectar la existencia de un ataque de presentación es mientras el viajero está esperando la respuesta en la tercera tarea.

Los métodos presentados en este estudio fueron cuidadosamente seleccionados para adaptarse al tiempo operativo disponible y a la capacidad computacional habitual de un sistema ABC. Además, la selección de los sensores fue un proceso en el que hubo que valorar diferentes aspectos, porque es crucial tener en cuenta los espectros de los sensores, que deben ser lo más complementarios posible.

Por esta razón, se consideraron tres tipos de sensores: cámara visible, infrarrojo cercano (NIR) y térmico. Cada sensor reacciona a diferentes espectros electromagnéticos: la longitud de onda visible se encuentra entre 400 y 700 nm, la longitud de onda NIR empleada es de 800 nm y la longitud de onda térmica se encuentra entre 3 y 1 μm.

Se deben considerar además otros factores, por ejemplo, el uso de sensores comerciales, que son fáciles de reemplazar, y el tiempo de adquisición de la imagen por el sensor, para no incrementar las esperas de los pasajeros. En el actual estudio, la hipótesis que se hizo fue que era posible utilizar varios sensores para detectar a un atacante plausible. La fusión de estos sensores proporciona un alto nivel de seguridad porque el delincuente tiene que sortear, simultáneamente, varios sensores con espectros de diferente rango. Por lo tanto, el ataque es casi imposible de llevar a cabo.

De acuerdo con los conceptos biométricos faciales, los ataques a sensores se producen habitualmente con herramientas o artefactos para hacerse pasar por otras personas, utilizando máscaras o imágenes impresas colocadas frente al sensor de la cámara. En este sentido, el sensor capturaría el artefacto en lugar de la cara de la persona. La ISO [12] define como instrumento de ataque de presentación (PAI) cualquier característica u objeto biométrico que se utilice en un ataque de presentación. Cabe señalar que el equipo necesario para atacar los sistemas debe ser de bajo precio y fácil de adquirir [14,15].

La relevancia de estos ataques se puede explicar en dos niveles: seguridad y costes de reputación. Los riesgos de seguridad se deben a delincuentes que no son

detectados cruzando la frontera y accediendo a un país sin problemas para delinquir. Mientras que los costes de reputación se producen con la posible difusión de vídeos de pasos fronterizos en los que delincuentes presentan con éxito diferentes artefactos de bajo coste en redes sociales y deep web, difundiendo información maliciosa en Internet.

En escenarios reales, los ataques a sistemas ABC se centran en el procesado de la imagen. En el cruce de la frontera, un sujeto se detiene frente al centinela o eGate sin moverse. La imagen se toma y se procesa mientras el sistema accede a las bases de datos oficiales para verificar la identidad del sujeto. Después de eso, el pasajero puede acceder o no al país.

Actualmente, se dispone comercialmente de otros tipos de sensores que trabajan en un rango de espectro diferente al visible y que pueden adquirir información multidimensional. La metodología presentada en este estudio utiliza los distintos sensores de imagen mencionados anteriormente (cámara de luz visible, sensores térmicos y de infrarrojo cercano) para realizar la PAD. En un sistema ABC, las imágenes de luz visible (VIS) se obtienen para facilitar la verificación facial frente a la imagen contenida en el eMRTD. Por lo tanto, las imágenes visibles se complementarán con otros sistemas de adquisición, típicamente infrarrojo cercano (NIR) e imágenes térmicas. El objetivo principal de este trabajo de investigación es el uso de sensores aislados o mixtos, considerando los costos y las restricciones computacionales. Los ataques considerados se realizan utilizando los siguientes artefactos: una fotografía impresa, una máscara, una máscara impresa con los ojos recortados, o una imagen mostrada en un dispositivo electrónico.

Se han propuesto varios métodos para detectar ataques de presentación a partir de imágenes por diferentes grupos de investigación. Básicamente, los enfoques se pueden agrupar en in situ, en movimiento, basados en texturas y en contexto, y basándose en inspecciones visuales [10,11].

Hay que tener en cuenta que las principales características seleccionadas están relacionadas con la textura y el análisis de color siguiendo un clasificador que puede separar imágenes de los usuarios auténticas de los ataques (clasificación dicotómica). Las tendencias más destacables en los últimos años son el uso de deep learning y redes neuronales profundas en tareas de procesamiento y clasificación. Las soluciones logradas en este estudio siguen un enfoque novedoso para los sistemas PAD, utilizando una red neuronal profunda, más específicamente, redes neuronales convolucionales (CNN) [16].

4.2. Estado del arte

La inteligencia artificial (IA) se ha vuelto importante en los últimos años debido a sus vastas aplicaciones en el mundo real [17]. Esta sección se centra en el uso de deep learning en PAD facial mediante imágenes multiespectrales. Un estudio exhaustivo de los últimos avances de la técnica general en el área de PAD está fuera del alcance de este documento, pero algunos ejemplos de la investigación actual incluyen huellas dactilares PAD [18,19], maquillaje PAD [20], morphing y demorphing PAD [21, 22], etc. Es de destacar que el principal problema de los diferentes enfoques presentados es que las bases de datos consideradas fueron adquiridas en condiciones de laboratorio [23]. Sin embargo, las situaciones reales dependen del estado de estrés, la situación fisiológica y las emociones del posible pasajero, que cambian los patrones faciales (por ejemplo, imágenes térmicas). Además, es lógico pensar que la detección de atacantes es más sencilla cuando se combinan varios sistemas. En consecuencia, la debilidad de cada sistema de adquisición se ve atenuada por la existencia de otros sistemas complementarios [24].

Kotwal et al. [25] investigaron el uso de datos multiespectrales (imágenes en color, imágenes en el infrarrojo cercano (NIR) e imágenes térmicas) para la PAD facial, específicamente contra los ataques personalizados de máscaras de silicona. Se emplearon veintiuna máscaras hechas a medida para establecer la eficiencia de varios métodos face-PAD de uso común, en los diferentes canales de imágenes, utilizando un nuevo conjunto de datos (XCSMAD).

George et al. [26] propusieron un enfoque multicanal basado en una red neuronal convolucional para uso en un PAD. También presentaron una nueva base de datos Wide Multi-Channel presentation Attack (WMCA) para face-PAD que contenía una amplia variedad de ataques 2D y 3D para suplantación y ofuscación. Los datos de los diferentes canales así como la profundidad, la temperatura, el color y el infrarrojo cercano estaban disponibles para avanzar en el estudio en la PAD facial. El método propuesto se comparó con los enfoques basados en caracterización y se encontró que los superaba, logrando una tasa de error de clasificación de presentación de ataques (APCER) del 0,3% en el conjunto de datos introducido.

En [27], los autores propusieron un detector inteligente de vivacidad facial para evitar que el sistema biométrico sea "engañado" por el video o la imagen de un usuario válido que el falsificador pudiese haber adquirido con un dispositivo de mano de alta definición (por ejemplo, tableta de alta gama).

Albakri y Alghowinem [28] evaluaron la detección de vivacidad para sugerir soluciones que explican las debilidades encontradas en la detección de ataques de suplantación de identidad. Realizaron un estudio inicial para evaluar la detección de vida de las cámaras 3D en tres dispositivos, donde los resultados sugieren que una mayor flexibilidad garantiza lograr una mejor tasa en la detección de ataques de suplantación de identidad.

En [29], los autores presentaron subsistemas visibles y NIR que fueron atacados de dos maneras diferentes. El primer ataque consistió en ataques con imágenes individuales visibles y NIR (una por una). El segundo ataque se compuso de un sistema multiespectral (VIS y NIR) que fue atacado por pares de imágenes visibles y NIR. En este tipo de ataques, se analiza el color de las imágenes visibles para detectar un ataque plausible, mientras que la textura se comprueba para detectar el ataque en el caso de las imágenes NIR. Un sistema multiespectral realiza en este caso un proceso de verificación de dos pasos en el que se realiza un análisis de color. La tasa de éxito obtenida fue muy alta (100% en los mismos casos), pero el tipo de ataque probado fue muy simple, considerando sólo imágenes estáticas.

Zhang et al. [30] propusieron un método para la PAD basado en evidencia de viveza con dispositivos de adquisición específicos. Los autores utilizaron longitudes de onda específicas para fotografiar y almacenar a los usuarios de buena fe y a los delincuentes. Se construyeron y utilizaron dispositivos específicos para adquirir información espectral. La reflectancia obtenida se utilizó para entrenar un clasificador SVM. Además, los autores lograron una precisión del cien por cien cuando se usaron videos de ataques, pero una precisión del 92% en las fotos de ataques en comparación con los usuarios de buena fe, empleando una imagen plana. Cuando compararon a los individuos de buena fe con ataques de la cara con una máscara, se obtuvo una exactitud del 89%.Por lo que los autores afirmaron que su método depende del material que cubre los rostros.

En [31], el enfoque de los autores consistió en distinguir si el objeto analizado era una persona real (denotada por el concepto de viveza). Una persona real presenta características fisiológicas como la sudoración, la presión arterial y la temperatura. Sin embargo, en un humano artificial, como puede ser un maniquí, este tipo de características no existen. Por esta razón, la propuesta se basó en el gradiente de las imágenes capturadas por un sistema multiespectral con diferentes longitudes de onda. El enfoque se probó en una base de datos que incluía fotos planas bidimensionales, maniquíes 3D y máscaras. En cuanto a la tarea de clasificación, se utilizó el clasificador SVM para aprender las características basadas en gradientes de caras genuinas y falsas. Los autores reportaron una tasa de verdaderos positivos (TPR) del 98,3% y una tasa de verdaderos negativos (TNR) del 98,7%.

Actualmente, las CNN se utilizan ampliamente en tareas de reconocimiento y clasificación, obteniendo resultados satisfactorios [32- 34].Una estructura canónica de CNN para la detección de falsificaciones faciales fue implementada por Yang et al. [35]. Los autores utilizaron videos grabados por una cámara de luz visible. El CNN adoptado fue el conocido AlexNet (que ganó ImageNet 2012) [36]. Los autores utilizaron las bases de datos CASIA (formada por 55 usuarios) y REPLAY-ATTACK (50 sujetos). La cara fue detectada en cada muestra y recortada con diferentes proporciones de escala. Los autores afirmaron que, a pesar de no seleccionar cuidadosamente los parámetros de la CNN, los resultados obtenidos fueron exitosos

(tasa de error total a la mitad (HTER) inferior al 5%). Sin embargo, los resultados entre pruebas no fueron tan satisfactorios.

Una modificación más compleja es una red neuronal recurrente (RNN) con grandes unidades de memoria a corto plazo (LSTM) previas a una arquitectura CNN (basada en AlexNet) [37]. Los autores informaron un HTER del 5,9% y una tasa de error igual (EER) del 5,17% con la base de datos CASIA (50 sujetos: 20 para entrenamiento y 30 para pruebas).

En [16], los autores implementaron un CNN + RNN para detectar ataques de presentación basados en mapas de profundidad de imágenes de luz visible.Los autores construyeron una base de datos, llamada “Spoof in the Wild”, que está compuesta por 165 sujetos (incluyendo videos de buena fe y de ataques). Los ataques realizados fueron imágenes impresas de diferentes calidades y vídeos exhibidos en diferentes dispositivos. La CNN se utilizó para obtener mapas de profundidad de los fotogramas individuales, y la RNN evaluó las características temporales.

Una nueva arquitectura fue desarrollada por Lucena et al. [38], que se denominó FASnet y se basó en la arquitectura VGG-16 [39]. La red neuronal se entrenó previamente con la base de datos ImageNet. Los autores utilizaron enfoque de transferencia de aprendizaje de transferencia, más específicamente, el ajuste fino, y usaron una CNN para abordar los métodos de detección de falsificaciones faciales. Los autores obtuvieron una precisión del 100% y un HTER del 0% utilizando la base de datos 3DMAD. Sin embargo, los resultados disminuyeron al 99.04% de precisión y 1.20% HTER usando un REPLAY-ATTACK.

Resulta complicado establecer una comparación entre los distintos resultados presentados hasta ahora, ya que las variaciones de las distintas métodologías son muy grandes. Este inconveniente ya ha sido puesto de manifiesto previamente por Ramachandra y Busch [40].

4.3. Base de datos

Se ha desarrollado una nueva base de datos denominada "FRAV-Attack" para escenarios próximos a la realidad con el objeto de evaluar los sistemas que deben evitar ataques de presentación. El propósito principal de esta base de datos es cubrir los ataques de presentación de pasos fronterizos con diferentes representaciones faciales proporcionadas por varios sensores que se pueden colocar en una eGate.

En cuanto a los datos considerados, las bases de datos de contra falsificaciones faciales de referencia se realizan con frecuencia utilizando imágenes de luz visible. Por lo tanto, las bases de datos de ataque de presentación facial gratuitas disponibles (con fines de investigación) contienen principalmente imágenes RGB. En los últimos años, se han incorporado más sensores en las bases de datos, pero el proceso de adquisición de las mismas siempre se ha realizado en condiciones de laboratorio como puede ser el caso de las bases de datos CASIA [41], REPLAY-ATTACK [42] y MFSD-MSU [14].

El proceso de adquisición se llevó a cabo durante varios días, y los usuarios fueron voluntarios que firmaron un formulario de consentimiento informado. Además, la base de datos sigue los estándares acordados con la normativa RPGD de la Agencia Española perteneciente a la Unión Europea. La base de datos está compuesta por 185 usuarios. Donde el 62% de los usuarios son hombres y 38% son mujeres. En general, el 75% tiene entre 18 y 40 años y el resto son mayores de 40 años. Las muestras se adquirieron bajo iluminación uniforme y controlada en un escenario de frontera [8,9] similares a la situación real.

Se utilizaron los siguientes sensores con cada usuario: una cámara de luz visible SONY ILCE-6000Y, una cámara de teléfono móvil con un sensor térmico FLIR ONE y una cámara de vigilancia HIKVISION. Cada cámara tomó varias muestras (tanto de usuarios de buena fe como de delincuentes).

Los sensores tienen las siguientes características:

- El sensor térmico tiene una resolución de 160 × 120 píxeles, realiza mediciones de temperatura de hasta 120 C, y puede detectar diferencias de temperatura de hasta 100 mK.
- El sensor NIR tiene un rango de funcionamiento (mín.-máx.) de 0,5–3,5 m y una resolución de 480 × 360 a 60 FPS.
- La cámara visible es una cámara USB estándar con características comunes.
-

Se seleccionan cuatro tipos de artefactos para atacar los sistemas ABC. Los ataques se seleccionaron teniendo en cuenta que podrían realizarse en una situación operativa real para un paso fronterizo en la cola del aeropuerto. Por esta razón, los artefactos que podían ser fácilmente detectados por los guardias fronterizos de un vistazo fueron descartados. Todos los artefactos seleccionados podrían ser puestos y retirados sin esfuerzo porque un ataque debe llevarse a cabo en un tiempo mínimo. Además, estos artefactos debían poder ocultarse rápidamente en el equipaje. El uso de máscaras rígidas 3D podría ser un contraejemplo notable ya que estos artefactos son más grandes que una cara normal y fácilmente detectables, incluso si el color de la máscara intenta imitar la piel humana o el criminal usa una bufanda, una gorra o gafas de sol. Con el añadido de que una máscara 3D es difícil de ocultar en el equipaje y colocarla posición correcta es más complejo y requiere más tiempo que en el caso de los artefactos seleccionados. Las máscaras flexibles 3D solucionan algunos de los inconvenientes anteriores, pero, de nuevo, la colocación adecuada de la máscara requiere mucho tiempo. De hecho, sólo se ha detectado un incidente real de este tipo, y el criminal fue arrestado por este tipo de ataque (vuelo de Hong Kong a Vancouver el 29 de octubre de 2010). Cabe señalar que el delincuente no pudo llegar a la puerta del ABC.

De esta manera, los artefactos que se utilizan comúnmente en ataques de presentación facial según la bibliografía [10,40] son:

- Ataques fotográficos impresos con una alta resolución
- Fotomáscara impresa
- Fotomáscara impresa sin ojos (simulando a un ser humano real realizando parpadeo ocular)
- Una imagen de alta resolución que se muestra en una tableta

El presente trabajo de investigación se pretende centrar en los resultados obtenidos en situaciones que puedan ser similares a la realidad. Por ello, se desarrollaron experimentos en dos localidades (aeropuerto Adolfo Suárez-Barajas de Madrid y puerto de Algeciras en Cádiz) en pasos fronterizos reales y entornos operativos. Como es lógico, fue necesaria (a la vez que obligatoria) una estrecha interacción con la guardia de fronteras para llevar a cabo este estudio.

Todas las imágenes de la base de datos siguen el estándar ICAO Doc 9303 [43] en el que se recomienda que la distancia del sensor de la cámara (CSD) este comprendida entre los siguientes límites: 1m<=CSD<=2,5m. En la figura 4.1 se representan para dos ejemplos todos los diferentes tipos de adquisiciones, las visibles, las de infrarrojo cercano y las térmicas, así como los diferentes ataques que se realizaron para los mismos.

La cara ha sido detectada y recortada en las muestras porque las bases de datos están compuestas principalmente por imágenes faciales, similares a estudios anteriores [35, 37]. Los tres subconjuntos particulares (VIS, NIR y térmico) se adquirieron del mismo eGate pero con diferentes sensores. Las imágenes se procesaron de forma independiente y se redimensionaron a 128 × 128 píxeles.

Figura 4.1. *Dos ejemplos de usuario genuino o de buena fe y sus correspondientes ataques.*

El esquema de la Figura 4.2, muestra como se adquiere la imagen del individuo. En primer lugar, la Figura 4.2a representa a un sujeto que se acerca a los sensores. A continuación, la figura 4.2b describe la situación del sistema en el paso fronterizo. El pasajero cruza el sistema, se toman y procesan imágenes, detectando a los delincuentes in situ. Por último, la figura 4.2c ilustra el sistema que obtiene las imágenes durante la aproximación de los pasajeros a la frontera.

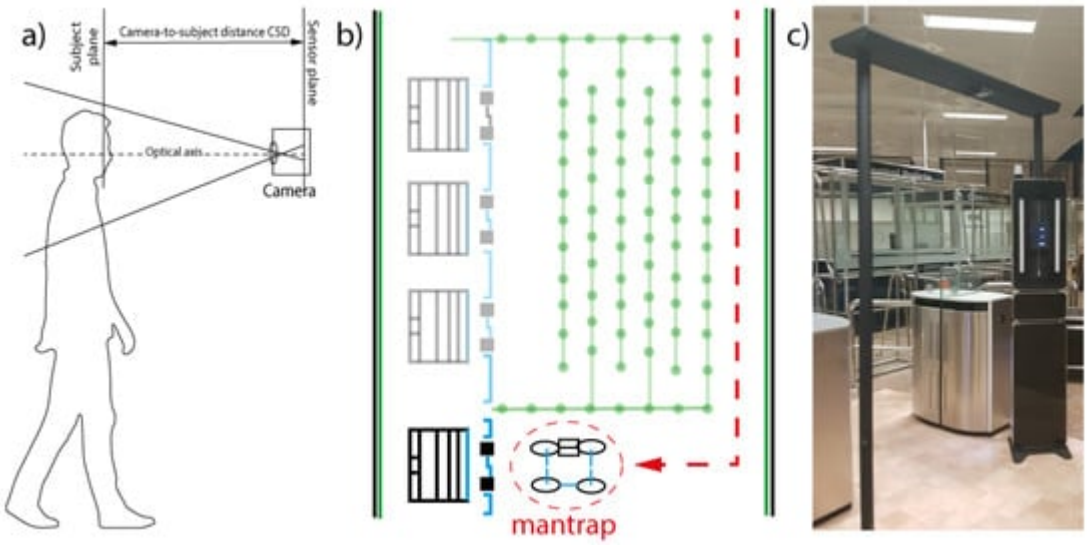

Figura 4.2. *a) El régimen de adquisiciones;* ***b)*** *representación del paso fronterizo con la puerta dedicada; y* ***c)*** *un sistema ABC real.*

4.4. Método y descripción experimental

4.4.1. Método

Los experimentos desarrollados se basaron en una red neuronal convolucional cuyos resultados fueron procesados posteriormente por diferentes clasificadores. Una vez que la CNN ha terminado, se obtienen las características de las imágenes más adecuadas y, a continuación, se devuelve un vector de características. Como entrada a la CNN, se utilizan tres subconjuntos de la base de datos (descritos en la sección anterior) siguiendo diferentes enfoques para la fusión de la información:

(1) visible (VIS), NIR y térmica proporcionadas por separado;

(2) VIS, NIR y térmico mezclados y añadidos a la red neuronal;

(3) los tres subconjuntos mezclados y agregados a un único clasificador.

Como se muestra en la Figura 4.3, la tarea de clasificación la realizan cinco clasificadores diferentes. Los clasificadores que fueron seleccionados son los siguientes:

- máquina de vectores de soporte (SVM) en dos versiones, una con función de base radial (RBF) y una segunda versión del mismo con kernel lineal.
- k-vecino más cercano (KNN).
- árbol de decisión.
- regresión logística.

Cada clasificador se entrenó de forma independiente.

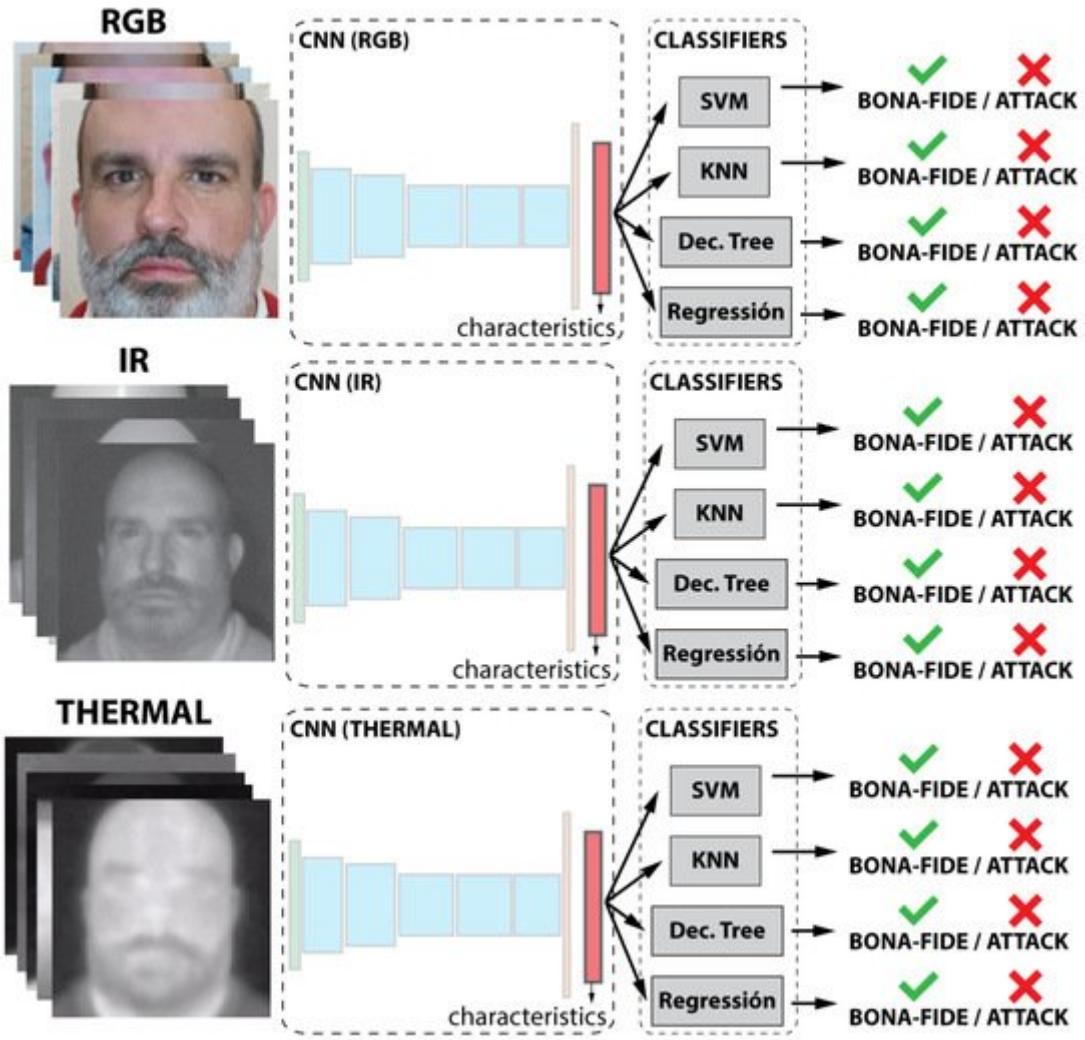

Figura 4.3. *Representación del primer caso de estudio.*

4.4.1.1. Arquitectura CNN

La arquitectura de la red neuronal convolucional se basa en la de AlexNet [36] con algunas adaptaciones al problema actual cuyo objeto es intentar mejorar el rendimiento. Las capas consideradas son las siguientes:

1. Capa convolucional (11 × 11) + capa MaxPool (2 × 2) + Capa de normalización
2. Capa convolucional (4 × 4)
3. Capa convolucional (3 × 3)
4. Capa convolucional (3 × 3)
5. Capa convolucional (3 × 3) + capa MaxPool (2 × 2)
6. Capa de abandono
7. Capa de abandono
8. Capa totalmente conectada

El número de núcleos utilizados para cada capa convolucional es 56 para la primera capa, 156 para la segunda, 256 para la tercera, 254 para la cuarta y 106 para la última capa convolucional. En la capa de abandono, se utilizaron 2512 neuronas y 500 estaban completamente conectadas, figura 4.4.

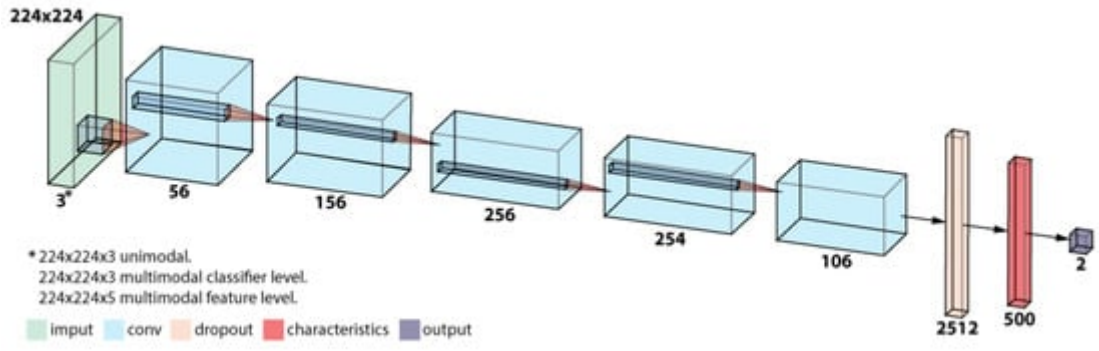

Figura 4.4. *Arquitectura de red.*

Se utilizó una unidad lineal rectificada (ReLU) como función de activación en cada capa convolucional y totalmente conectada. Se utilizó la función de regresión logística para entrenar la red; la probabilidad negativa se utilizó para calcular la función de costo durante el entrenamiento. En cuanto a los pesos de inicialización, se probó una distribución gaussiana (con std = 0,01 y valor medio = 0) y una distribución uniforme para la inicialización (también llamada "inicialización de Xavier"). Los mejores resultados se obtuvieron con el peso de inicialización gaussiano.

Se probaron diferentes tasas de aprendizaje: 0,01 de tasa de aprendizaje estático y una tasa de aprendizaje dinámico con un valor inicial de 0,0095 y un decaimiento proporcional de 0,995 por iteración.

La mejor configuración de tasa de aprendizaje utilizada en los experimentos descritos es la tasa de aprendizaje dinámico. Se utilizaron treinta muestras por lote. El procedimiento del entrenamiento fue de 200 iteraciones. Los pesos de las capas convolucionales y las capas totalmente conectadas se aprendieron y adaptaron a este proceso. El código fue desarrollado usando el framework Theano [44] para optimizar el modelo en la GPU.

4.4.1.2. Clasificación

Las muestras se etiquetaron como 0 ó 1, dependiendo de si las muestras pertenecen a casos de buena fe o ataques reales. Por lo tanto, se trata de una tarea de clasificación dicotómica, y no se realiza ninguna distinción entre los ataques. Como resultado, el número de muestras negativas es mayor que el número de muestras positivas. Después del proceso de entrenamiento de la CNN, los clasificadores fueron entrenados y validados, obteniendo los hiperparámetros que mejor se adaptan a la tarea de clasificación.

4.4.2. Descripción experimental

Para probar el sistema propuesto, se diseñó un conjunto de experimentos completo. Los diferentes escenarios se presentan a continuación. El primero es la evaluación individual, y luego, se realizan dos enfoques de concatenación: concatenar las tres bases de datos antes del procedimiento CNN (en el nivel de característica) y concatenar los vectores de características después de cada procedimiento CNN independiente (en el nivel de clasificación).

La base de datos se dividió en dos partes. La primera parte se componía de 185 imágenes/sujetos de buena fe. Por otro lado, la combinación del número de ataques (cuatro) y las imágenes de buena fe arrojó un total de 740 imágenes de ataques. Por otra parte, el total de la base de datos para el entrenamiento de la CNN se dividió en dos subconjuntos (entrenamiento y validación). El ochenta y cinco por ciento de las imágenes se utilizaron para el proceso de entrenamiento y las imágenes restantes el 15% se seleccionaron para el proceso de validación. Por lo tanto, 157 de 185 de las imágenes de buena fe y 629 de las 740 imágenes de ataque se utilizaron para entrenar la red. Asimismo, 28 de las 185 imágenes de buena fe y 111 de 740 fueron seleccionadas para el proceso de validación.

En cuanto al entrenamiento de los clasificadores, hay que destacar que todas las imágenes se utilizaron para este proceso, y la distribución de imágenes fue desequilibrada. Para evitar este hecho, era obligatorio llevar a cabo un método de remuestreo. Este método se basó en el uso del proceso de validación cruzada k-fold, como se describe en [45]. Los dos subconjuntos, entrenamiento y prueba, se construyeron n veces. Se seleccionaron muestras aleatorias de entrenamiento y prueba del total de muestras, siguiendo la proporción de 75% para el entrenamiento y 25% para la prueba. A continuación, se calcularon la tasa de error de clasificación de presentación de ataques (APCER) y la tasa de error de clasificación de presentación de buena fe (BPCER). Como era de esperar, una valor de K igual a 5 es adecuado para evitar este desequilibrio.

La evaluación de un sistema PAD se ha descrito en los últimos años utilizando diferentes mediciones, pero se ha alcanzado un punto de vista común con la definición de la norma ISO (IEC 30107-3:2016). En este estándar, la detección de la capacidad de ataque se mide con errores: tasa de error de clasificación de presentación de ataque (APCER), tasa de error de clasificación de presentación de buena fe (BPCER) y tasa de error de clasificación promedio (ACER), tal y como se define a continuación.

- **La tasa de error de clasificación de presentación de ataques (APCER)** se define como la proporción de ataques de presentación que se clasificaron incorrectamente (como de buena fe) [46] Ecuación 4.1.

$$\text{APCER}_{\text{PAIs}}=1-\left(\frac{1}{|PAI|}\right)\sum_{w=1}^{|PAI|}(RES_w)$$

Ecuación 4.1. *Fórmula para el cálculo de APCER*

Dónde |PAI| es el número de instrumentos de ataque de presentación (PAI) y RESω toma el valor 1 si la presentación ω se evalúa como un ataque y 0 si se evalúa como de buena fe. Un PAI se define como un objeto o rasgo biométrico utilizado en un ataque de presentación.

- **La tasa de error de clasificación de presentación de buena fe (**BPCER) se define como la proporción de presentación de buena fe clasificada incorrectamente como ataques de presentación [46] Ecuación 4.2.

$$\text{BPCER}_{\text{PAIs}} =\frac{\sum_{w=1}^{|BF|}(RES_w)}{|BF|}$$

Ecuación 4.2. *Fórmula para el cálculo de BPCER*

Dónde | BF| es el número de las presentaciones de buena fe y RES_i Devuelve el valor 1 si la presentación ω se asigna como ataque ab y 0 si se analiza como de buena fe.

- **Tasa de error medio de clasificación (ACER):**es la media ponderada entre APCER y BPCER

$$\text{APCER}_{\text{PAIs}} = \frac{\text{APCER}_{\text{PAIs}} + \text{BPCER}_{\text{PAIs}}}{2},$$

Ecuación 4.3. *Fórmula para el cálculo de APCER*

4.4.2.1. Primer caso de estudio: evaluación unimodal

En este experimento, el objetivo fue evaluar los resultados obtenidos utilizando cada tipo de información individualmente, por lo que cada subconjunto (visible, térmico y NIR) se probó por separado**.** Cada subconjunto se utiliza para entrenar de forma independiente una CNN y los clasificadores. En este caso, se obtuvieron tres resultados diferentes e independientes, uno por subconjunto que se reflejan en la figura 4.3.

4.4.2.2. Segundo caso de estudio: fusión multimodal a nivel de clasificador

En este experimento, cada subconjunto se usó para entrenar una CNN separada; por lo tanto, se entrenaron tres CNN con la misma arquitectura pero con diferentes datos de entrenamiento. Cada red neuronal se entrenó con un tipo de imagen (VIS, NIR o térmica). Las salidas de las tres redes neuronales se concatenaron en un vector característico, cuya dimensión es la suma de los subconjuntos tridimensionales, como se representa en figura 4.5. Se alimentó un vector concatenado a los clasificadores. Al igual que en el primer caso, los clasificadores fueron entrenados a continiuación.

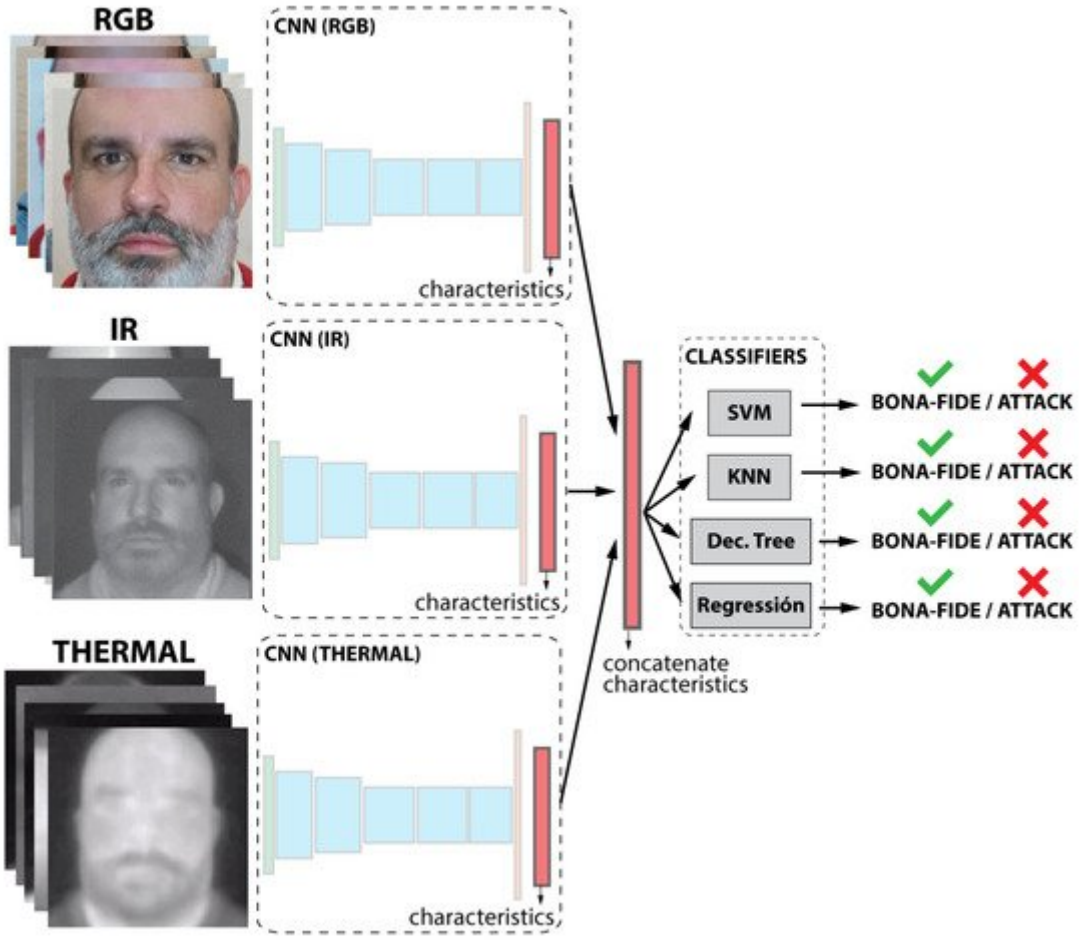

Figura 4.5. *Representación del segundo caso de estudio. Fusión multimodal a nivel de clasificador.*

4.4.2.3. Tercer caso de estudio: fusión multimodal a nivel de caracterización

Este experimento describe el nivel de característica. Las imágenes de luz visible proporcionan tres canales diferentes. El primer canal es para el color rojo (R), el segundo canal es para el color verde (G) y el último canal es para el color azul (B). Sin embargo, las imágenes térmicas y NIR presentan sólo un canal en escala de grises; por esta razón, este tipo de imágenes contribuyen con un único canal adicional a cada una. Por lo tanto, la salida resultante de concatenar imágenes de tres subsistemas diferentes devuelve una imagen de cinco canales (R, G, B, NIR y Térmica), como se muestra en figura 4.6. Esta imagen compuesta por cinco canales (R, G, B, NIR y Térmica) es la entrada a la red neuronal.

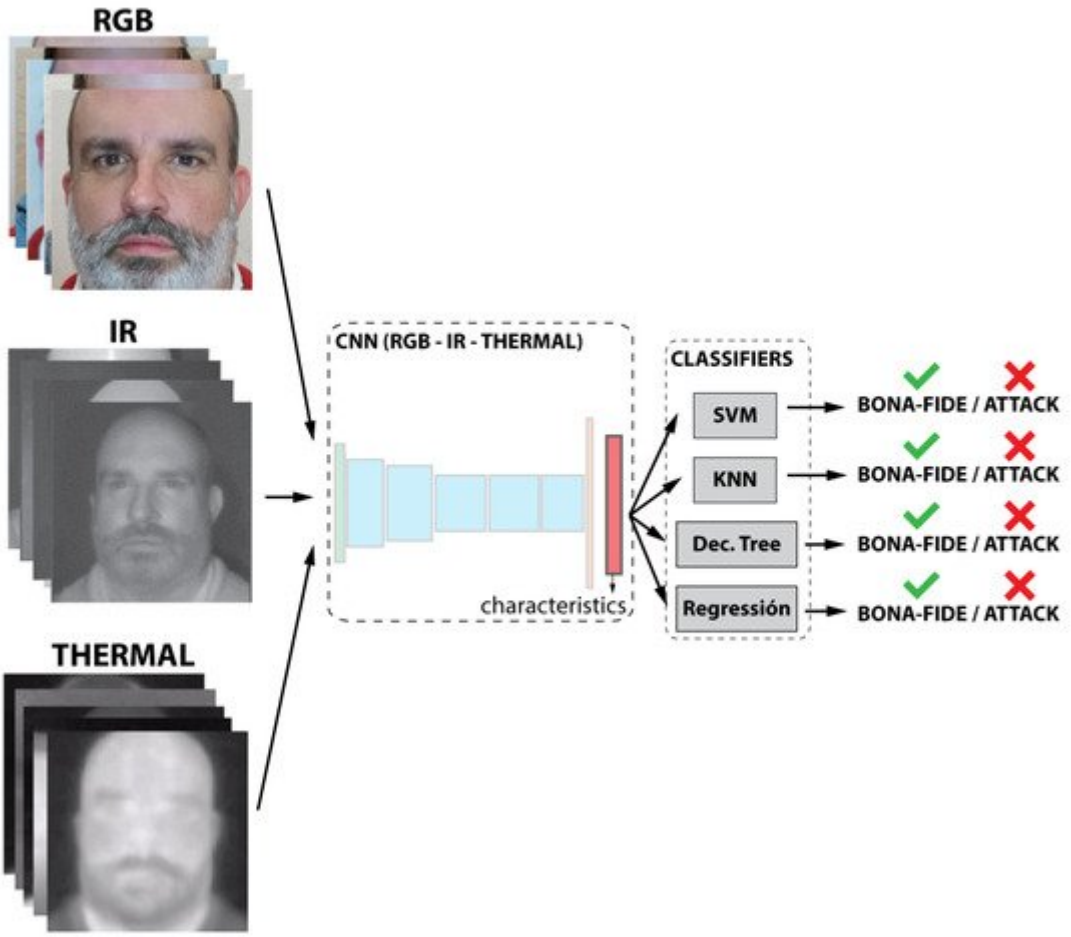

Figura 4.6. *Representación del tercer caso de estudio. Fusión multimodal a nivel de característica.*

En resumen, se obtuvieron cinco resultados diferentes: visible, NIR y térmico (independientemente); las tres bases de datos agregadas en el nivel de caracterización; y las tres bases de datos agregadas a nivel de clasificación.

4.5. Resultados y discusión

4.5.1. Resultados de la evaluación unimodal

El modelo se probó por separado para cada subconjunto (visible, NIR y térmico). Los resultados se resumen en la tabla 4.1, mostrando los parámetros APCER y BPCER obtenidos con el mejor resultado de cada clasificador.

	Visible			NIR			Térmico		
	APCER (%)	BPCER (%)	ACER (%)	APCER (%)	BPCER (%)	ACER (%)	APCER (%)	BPCER (%)	ACER (%)
SVM RBF	2.45	23.40	12.9	1.23	0	0.61	0	6.38	3.19
SVM Lineal	2.45	21.27	11.8	1.23	0	0.61	0	6.38	3.19
KNN	2.45	21.27	11.8	1.23	0	0.61	1.84	6.38	4.11
Árbol de decisión	4.29	21.27	12.7	18.4	0	9.2	0	8.51	4.25
Regresión logística	3.06	7.41	5.23	18.4	0	9.2	0	4.26	2.13

Tabla 4.1. *Resultados unimodales*

Como se muestra en la tabla de salida, en general, los resultados obtenidos con el NIR y la cámara térmica fueron mejores que los resultados logrados con el sensor visible. Para la NIR, todas las muestras positivas (BPCER) se clasificaron correctamente, mientras que el número de muestras negativas (APCER) clasificadas incorrectamente fue muy bajo. Teniendo en cuenta las imágenes térmicas, se logró casi un 100% de precisión para los ataques, y el número de muestras positivas clasificadas erróneamente fue relativamente bajo. El sensor visible tuvo también un rendimiento aceptable, y los resultados obtenidos no fueron mucho peores que con los otros sensores. Lógicamente los ataques se diseñan por humanos, por lo que era de esperar que funcionasen especialmente bien con una cámara con características similares al ojo humano.

Vale la pena destacar que el sensor térmico es el más adecuado para escenarios de alta seguridad, ya que los todos los ataques están correctamente clasificados. La desventaja de los sensores térmicos es que, en algunos casos, un pasajero de buena fe que se detecta como un delincuente, y esta situación implica molestar a los pasajeros y ocupar a los guardias fronterizos dedicados a confirmar estos falsos positivos. Además, los sensores térmicos suelen ser más caros que los sensores visuales. Es por ello que esta es una situación adecuada para aquellos casos en que la directiva principal sea la de que ningún pasajero de mala fe pueda escapar.

Sin embargo, el sensor NIR es el mejor equipado para entornos no tan restrictivos en los que se garantiza el bajo riesgo, ya que los pasajeros de buena fe siempre están correctamente clasificados. Los resultados del NIR muestran que algunos delincuentes pueden ser clasificados como pasajeros de buena fe. En estas situaciones de bajo riesgo, se puede tomar una decisión a partir de imágenes obtenidas por este sensor, ya que los guardias fronterizos no serán interrumpidos con falsos positivos y la ausencia de molestias a los pasajeros de buena fe es más relevante que la detección de los ataques realizados.

El sensor visible es un buen compromiso entre ambas opciones. Los resultados de APCER y BPCER se mantienen relativamente bajos; por lo tanto, el comportamiento del clasificador es muy uniforme en todos los escenarios. De hecho, el sensor visible tiene una ventaja notable en situaciones ABC porque las imágenes proporcionadas para el sensor se pueden enviar directamente a la guardia de fronteras. Además, el coste de adquisición y reemplazo de los sensores visibles es insignificante en comparación con el costo de los otros sensores.

Los resultados del clasificador dependen en gran medida del conjunto de datos seleccionado. En cuanto a las imágenes visibles, KNN y SVM obtienen el mismo resultado; por lo tanto, ambos pueden ser considerados los mejores clasificadores. Tal vez la razón principal se basa en que CNN devuelve un buen conjunto de características, y un clasificador simple como KNN puede obtener una buena puntuación de clasificación. Aumentar la complejidad del algoritmo mediante un árbol de decisión o una regresión logística no mejorará drásticamente las puntuaciones obtenidas previamente. Se pueden hacer comentarios similares relacionados con las imágenes térmicas o NIR. Finalmente, en el sensor térmico, el clasificador SVM obtiene un resultado ligeramente mejor que KNN.

4.5.2. Resultados para la fusión a nivel de clasificador

Este modelo se probó analizando una imagen de cada sensor con una CNN específica y se concatenaron las tres salidas en un único vector. Este nuevo vector es la entrada de la etapa clasificadora, en la que se prueban los clasificadores. Los resultados obtenidos de cada clasificador se resumen en la Tabla 4.2. En la mayoría de los clasificadores, las muestras de ataque se clasificaron con buena precisión y las muestras clasificadas erróneas corresponden a usuarios genuinos. Por lo tanto, estos resultados muestran el uso potencial en entornos en los que la seguridad es el objetivo principal. En un ABC, cuando el nivel de seguridad necesario es alto, se puede considerar este tipo de fusión. La única excepción a estos resultados es el algoritmo de árbol de decisión que proporciona el comportamiento opuesto. Para esta situación, la SVM lineal supera a la KNN y a la función de base radial SVM. Hay que tener en cuenta que la SVM lineal y la regresión logística logran el mismo rendimiento. Comparando las puntuaciones de la Tabla 4.2 con la evaluación unimodal de la Tabla 4.1, se puede observar que la adición de más información a los clasificadores proporciona un mejor resultado. La principal desventaja de esta fusión se debe principalmente al coste de agregar tres sensores a la puerta ABC. Más sensores aumentan el presupuesto del sistema en general y producen mayores gastos de mantenimiento.

	APCER (%)	**BPCER (%)**	**ACER (%)**
SVM RBF	0	6.38	3.19
SVM Lineal	0	2.13	1.06
KNN	0	4.26	2.13
Árbol de decisión	1.84	0	0.92
Regresión logística	0	2.13	1.06

Tabla 4.2. *Fusión a nivel de clasificador.*

4.5.3. Resultados para la fusión a nivel de caracterización

El último clúster de resultados corresponde a los tres subconjuntos agregados en el nivel de caracterización, es decir, las imágenes se concatenan y luego alimentan a

una CNN. El resultado se muestra en la Tabla 4.3 para cada clasificador. Los pasajeros de buena fe siempre se clasifican correctamente, mientras que, en algunos casos, los delincuentes se clasifican incorrectamente como pasajeros de buena fe. Este resultado desaconseja el uso de este método de fusión para una situación en la que se requiere una alta seguridad. Este modelo se puede utilizar en situaciones normales o permisivas, en las que el objetivo principal es mantener un flujo de pasajeros fluido.

El mejor resultado se obtiene con el algoritmo de regresión logística. Los clasificadores KNN y SVM logran resultados similares. Comparando los resultados entre el resultado unimodal NIR y la fusión a nivel de caracterización, se puede observar que las salidas son muy similares en este caso particular. Por lo tanto, este último caso de estudio no parece mejorar los resultados unimodales.

	APCER (%)	**BPCER (%)**	**ACER (%)**
SVM RBF	1.23	0	0.61
SVM Lineal	1.23	0	0.61
KNN	1.23	0	0.61
Árbol de decisión	1.84	0	0.92
Regresión logística	0.61	0	0.31

Tabla 4.3. *Fusión a nivel de característica.*

4.5.4. Discusión de los resultados

El enfoque propuesto se evaluó con diferentes sensores y sistemas de adquisición escenarios muy similares a ABC reales. Los resultados obtenidos con las cámaras, en el rango visible, son una buena solución para ABC en situaciones normales de funcionamiento. Si se trata de escenarios de alta seguridad, el sensor térmico muestra los mejores resultados, ya que todos los delincuentes están correctamente clasificados. El agrupamiento de las diferentes imágenes muestra que la fusión a nivel de clasificador es la mejor opción para escenarios de alta seguridad, mientras que, en una situación de seguridad relajada, la fusión a nivel de característica parece ser una opción adecuada.

El uso combinado de todas las fuentes de de información proporciona mejores resultados en la mayoría de los casos, utilizando el clasificador o la fusión de nivel de caracterización. Sin embargo, utilizando la información de NIR individualmente, la fusión a nivel de caracterización no parece mejorar excesivamente el rendimiento unimodal. Finalmente, los clasificadores como KNN o SVM muestran potencial discriminativo, manteniendo una complejidad lo suficientemente baja.

Comparando los resultados obtenidos con estudios similares en la literatura, parece que estos son competitivos. En la Tabla 4.4, se muestra una comparación exhaustiva entre el enfoque de este trabajo y una selección de trabajos de investigación recientes con experimentos similares.

Como ya se ha puesto de manifiesto, no resulta sencilla la selección de estudios con conjuntos de datos y condiciones de evaluación similares. Algunos puntos esenciales que hay que tener en cuenta son las métricas de evaluación, las bases de datos, el número de ataques y los sensores involucrados. Por otra parte, no todos los trabajos de investigación han adoptado métricas ISO estándar [12], que se han establecido, recientemente. Asimismo, en la mayoría de los casos, los las bases de datos utilizadas han sido adquiridas en un ambiente controlado, mientras que el presente estudio obtuvo todas las imágenes en un escenario real. Por último, dos trabajos de investigación recientes podrían considerarse como los estudios con condiciones más similares [26, 47].

BASE DE DATOS	PERSONAS/ATAQUES	ALGORITMO	SENSOR	APCER %	BPCER %	ACER %
CASIA - SURF	1000/Imágenes y máscaras	Basado en RESNET-18	RGB	40.3	1.6	21.0
			PROFUNDIDAD	6.0	1.2	3.6
			NIR	38.6	0.4	19.4
			RGB+PROFUNDIDAD	5.8	0.8	3.3
			RGB+NIR	36.5	0.005	18.3
			PROFUNDIDAD + NIR	2.0	0.3	1.1
			RGB+PROFUNDIDAD+NIR	1.9	0.1	1.0
WMCA	72/ Imágenes, gafas, reproducción y máscaras	Red Neuronal Convolucional Multicanal	ESCALA DE GRISES+PROFUNDIDAD+ NIR+TÉRMICO	0.6	0	0.3
			ESCALA DE GRISES+PROFUNDIDAD+ NIR	2.07	0	1.04
			ESCALA DE GRISES	65.65	0	32.82
			PROFUNDIDAD	11.77	0.31	6.04
			NIR	5.03	0	2.51
			TÉRMICO	3.14	0.56	1.85
FRAV - ATTACK	Este trabajo	SVM RBF	RGB	2.45	23.4	12.9
		SVM RBF	NIR	1.23	0	0.61
		SVM RBF	TÉRMICO	0	6.38	3.19
		SVM RBF (Fusión a nivel de Clasificador)	RGB+NIR+TÉRMICO	0	6.38	3.19
		SVM RBF (Fusión a nivel de Caracterización)	RGB+NIR+TÉRMICO	1.23	0	0.61
		SVM Lineal	RGB	2.45	21.27	11.8
		SVM Lineal	NIR	1.23	0	0.61
		SVM Lineal	TÉRMICO	0	6.38	3.19
		SVM Lineal (Fusión a nivel de Clasificador)	RGB+NIR+TÉRMICO	0	2.13	1.06
		SVM Lineal (Fusión a nivel de Caracterización)	RGB+NIR+TÉRMICO	1.23	0	0.61
		KNN	RGB	2.45	21.27	11.8
		KNN	NIR	1.23	0	0.61
		KNN	TÉRMICO	1.84	6.38	4.11
		KNN (Fusión a nivel de Clasificador)	RGB+NIR+TÉRMICO	0	4.26	2.13
		KNN (Fusión a nivel de Caracterización)	RGB+NIR+TÉRMICO	1.23	0	0.61
		Árbol de Decisión	RGB	4.29	21.27	12.7
		Árbol de Decisión	NIR	18.4	0	9.2
		Árbol de Decisión	TÉRMICO	0	8.51	4.25
		Árbol de Decisión (Fusión a nivel de Clasificador)	RGB+NIR+TÉRMICO	1.84	0	0.92
		Árbol de Decisión (Fusión a nivel de Caracterización)	RGB+NIR+TÉRMICO	1.84	0	0.92
		Regresión Logística	RGB	3.06	7.41	5.23
		Regresión Logística	NIR	18.4	0	9.2
		Regresión Logística	TÉRMICO	0	4.26	2.13
		Regresión Logística (Fusión a nivel de Clasificador)	RGB+NIR+TÉRMICO	0	2.13	1.06
		Regresión Logística (Fusión a nivel de Caracterización)	RGB+NIR+TÉRMICO	0.61	0	0.31

Tabla 4.4. *Comparación entre los resultados actuales del estudio y otros trabajos de investigación.*

Comparando los resultados de las imágenes RGB o en escala de grises, se puede observar que el trabajo de investigación presentado disminuye un orden de magnitud el valor de APCER (de 40,3 en [47], ó 65,65 en [26] a 2,45 APCER en este trabajo). Por el contrario, en el caso de BPCER, el presente estudio aumenta los resultados de BPCER. Este resultado sería aceptable, ya que una frontera ABC es un entorno de alta seguridad. Este entorno está diseñado principalmente para evitar los APCER. En resumen, la media ACER se reduce a 11,8, un mejor resultado que en los otros dos estudios.

En cuanto a los resultados de la NIR, este trabajo mejora los estudios previos seleccionados. El APCER se disminuye de 5,03 [26] a 1,23, mientras que el BPCER obtenido logra un valor de 0 en ambos casos. Centrándose en los resultados térmicos, sólo George [26] presentó este tipo de sensor. Una vez más, el trabajo presentado muestra una reducción notable de APCER (0 contra 3,15) pero aumentando BPCER (4,26 contra 0,56) y ACER.

Finalmente, analizando los resultados de fusión presentados con los de los trabajos seleccionados, los autores han elegido dos configuraciones diferentes en su aproximación al problema.

Por un lado, la regresión logística (fusión en el nivel de clasificador) alcanzó el APCER y el registro más bajos. La regresión (fusión en el nivel de Caracterización) logró el BPCER más bajo. Por otro lado, el primer sistema devuelve valores muy similares a los obtenidos en el estudio de George [26],en el que APCER es 0,6, y ACER es 0,3 en ambos casos. En el segundo estudio, los resultados obtienen 0 APCER, aumentando el BPCER de 0 a 2,13.

4.6. Conclusiones

Este trabajo propone un método de detección de ataques de presentación facial en el control de fronteras prácticamente apto para un uso real. Se concluye que las CNN son capaces de aprender características para la detección de ataques de presentación facial multiespectral (visible, infrarrojo cercano y térmico). Se utilizaron varios clasificadores diferentes (lineal y RBF SVM, KNN, árbol de decisión y regresión logística).

Se consideraron cuatro tipos ataques diferentes utilizando los siguientes artefactos: foto impresa, máscara impresa, máscara impresa con los ojos recortados y una imagen mostrada en una tableta. Considerando los resultados experimentales, parece que el modelo de las CNN es ideal para la detección de ataques de presentación, obteniendo buenos resultados.

En situaciones operativas normales para los sistemas ABC, las imágenes visibles se pueden considerar una buena opción, tal y como se muestra en los resultados, especialmente considerando que suelen ser los principales o únicos sensores que se pueden adquirir fácilmente a bajo coste.

Para situaciones de alta seguridad, el sensor térmico muestra un mejor resultado ya que todos los atacantes fueron correctamente clasificados, mientras que en sólo en unos pocos casos se puede detectar como delincuente a un pasajero de buena fe. La situación opuesta se logra en el caso del sensor NIR: todos los viajeros de buena fe están correctamente categorizados, pero algunos ataques no fueron detectados y se clasificaron erróneamente como auténticos.

La fusión de las diferentes imágenes muestra que la fusión a nivel de clasificador es la mejor opción para escenarios de alta seguridad. Sin embargo, en una operación de seguridad más permisiva, la fusión a nivel de caracterización parece ser la

mejor opción. El uso de todas las fuentes de información muestra mejores resultados que el uso de sensores aislados, tanto cuando la fusión se realiza en el nivel de clasificador como las fusiones a nivel de caracterización. Finalmente, clasificadores como KNN o SVM presentan suficiente poder discriminativo para mantener baja la complejidad del sistema.

Este trabajo ha dado lugar a una publicación cuyos datos son los siguientes.

Convolutional neural network approach for multispectral facial presentation attack detection in automated border control systems

Araceli Sánchez-Sánchez, M., Conde, C., Gómez-Ayllón, B.,.Palacios-Alonso, D., Cabello, E.

Entropy, 2020, 22(11), pp. 1–18, 1296

CAPITULO 5

DETECTOR AUTOMÁTICO DE CONFLICTOS DE TRÁFICO BASADO EN VISIÓN ARTIFICIAL

5.1. Introducción y estado del arte

En el ámbito de la seguridad vial es evidente que los peatones son las víctimas más frágiles en casos de accidentes de tráfico, y como tal existen múltiples informes y estadísticas que lo confirman. Esto es especialmente relevante en aquellos que tienen lugar en áreas consideradas seguras para ellos, por ejemplo pasos de cebra y semáforos. De hecho, el ejercicio de caminar se reconoce como un ejercicio saludable sin apenas consecuencias desfavorables, excepto aquellas que son causadas por el tráfico rodado.

En las últimas décadas numerosos proyectos han enfocado sus esfuerzos a estudiar y mejorar las condiciones de los pasos de peatones en sus múltiples posibilidades. Evaluaciones de los riesgos de los peatones [1], estudios acerca del comportamiento de la velocidad de los conductores en los pasos de cebra [2] o análisis de situaciones particulares como parejas de adultos con niños en los pasos de peatones [3] son algunos de los estudios recientes que se han realizado para mejorar la seguridad de los mismos. La mayoría de estos estudios obtienen sus datos a partir de bases de datos de accidentes de peatones, que se consideran como una de los elementos más importantes para establecer una estrategia para el desarrollo de sistemas de seguridad integrados en las carreteras [4]. Otros usan observaciones puntuales realizadas por observadores entrenados que miden, categorizan y registran situaciones que de las que son testigos [5].

Por fortuna, los accidentes en pasos de cebra o en pasos de cebra con semáforos ocurren pocas veces, y si alguna vez tienen lugar, los datos obtenidos de la escena suelen ser poco claros. Con objeto de proporcionar una seguridad proactiva, debería existir una herramienta capaz para realizar un análisis preventivo en cada cruce con paso de peatones.

5.2. Conflictos de tráfico

La percepción que tienen los peatones y los conductores se ha usado como una medida de seguridad proactiva, comparando los lugares percibidos de riesgo por ellos con los identificados en los partes de accidentes reportados por la policía [6] o estudiando el comportamiento de los conductores [7]. Otras aproximaciones han enfocado sus esfuerzos en el concepto de "conflicto", definido como aquellas situaciones que involucran a 1 ó más peatones o vehículos en los que el peligro de colisión o atropello está presente [8]. La frecuencia de los conflictos es mucho mayor que la de los accidentes y, como es lógico, evitar situaciones peligrosas mejoraría la seguridad.

La técnica del conflicto se desarrolló inicialmente en el laboratorio de Detroit de la General Motors a finales de los años 60 [9], se basaba en el juicio de observadores que normalmente estudiaban y evaluaban grabaciones, lo que supone una técnica muy costosa en tiempo. Los fundamentos de sus investigaciones se presentaron en el Lund Institute of Technology (LTH) en Suecia [5]. La técnica sueca se centra en situaciones donde dos usuarios de la carretera hubieran colisionado sin que ninguno de ellos hubiera tomado una maniobra evasiva. Variables como el punto donde la maniobra evasiva se tomó y el tiempo hasta que el accidente se hubiera producido, estimado por observadores entrenados son útiles para determinar si un conflicto es serio o no.

Dependiendo de la posición del observador y del punto de vista, podemos identificar un conflicto por la posición de las partes que están involucradas en él, su tipo, trayectoria, velocidad y otras propiedades subjetivas que un observador entrenado

puede incluir en sus notas. En trabajos previos [10], observadores entrenados han empleado para el registro de estos conflictos materiales como cinta de medida, tiza, sprays y portátiles para recoger los datos [11].

Nuestro objetivo era el desarrollo de un grabador automático de conflictos y analizador basado en visión artificial que nos proporcione una amplia base de datos de conflictos entre peatones y vehículos o entre de varios vehículos, así como los datos técnicos tales como la posición y velocidad de las partes involucradas. También se recogen las imágenes y se guarda la hora exacta a la que los posibles conflictos han tenido lugar para que observadores entrenados pueda puedan repasar esas situaciones y centrar su atención en esos puntos exactos sin perder el tiempo observando comportamientos que no son relevantes. De esta forma la cantidad de las observaciones útiles puede ser incrementada sin elevar significativamente el tiempo y los recursos necesarios para el tratamiento de los datos.

5.3. Base de datos

En colaboración con la sección de la DGT de la ciudad de Salamanca fueron seleccionados dos pasos de cebra.

Salamanca es una ciudad de tamaño medio en el noroeste de España con una población de 140000 habitantes donde el centro histórico está mayoritariamente peatonalizado y los vehículos tienen que rodearlo. La Policía Local informa de que cada año se producen alrededor de 1000 accidentes, 100 de ellos implican atropellos que involucran a unos 120 peatones. Los datos disponibles de estos accidentes incluyen edad, género y gravedad (4 personas muertas, 36 heridos graves y 80 leves). Las características de los dos cruces analizados se muestran en la siguiente figura:

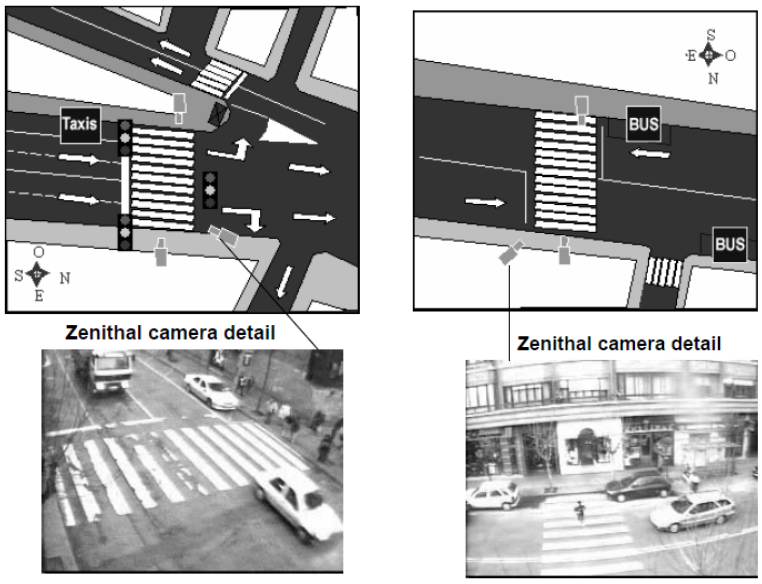

Figura 5.1. *Cruces seleccionados en la ciudad de Salamanca y la disposición de las cámaras.*

El primero de ellos se localiza en un paso de cebra en una calle de doble sentido con una parada de autobús cercana al paso de cebra. No está regulado por ningún semáforo y los dos sentidos están permitidos para los vehículos, por lo tanto los peatones tienen que parar y estar atentos a ambos sentidos para cruzar. El aparcamiento también está permitido en ambos sentidos de la carretera. La parada de autobús está situada antes del paso de peatones, justo enfrente del emplazamiento de la cámara. Cuando el autobús realiza la parada es un obstáculo para los vehículos que se mueven de derecha a izquierda.

El otro cruce está controlado por un semáforo en una calle de una sola dirección con una parada de taxi en uno de los lados de la calle. Los vehículos disponen de dos carriles de circulación y la parada de taxis está justo delante del paso de peatones de forma que el primer taxi está esperando en el borde del paso de cebra. La configuración de este punto ha sido cambiada recientemente de dos a un único sentido debido a la alta densidad del tráfico de la zona. El cambio de disposición del tráfico no ha sido significativo respecto al número de accidentes ocurridos.

Estos puntos se seleccionaron entre cientos de posibles candidatos para estudiar la influencia de la parada de autobús y de taxis en el comportamiento de los peatones y

de los vehículos, para valorar si la información recogida en este trabajo evidenciaba que era necesario recurrir a posibles alternativas en estos servicios para mejorar la seguridad.

Un ordenador y tres cámaras fueron situados de forma estratégica en los puntos de cruce para causar las menores alteraciones al entorno. Dos cámaras se situaron en ambos lados, capturando una visión frontal del cruce, cada una en un lado de la calle. La otra estaba a 10 metros de altura sobre el cruce capturando una visión superior del mismo, de esta forma el comportamiento de los peatones y su trayectoria se podía seguir antes, durante y después de que hubieran cruzado la carretera.

Las imágenes se grabaron y se almacenaron en niveles de gris con una resolución de 240 x 180 píxeles, a una velocidad de 4 imágenes por segundo en cada cámara, suficiente para llevar a cabo el análisis. Los periodos de grabación fueron desde el 18 de enero hasta el 14 de febrero y desde el 9 de marzo al 30 de abril durante el año 2002. Se grabaron situaciones tanto durante el día como la noche, la iluminación artificial nocturna resultó ser suficiente para permitir al sistema funcionar correctamente. Este experimento contrasta con otros intentos previos de analizar periodos de días, partes de días o con un número limitado de peatones, ya que en este caso se cuenta con información más continuada y en condiciones de iluminación y atmosféricas muy variadas dado el periodo de tiempo recogido.

Se han identificado dos principales causas del ruido y el proceso de análisis de las imágenes se ha diseñado para minimizar sus consecuencias.

El primero es debido al proceso de adquisición, en este nivel se han considerado todos los efectos de ruido que provienen de las condiciones técnicas y experimentales. Debido a la disposición técnica (más de 5 metros entre la cámara y el ordenador) las imágenes sufren serios efectos de ruido principalmente causadas por presencia de campos electromagnéticos y la atenuación de la señal transmitida. Se emplean algoritmos para procesar las imágenes que minimizan estos efectos.

El segundo tipo son los originados por la saturación de la cámara producidos por fuertes cambios de iluminación. En las localizaciones seleccionadas para las CCD se saturan de luz cuando hay un fuerte reflejo, por ejemplo al pasar un autobús por la escena. El filtro de Kalman que se usa puede realizar el seguimiento de objetos con una pequeña cantidad de imágenes disponibles.

En lo referido a la distancia, la presencia de sombras no influye ni en el seguimiento ni en el cálculo de la trayectoria, pero se han tenido en cuenta las pequeñas diferencias entre el tamaño real y el percibido de los objetos. El centro de gravedad de cada objeto se toma como punto de referencia para el seguimiento ya que permanece razonablemente estable durante el desplazamiento del objeto por la escena.

El número de imágenes adquiridas por segundo es una variable crítica. El número seleccionado proporciona un equilibrio entre la información obtenida y los requerimientos del seguimiento. Por su mayor velocidad, los vehículos necesitan más imágenes por segundo que los peatones para llevar a cabo un seguimiento preciso. El filtro de Kalman empleado nos ofrece suficiente fiabilidad incluso con pocas imágenes.

Además, se tiene que tener en cuenta que una imagen es una proyección bidimensional de una escena tridimensional, de esta forma, la posición de la cámara determina la apariencia del objeto. La estimación de Kalman de la velocidad y la aceleración se calcula en las imágenes adquiridas, que son proyecciones bidimensionales y por lo tanto estas estimaciones no son la velocidad real del objeto y la aceleración, pero es una estimación bastante aproximada. Las ecuaciones simplificadas para esta proyección perspectiva son las siguientes:

$$u = \frac{X_{3\mathrm{d}}}{\frac{Z_{3\mathrm{d}}}{D} + 1}$$

$$v = \frac{Y_{3\mathrm{d}}}{\frac{Z_{3\mathrm{d}}}{D} + 1}$$

Ecuación 5.1. *Ecuaciones utilizadas.*

Dónde (u,v) son las coordenadas de imágenes tomadas y (X,Y,Z) son las coordenadas reales del objeto en 3D y D es la distancia focal (empleando un modelo simplificado de cámara pin-hole). La relación entre la velocidad y la aceleración estimada en la imagen y el objeto real no es una aproximación trivial. Para los objetos lejanos la velocidad estimada y la aceleración son diferentes de la real, sin embargo, cuando el objeto se aproxima a la cámara, la velocidad y la aceleración real y la estimada se aproximan. Para estimar la velocidad real y la aceleración se pueden utilizar varios métodos, como por ejemplo la hipótesis del plano del suelo [12] o el uso de cámaras no calibradas con distintos campos de vista que se superponen [13].

5.4. Análisis de las imágenes

En este caso, al no tratarse de un sistema en tiempo real, el procesado y el análisis de las imágenes se realiza a posteriori, una vez que estas se capturaron y almacenaron.

Es necesario distinguir exactamente los objetos en movimiento de los estáticos. Para evaluar y seguir los objetos en movimiento se realiza la resta del fondo [14,15].

Para obtener una imagen de fondo actualizada continuamente, un conjunto seguido de imágenes se organizan en una cola LIFO. Se consideraron inicialmente dos métodos: evaluación de la media o de la moda del nivel de gris de cada pixel. El número de imágenes para formar este conjunto era la variable que debía ser elegida. Se hicieron los test con conjuntos de 5 a 20 imágenes. El experimento nos llevó a la conclusión de que emplear una selección de 10 imágenes y el empleo de la moda como método de evaluación de cada nivel de gris de cada pixel nos permitía realizar el cálculo más adecuado del fondo de la escena. Sin embargo, se hizo necesario un estudio preciso para cada cruce, ya que nos podía llevar a unos mejores resultados para cada situación seleccionada.

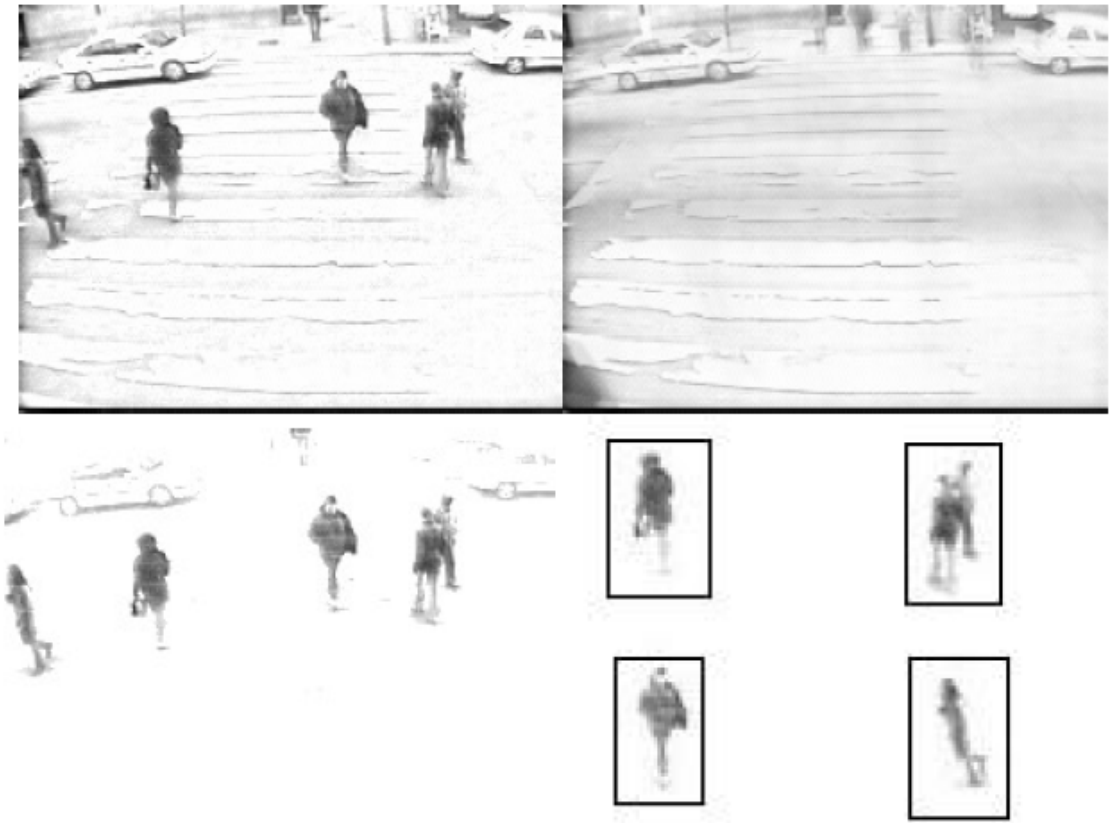

Figura 5.2. *Arriba izquierda, imagen original capturada. Arriba derecha, el fondo de la imagen. Abajo izquierda, la imagen restada. Abajo derecha, los 4 peatones identificados en la escena.*

Después de obtener las imágenes de los objetos en movimiento sobre un fondo en blanco, el siguiente paso era delimitar esos objetos, que a partir de ahora se referirán como “componentes” (Figura 5.2). El cálculo de la envolvente convexa de estos componentes finalmente se estableció a través de un etiquetado secuencial basado en la propiedad de proximidad, mientras que se rechazó un algoritmo de etiquetado matricial debido al elevado coste computacional y la poca mejora que se obtenía con el mismo [16,17].

Mediante la cobertura consecutiva de cada píxel en la imagen, la comprobación y etiquetado de cada píxel activo en correspondencia con sus vecinos, se obtiene una imagen etiquetada donde cada componente será identificado.

Una vez que los componentes están etiquetados, se aplica un filtro para ignorar aquellos componentes que no alcanzaron una cantidad mínima de pixeles y que se consideran como ruido producido por el proceso de sustracción de imágenes. A continuación cada componente es extraído en una imagen separada. Los vehículos y los peatones se distinguían claramente debido a su diferente tamaño y a sus trayectorias ortogonales. Es en este punto cuando comienza el seguimiento y el proceso de grabación de las características de cada componente.

5.5. Seguimiento a través del filtro de Kalman

El filtro de Kalman es una técnica iterativa diseñada para predecir en una serie de modelos de series temporales [18]. Cada trayectoria de peatones y de vehículos se ha modelado de acuerdo un sistema dinámico lineal sometido a ruido gaussiano. Se desarrolló un filtro modificado de Kalman [15, 19]. El algoritmo adapta su modelo a cada paso para mejorar la estimación del movimiento de cada componente. La posición, velocidad y valores de aceleración en diferentes pasos se usan para predecir la localización futura del objeto.

El proceso de seguimiento se puede definir de este modo. Para modelar trayectorias suaves se emplea un vector de estado que proporciona información de cada componente en un instante de tiempo $t\ (p_t, v_t, a_t, a_{t-1}, a_{t-2})^t$, dónde p, v y a representan la posición, velocidad y aceleración en el paso descrito por el subíndice. Los vectores de estado proporcionados por las estimaciones de Kalman se denominarán como x_t mientras que los vectores de estado real obtenidos de las observaciones de las imágenes se denominarán como y_t.

El proceso iterativo predice vectores de estado a través de la matriz de transición *A* que relaciona el vector de estado de instante de tiempo previo *k-1* al estado correspondiente al instante *k.* Donde *C* es la matriz de coeficientes que relaciona el estado predicho con el medido y_k. El proceso y la medida del ruido, q_k y r_k se asume que

son independientes entre sí, es blanco y con distribuciones de probabilidad normal, media 0 y matrices de covarianza *Q* y *R* respectivamente.

$$x_k = Ax_{k-1} + q_k, \quad q \sim N(0,Q).$$
$$y_k = Cx_k + r_k, \quad r \sim N(0,R).$$

El filtro de Kalman es esencialmente un conjunto de ecuaciones matemáticas que implementan un estimador predictivo del tipo predictor-corrector que minimiza el error estimado en la covarianza. La forma en que funciona es a través de procesos de retroalimentación, el filtro estima el vector de estado en un instante de tiempo y obtiene retroalimentación en forma de medidas ruidosas. Por lo tanto, las ecuaciones de Kalman se pueden clasificar en dos conjuntos, estimaciones previas y posteriores correcciones.

La parte anterior está formada por un vector de estado predicho $\hat{x}_{\bar{k}}$ y un error de covarianza predicho, $S_{\bar{k}}$, ambos en un instante de tiempo futuro *k*.

$$\hat{x}_{\bar{k}} = Ax_{k-1}$$

Entonces, en un instante de tiempo *k*, cuando se obtiene la medida correcta y_k , se calcula la ganancia de Kalman, K y la estimación anterior se actualiza, $\hat{x}_k$, proporcionalmente al error previo obtenido. También la covarianza del error previo lleva a una nueva actualización del error de covarianza, S_k.

$$K_k = S_{\bar{k}} C^t (C S_{\bar{k}} C^t + R)^{-1}$$
$$\hat{x}_k = \hat{x}_{\bar{k}} + K_k (y_k - C\hat{x}_{\bar{k}})$$
$$S_k = (I - K_k C) S_{\bar{k}}$$

Al final de cada iteración, se obtienen el error previo, $e_{\bar{k}} = (y_k - \hat{x}_{\bar{k}})$ y el error posterior, $e_k = (y_k - \hat{x}_k)$.

Las matrices de transición A y C, la covarianza Q del ruido del proceso y la covarianza del ruido medido R son parámetros importantes del filtro de Kalman. Su eficiencia está fuertemente determinada por cómo de bien estos factores se ajustan al problema.

En nuestro sistema se emplea un filtro de Kalman de un solo paso. Esto significa que sólo es necesario realizar un paso adelante en la predicción. La matriz de transición se convierte fácilmente en C=Id.

La matriz de transición A se calcula utilizando las consideraciones físicas clásicas de posición, velocidad y aceleración.

$$\begin{pmatrix} p_k \\ v_k \\ a_k \\ a_{k-1} \\ a_{k-2} \end{pmatrix} = \begin{pmatrix} 1 & t & \frac{t^2 w_1}{2} & \frac{t^2 w_2}{2} & \frac{t^2 w_3}{2} \\ 0 & 1 & tw_1 & tw_2 & tw_3 \\ 0 & 0 & w_1 & w_2 & w_3 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} p_{k-1} \\ v_{k-1} \\ a_{k-1} \\ a_{k-2} \\ a_{k-3} \end{pmatrix}$$

Los pesos (w_1, w_2, w_3) son positivos y verifican que $\sum_i w_i = 1$. Obteniendo una media de aceleración promedio.

Como el movimiento espacial está explicado a través de información bidimensional proporcionada por las cámaras, el parámetro A es corregido después de cada paso de acuerdo a la ganancia de de Kalman, K_k, y el posterior error e_k.

$$A_{k+1} = A_k + \eta(K_k, e_k)$$

Cuando los componentes se han etiquetado y filtrado, la estimación de trayectorias futuras mediante el filtro de Kalman, proporciona al sistema información

crucial para predecir posibles conflictos y para calcular el movimiento de los objetos en el siguiente instante de tiempo. Esto nos permite seguir la trayectoria de cada componente presente en el punto de cruce. (Figura 5.3)

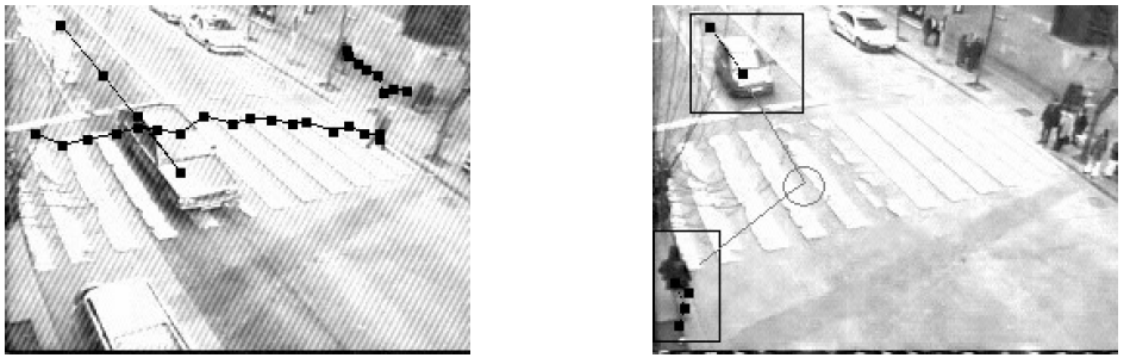

Figura 5.3. *Izquierda: Las trayectorias a lo largo del tiempo de tres componentes. El conocimiento y el análisis de estas trayectorias proporcionan la información necesaria para detectar un posible conflicto. Derecha: Posible conflicto si la velocidad y la trayectoria de uno o de ambos componentes permanece constante.*

5.6. Predicción de conflictos

Una vez que el sistema ha registrado y almacenado la trayectoria, la velocidad y diferenciado entre vehículos y peatones, el análisis de estos datos se realiza de forma automática. El objetivo era detectar posibles conflictos entre peatones y vehículos si sus trayectorias y velocidades no varían de forma significativa, figura 5.3.

Para realizar estas estimaciones se calcula un valor medio de la última trayectoria guardada y de la velocidad. Estamos considerando un movimiento rectilíneo y uniforme a lo largo del punto de cruce tanto para los vehículos como para los peatones. A continuación, se realiza una estimación del número de futuros instantes especificados y se mide la distancia entre los componentes. Si esa distancia tiene un valor menor que un parámetro especificado, se clasifica como conflicto potencial y el tiempo, la posición, la velocidad, la aceleración y la imagen de ocurrencia y se guarda para posterior análisis.

5.7. Resultados

El objeto de este proyecto era comprobar si un sistema de visión por ordenador es lo suficientemente robusto para trabajar en situaciones reales para la detección de conflictos de tráfico. Este objetivo se ha conseguido y se ha desarrollado un sistema de visión artificial por ordenador para procesar secuencias reales. El sistema desarrollado puede ayudar a un operador humano entrenado a estimar conflictos procesando grandes cantidades de video y etiquetando sólo aquellas secuencias en las que se puede producir un conflicto.

Una vez que el video se ha procesado, las situaciones conteniendo los posibles conflictos se etiquetan y un operador puede analizarlas. Esta aproximación representa una enorme ganancia de tiempo, ya que la tarea del operador entrenado se reduce a analizar y confirmar o descartar las potenciales situaciones de conflicto. La cantidad de tiempo necesaria para que la tarea completa la realizase un único observador entrenado hubiera aumentado de forma enorme. El sistema de visión por ordenador puede ayudar al usuario incrementando su eficiencia porque sólo los elementos significativos requieren su atención.

Una vez que las secuencias fueron etiquetadas, se enviaron a la DGT, cuyos observadores entrenados realizaron el de los conflictos detectados. En nuestro caso, en vista de los resultados arrojados por este trabajo el departamento de tráfico realizó cambios en los pasos de peatones que fueron estudiados para mejorar la seguridad.

Para el primer escenario, la parada de autobús resulta ser claramente un importante obstáculo visual para tanto peatones como vehículos, así que se movió a una localización posterior al paso de cebra. Los conflictos en este cruce se debían a los peatones que cruzan delante del autobús o a automoviles que intentan evitar el autobús, con visibilidad reducida de la vía tanto para unos como para otros. La nueva configuración hace que los peatones crucen la calle por detrás del autobús de forma que

se convierten en un objeto claramente visible para los vehículos que se aproximan al paso de cebra.

Para el segundo escenario, la parada de taxis es responsable de la mayor cantidad de conflictos. Cuando el primer taxi realiza la salida de la parada, el escenario cambia de forma imprevista la configuración de la calle, que se transforma de una vía de 2 carriles a 3 carriles. Este cambio repentino puede afectar a los peatones que están esperando en el cruce o a los que están aproximándose a él, ya que no esperan que el taxi entre en movimiento. La parada de taxis se movió lejos del cruce y se permitió el aparcamiento en el lugar donde estaba la parada de taxis, con lo que la frecuencia con la que el carril de aparcamiento se vuelve activo disminuye drásticamente.

El sistema de visión artificial se ha aplicado también a imágenes proporcionadas por la Universidad de Lund [20]. Se estudiaron cinco secuencias de vídeo que contenían conflictos, figura 5.4. Sólo se requirieron pequeños ajustes de parámetros para adaptar el sistema a las nuevas imágenes. No se hicieron cambios en la implementación y los 5 conflictos fueron correctamente identificados y etiquetados. Este es un resultado prometedor aunque deben realizarse más experimentos para probar estos algoritmos.

Figura 5.4. *Seguimiento y detección de dos conflictos de tráfico presentes en el video de la Universidad de Lund. Izquierda: Un coche que llega a la intersección en T no se percata de la aproximación de otro vehículo desde su izquierda. Derecha: Una bicicleta cruzando la calle principal casi colisiona con un coche acercándose por su izquierda.*

Las estadísticas descriptivas sobre los conflictos son ahora una tarea más fácil que realizar. Una vez que los conflictos han sido determinados y especificados durante el análisis automático de las grabaciones, un observador entrenado puede fácilmente comprobar las imágenes y determinar si el peatón o el vehículo son responsables del accidente.

5.8. Conclusiones

La detección y categorización automática mediante visión artificial ha demostrado ser una herramienta poderosa y valiosa para detectar la existencia de conflictos entre vehículos y peatones. Al centrar los esfuerzos en la seguridad proactiva y no en el análisis de las bases de datos, la complejidad del problema se incrementa enormemente hasta convertirse en inabordable sin un procedimiento automatizado.

Por otra parte resulta más eficiente la ausencia de observadores humanos cuya presencia física puede alterar los comportamientos normales de los vehículos y los peatones en la escena. La utilización de pequeñas cámaras, que se colocan y resultan inapreciables para la mayoría de los vehículos y los peatones, mejora la obtención de resultados al mismo tiempo que se reducen los costes y aumentan los períodos de observación.

Los resultados muestran en general que las condiciones específicas de cada cruce son las máximas responsables en el tipo de conflictos que tienen lugar en los alrededores.

En resumen, la herramienta aquí desarrollada puede ser adaptada e instalada en diferentes puntos de cruce con diferentes geometrías y características proporcionándonos una clara y creíble ayuda para la toma de decisiones concretas con el fin de mejorar la seguridad.

Para finalizar, comentar que los resultados obtenidos con este trabajo han dado lugar a una serie ponencias a congresos que se citan a continuación.

Título: Multiple object detection and tracking in a non-constrained environment.

Nombre del congreso: II Workshop Hispano Luso de Agentes Físicos.

Ciudad de realización: Móstoles, España

Fecha de realización: 03/2001

Ciudad: España

Antonio Sanz; Iñigo Martín; M. Araceli Sánchez; Enrique Cabello.03/2001.

Título: Artificial Vision for Road Safety Improvement

Nombre del congreso: II European Congress on Intelligent Transport Systems

Ciudad de realización: Bilbao, España

Fecha de realización: 06/2001

Enrique Cabello Pardos; M. Araceli Sánchez Sánchez; Laura Agudo Mérida; Antonio Sanz Montemayor; Iñigo Martín Sánchez.06/2001.

CAPITULO 6

CONCLUSIONES

En este trabajo se ha mostrado la capacidad de los sistemas de visión artificial para ser aplicados a diferentes problemas reales, partiendo de ideas teóricas es posible llegar a un sistema que tiene aplicación práctica en la vida real en condiciones muy variadas de funcionamiento. Aunque las conclusiones del trabajo han quedado expuestas a lo largo de los distintos capítulos, en este último capítulo se resumen las conclusiones obtenidas en cada uno de ellos

Identificación de personas

Es posible llevar a cabo la identificación de personas en imágenes de reducido tamaño mediante el uso de diferentes tipos de redes neuronales En particular ha sido posible realizar identificaciones de imágenes de caras en diferentes ángulos con tamaños tan pequeños como 32x22 píxeles con porcentajes de aciertos de hasta el 96.67%.

Identificación de rocas en minería

Se ha conseguido desarrollar un sistema de detección de rocas de gran tamaño a la entrada del sistema de machacado que se ha puesto en funcionamiento habitual. Durante los primeros meses de uso no controlado ha conseguido proporcionar niveles de acierto en las detecciones superiores al 70% demandado por la empresa, que además

comunica la percepción de una reducción a la mitad de los tiempos de parada del sistema.

Detección de suplantación de personalidad en una frontera

Se ha desarrollado un sistema PAD operativo en condiciones cuasi-reales de gran versatilidad que puede ser aplicado tanto en entornos de alta seguridad como de seguridad más relajada, existiendo en ambos casos configuraciones capaces de producir un 100% de efectividad en la detección de ataques o en la detección de pasajeros de buena fe, aunque no ha sido posible combinar esos dos resultados en un único sistema.

Conflictos peatón-vehículo

Se ha implementado un sistema de detección de conflictos peatón-automóvil que funciona de una forma prácticamente no detectable, capaz de almacenar una gran cantidad de datos, procesarlos e indicar a un usuario entrenado cuales son los momentos en los que se producen los conflictos. Como consecuencia de los resultados obtenidos por el sistema, la regulación de varios pasos de peatones de la ciudad de Salamanca han sido modificados para reducir la siniestralidad de los mismos.

Como conclusión final hay que destacar que casi todos los estudios realizados en este trabajo han conseguido demostrar su funcionamiento fuera del ambiente del laboratorio, lo que demuestra que los estudios de visión artificial tienen un futuro aun más prometedor.

Por último recopilar todas las aportaciones que los trabajos presentados en esta tesis han supuesto en forma de capítulos de libros, artículos en revistas y ponencias en congresos.

Libros y revistas:

Conde, C.; Sanchez, A.; Cabello, E. “Influence of location over several classifiers in 2D and 3D face verification”. 3161, pp. 153 - 158. Springer Lecture Notes Computer Science GuidelinesSpringer Verlag, 05/2005 .

Tipo de producción: Capítulos de libros

Tipo de soporte: Libro

Enrique Cabello; M. Araceli Sánchez; Luis Pastor.” Some experiments on face recognition with neural networks”. pp. 589 - 599.Springer Verlag, 1998.

Tipo de producción: Capítulos de libros

Tipo de soporte: Libro

Enrique Cabello; M. Araceli Sánchez; Luis Pastor.”Reconocimiento de caras humanas mediante una red neuronal con Ada95”. Revista Ada Spain. 35, pp. 29 - 38.1998.

Tipo de producción: Artículo

Tipo de soporte: Revista

Enrique Cabello; M. Araceli Sánchez; Javier Delgado. “A New Approach to Identify Big Rocks with Applications to the Mining Industry”. Real Time Imaging.8, pp. 1 - 9.02/2002.

Tipo de producción: Artículo

Tipo de soporte: Revista

Enrique Cabello; M. Araceli Sánchez; Julian Nieto; Jesús M. Berrocal; Guido Castro and Javier Delgado. “A Computer Vision Application in Real-Time to

Identifying Big Rocks with Applications to the Minning Industry". pp. 1 - 12.04/2000.

Tipo de producción: Artículo

Tipo de soporte: Revista

Enrique Cabello; M. Araceli Sánchez; Julian Nieto; Jesús M. Berrocal; Guido Castro and Javier Delgado. "A real-time vision system for on-line rock size estimation". pp. 86 - 91. International Institute for Advanced Studies,1999.

Tipo de producción: Capítulos de libros

Tipo de soporte: Libro

A. Sánchez; J. Nieto; J. M. Berrocal; G. Castro; E. Cabello; J. Delgado. "Una aplicación de visión artificial para medir el tamaño de rocas en tiempo real". Canteras y Explotaciones. 380, pp. 52 - 57.1999.

Tipo de producción: Artículo

Tipo de soporte: Revista

J. Nieto; J. M. Berrocal; A. Sánchez; G. Castro; and Enrique Cabello.A real time computer vision system for rock size estimation". La Lettre de l'IA.134 - 135-136, pp. 282 - 284.EC2, 1998.

Tipo de producción: Artículo

Tipo de soporte: Revista

Convolutional neural network approach for multispectral facial presentation attack detection in automated border control systems

Araceli Sánchez-Sánchez, M., Conde, C., Gómez-Ayllón, B.,.Palacios-Alonso, D., Cabello, E.

Entropy, 2020, 22(11), pp. 1–18, 1296

Congresos:

Título: Adaptación de modelos geométricos a imágenes: aplicación a caras humanas

Nombre del congreso: XX Jornadas de Automática.

Ciudad de realización: Salamanca, España

Fecha de realización: 09/1999

Javier Gómez; M. Araceli Sánchez; Enrique Cabello.09/1999.

Título: Supervised methods for face recognition using geometric characteristics.

Nombre del congreso: IASTED International Conference on Signal Processing and Communications Sponsored by IASTED, ULPGC, IAC, IEEE.

Ciudad de realización: Canarias, España

Fecha de realización: 02/1998

Ciudad: España

Enrique Cabello; M. Araceli Sánchez; Angel Luis Labajo; Luis Pastor; Juan Alonso.02/1998.

Título: Modelos conexionistas para el reconocimiento de caras.

Nombre del congreso: IX Congreso de la sociedad española de psicología comparada.

Ciudad de realización: Salamanca, España

Fecha de realización: 09/1997

Enrique Cabello; Araceli Sánchez; Luis Pastor.09/1997.

Título: Reconocimiento de caras humanas: una aproximación por medio de redes neuronales.

Nombre del congreso: VII RPIC (Reunión de Trabajo en Procesamiento de la Información y Control). **Fecha de realización:** 09/1997

Araceli Sánchez; Enrique Cabello; Luis Pastor. Lugar: San Juan; Argentina.09/1997.

Título: Automatic face recognition using neural networks: gray level images versus geometric characteristics.

Nombre del congreso: 15 IMACS WORLD CONGRESS on Scientific Computation, Modelling and Applied Mathematics. Sponsored by IMACS, DFG, IEEE, IFAC, IFIP, IFORS, IMEKO.

Ciudad de realización: Berlín, Alemania

Fecha de realización: 08/1997

Enrique Cabello; M. Araceli Sánchez; Luis Pastor; Juan Alonso.08/1997.

Título: Automatic face recognition using neural networks: gray level images versus geometric characteristics.

Nombre del congreso: FACE RECOGNITION. **Ciudad de realización:** Sirling, Reino Unido

Fecha de realización: 07/1997

Entidad organizadora: OTAN

Enrique Cabello; M. Araceli Sánchez; Luis Pastor; Juan Alonso.07/1997.

Título: Una red neuronal con Ada95, aplicación al reconocimiento de caras humanas.

Nombre del congreso: Jornadas Técnicas de Ada Spain.

Ciudad de realización: Madrid, España

Fecha de realización: 02/1997

Araceli Sánchez; Enrique Cabello; Luis Pastor.02/1997.

Título: Procesamiento Digital de Imágenes: aplicación de redes neuronales al reconocimiento de caras humanas.

Nombre del congreso: XIV Congreso de la Sociedad Española de Ingeniería Biomédica

Ciudad de realización: Navarra, España

Fecha de realización: 09/1996

Enrique Cabello; Araceli Sánchez; Luis Pastor.09/1996.

Título: Un sistema de visión para detectar y estimar el tamaño de rocas.

Nombre del congreso: XX Jornadas de Automática.

Ciudad de realización: Salamanca, España

Fecha de realización: 09/1999

M. Araceli Sánchez; Julian Nieto; Jesús M. Berrocal; Guido Castro; Enrique Cabello; Javier Delgado.09/1999.

Título: A real-time vision system for on-line rock size estimation.

Nombre del congreso: Special session on Advanced Concepts for Intelligent Vision Systems. (XI International Conference on Systems Research, Informatics and Cybernetics).

Ciudad de realización: Baden-Baden, Alemania

Fecha de realización: 08/1999

Enrique Cabello; M. Araceli Sánchez; Julián Nieto; Jesús M. Berrocal; Guido Castro and Javier Delgado.08/1999.

Título: A real time computer vision system for rock size estimation.

Nombre del congreso: EC2 and Developpement.

Ciudad de realización: Nimes, Francia

Fecha de realización: 05/1998

J. Nieto; J. M. Berrocal; A. Sánchez; Guido Castro; Enrique Cabello; Francia.05/1998.

Título: Multiple object detection and tracking in a non-constrained environment.

Nombre del congreso: II Workshop Hispano Luso de Agentes Físicos.

Ciudad de realización: Móstoles, España

Fecha de realización: 03/2001

Ciudad: España

Antonio Sanz; Iñigo Martín; M. Araceli Sánchez; Enrique Cabello.03/2001.

Título: Artificial Vision for Road Safety Improvement

Nombre del congreso: II European Congress on Intelligent Transport Systems

Ciudad de realización: Bilbao, España

Fecha de realización: 06/2001

Enrique Cabello Pardos; M. Araceli Sánchez Sánchez; Laura Agudo Mérida; Antonio Sanz Montemayor; Iñigo Martín Sánchez.06/2001.

BIBLIOGRAFÍA

Bibliografía Capítulo 2

[1] Galton, F. Numeralized Profiles for Classification and Recognition. *Nature* **1910**, *83*, 127-130.

[2] Chellappa, W.; Sirohey. Human and Machine Recognition of Faces: A Survey. *Proc. of the IEEE* **1995**, *88*.

[3] Ashok Samal and Prasana A. Iyengar. Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey. *Pattern Recognition* **1992**, *25*, 65 - 77.

[4] Kanade, T. Picture Processing by Computer Complex and Recognition of Human Face. *PhD Thesis*, **1973** Kioto University.

[5] Brunelli, R.; Poggio, T. Face Recognition Features Versus Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1993** *15*, 1042 - 1052.

[6] Cox, I.J.; Ghosn, J.; Yianilos, P.N. Feature - Based Face Recognition Using Mixture Distances. *Technical Report, NEC Research Institute*, **1995** Princeton, NJ.

[7] Burt, P. Smart Sensing within a Pyramid Vision Machine, *Proc. IEEE*, **1998**, *76*.

[8] Yuille, A. L.; Cohen; D. S.; Hallinan P. W. Feature Extraction From Faces Using Deformable Templates. *Proc. CVPR*, **1989**, San Diego, CA.

[9] Turk, M.; Pentland, A. Eigenfaces for Recognition. *J. of Congnitive Neuroscience*, **1991**, *3*, 71 – 86.

[10] Turk, M.; Pentland, A. Face Recognition Using Eigenfaces. *Procedings Computer Vision and Pattern Recognition 91*, 586 - 591.

[11] Jian, X.; Nixon, M.S. Extending the Feature Vector for Automatic Face Recognition. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, **1995**, *17*, 1167 - 1176.

[12] Pentland, A.; Moghaddam, B.; Starner, T. View-Based and Modular Eigenspaces for Face Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, **1994**.

[13] Pentland, A.;Starner, T.; Etcoff, N.; Masoiu, A.;Oliyide, O.; Turk, M. Experiments with Eigenfaces. *Looking at People Workshop, International Joint Conference on Artificial Intelligence*, **1993**. Chamberry, France.

[14] Moghaddam, B.; A. Pentland, A. Face Recognition using View-Based and Modular Eigenspaces. *Automatic Systems for the Identification and Inspection of Humans, SPIE*, **1994**, *2257*.

[15] Marr, D. *Vision*, **1982**, W.H. Freeman, San Francisco.

[16] Perret; Rolls; Caan. Visual Neurones Responsive to Face in the Monkey Temporal Cortex. *Experimental Brain Research*, **1982**, *47*, 329 - 342.

[17] Lawrence, S.; Lee Giles, C.; Tsoi, A.C.; Back, A.D. Face Recognition: A Hibrid Neural Network Approach. *Technical Report. CS – TR*, **1996**, *3608*.

[18] Fleming, M.; Cottrell, G. Categorization of Faces Using Unsupervised Feature Extraction. *Proc. ISCNN 2*, 90.

[19] Kohonen, T.; Lehtio, P. Storage and Processing of Information in Distributed Associative Memory Systems, *in G. E. Hinton and J.A. Anderson, Parallel Models of Associative Memory, Hillsdale*, **1981**, NJ, Lawrence Erlbam Asociates, 105 - 143.

[20] Stonham, T.J. Practical Face Recognition and Verification with WISARD, *in H. Ellis, M. Jeeves, F. Newcombe and A. Young (eds), Aspects of Face Processing, Martinus Nijhoff Publishers*, **1986**, Dordrecht.

[21] Facial Data Base. University of Bern (iamftp.unibe.ch). Switzerland, **1995**.

[22] Rumelhart; McClelland, D.J. PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, **1986**, MIT Press/Bradford Books.

[23] LeCun, Y.; Boser, B.; Denker, J.; Hendris, D.; Howard, R.; Hubbard, W.; Jackel, L. Handwritten Digit Recognition with a Back - Propagation Network. *Advances in Neural Information Procession Systems II*, **1990**, Morgan Kauffmann.

[24] Bueno Sanz, A. Mecanismo Atencional Multinivel para la Correspondencia de Imágenes, **1995**, Grado de Salamanca.

[25] Koepfler, G.; López Ch.; Morel, M. A Multiscale Algorithm for Image Segmentation by Variational Method. *Internal Report num. 9253.* **1992**, CEREMADE. Université de Paris IX - Dauphine.

[26] Bueno Sanz, A.; Cabello Pardos, E. A Matching Algorithm for a Stereo System Simulating Human Vision. *Proc. de las conferencias del XIV IASTED Internacional*, **1996**, Insbruck, Austria.

[27] Reppas, J.B.; Dale, A.M.; Sereno, M.I.; Tootell, R.B.H. La Visión, Una Percepción Subjetiva. **1996**, Mundo Científico.

[28] Institute for Parallel and Distributed High Performance Systems. Manual SNNS. University of Stuttgart. Report No. 3/94.

[29] Carter, J.R.; Sanden, B.I. Ada Design of a Neural Network. *ACM Ada Letters*, **1994**, *14*, 61 - 73.

[30] Barnes, J. Programming in Ada95. **1995**, Addison - Wesley.

[31] Barnes, J. G. P. Programación en Ada. **1987**, Ediciones Díaz de Santos, S.A. Addison - Wesley.

[32] LVQ Programming Team. The Lvq Program Package. University of Technology. Helsinki. **1995**.

[33] Kohonen, T. Self - Organizing Maps, **1995** Ed. Springer.

[34] Cabello Pardos, E.; Sánchez Sánchez, M.A.; Pastor Pérez, L. Procesamiento Digital de Imágenes: Aplicación de Redes Neuronales al Reconocimiento de

Caras Humanas. **1996**, XIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica. Pamplona.

[35] Cabello Pardos, E.; Sánchez Sánchez, M.A.; Pastor Pérez, L. Una Red Neuronal con Ada95. Aplicación al Reconocimiento de Caras Humanas, **1997** Jornadas Técnicas de AdaSpain. Madrid.

Bibliografía Capítulo 3

[1] Corke, P.R.; Winstanley, G.J. Vision based control for mining automation. *IEEE Robotics and Automation Magazine* **1998**, *5*, 44–49.

[2] Wang, W.X.; Stephansson, O. Automatic selection of fragment images from a moving conveyor belt. *Mining Engineering*. **1996,** 83–88.

[3] Wu, X.; Devgan, A.; Hagaman, R.; Kemeny, J.M. Analysis of rock fragmentation using digital image processing. *Journal of Geotechnical Engineering*, **1993**, *119* 1144–1160.

[4] Crida, R.C.; De Jager, G. Rock recognition using feature classification. *Proceedings of the IEEE South African Symposium on Communications and Signal Processing October* **1994**, 152–157.

[5] Crida, R.C.; De Jager, G. Multiscalar rock recognition using active vision. *Proceedings of the IEEE International Conference on Image Processing*, **1996**.

[6] Crida, R.C.; De Jager, G. Rock detection using neural networks. *Proceedings of the Third South African Workshop on Pattern Recognition*, **1992**.

[7] Fernandez, R.; Viennet, E.; Goles, E.; Barrientos, R.; Telias, M. On-line coarse ore granulometric analyzer using neural networks. *Proceeding of ICANN* **1995**.

[8] Russ, J.C. *The Image Processing Handbook*. **1995** CRC Press.

[9] Baxes, G.A. *Digital Image Processing Principles and Applications*. **1994** John Wiley.

[10] Tsai, W.H. Moment preserving threshold: A new approach. *Computer Vision Graphics and Image processing* **1985**, *29*, 377–393.

[11] Otsu, N. A threshold selection method for gray- level histograms. *IEEE Transactions on System, Man And Cybernetics 9*, **1979**, 62–66.

[12] Jain, A.K. *Fundamentals of Digital Image Processing*. **1989** Prentice Hall.

[13] Sahoo, P.K.: Soltani, S.; Wong, A.K.C.; Chen, Y.C. A survey of thresholding techniques. *Computer Vision Graphics Image Process*. **1988**, *41*, 233–260.

[14] Lee, S.U.; Chung, S.Y.;Park, R.H. A comparative performance study of several global thresholding techniques for segmentation. *Computer Vision Graphics Image Process*. *52*, **1990**, 171–190.

[15] Hannah, I.; Patel, D.; Davies, R. The use of variance and entropic thresholding methods for image segmentation. *Pattern Recognition* **1995**, *28*, 1135–1143.

[16] Vernon, D. *Machine Vision*. **1991** Prentice Hall.

[17] Ripley, B.D. *Pattern Recognition and Neural Networks*. **1996** Cambridge University Press.

[18] Bishop, C.M. *Neural Networks for Pattern Recognition.* **1995** Oxford University Press.

[19] Looney, C.G. *Pattern Recognition Using Neural Networks*. **1997** Oxford University Press.

[20] Jolliffe, I.T. *Principal Component Analysis*. **1986** Springer-Verlag.

[21] Fukunaga, K. *Statistical Pattern Recognition*. **1989** New York: Academic Press.

[22] Daugman, J.D. Complete discrete 2-D Gabor transforms by Neural networks for image analysis and compression. *IEEE Trans. On Acoustics, Speech and Signal Processing 36* 1169–1179.

Bibliografía Capítulo 4

[1] Delac, K.; Grgic, M. A survey of biometric recognition methods. *Proceedings of the Elmar-2004, 46th International Symposium on Electronics in Marine*, Zadar, Croatia, **2004**; 184–193.

[2] del Campo, D.O.; Conde, C.; Serrano, Á.; de Diego, I.M.; Cabello, E. Face Recognition-based Presentation Attack Detection in a Two-step Segregated Automated Border Control e-Gate—Results of a Pilot Experience at Adolfo Suárez Madrid-Barajas Airport. *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications*. **2017**, *4*, 129–138 SECRYPT, (ICETE 2017), Madrid, Spain,.

[3] Robertson, J.J.; Guest, R.M.; Elliott, S.J.; O'Connor, K. A Framework for Biometric and Interaction Performance Assessment of Automated Border Control Processes. *IEEE Trans. Hum. Mach. Syst.* **2017**, *47*, 983–993.

[4] Sanchez del Rio, J.; Moctezuma, D.; Conde, C.; Martin de Diego, I.; Cabello, E. Automated border control e-gates and facial recognition systems. *Comput. Secur.* **2016**, *62*, 49–72.

[5] Labati, R.; Genovese, A.; Muñoz, E.; Piuri, V.; Scotti, F.; Sforza, G. *Automated Border Control Systems: Biometric Challenges and Research Trends* **2015**, *9478*, 11–20, Springer: Berlin, Germany.

[6] Frontex. *Best Practice Operational Guidelines for Automated Border Control (ABC) Systems*; Technical Report; Frontex: Warsaw, Poland, **2016**.

[7] Frontex. *Best Practice Technical Guidelines for Automated Border Control (ABC) Systems*; Technical Report; Frontex: Warsaw, Poland, **2016**.

[8] ABC4EU. Automated Border Control Gates for Europe Project, 2014–2018. *European Union's Seventh Framework Programme for Research, Technological Development and Demonstration under Grant Agreement No 312797*; **2020** ABC4EU: Geneva, Switzerland,

[9] BIOinPAD. Bio-inspired face recognition from multiple viewpoints. Evaluation in a presentation attack detection environment Project, 2016–2020. In *Funded by*

Spanish National Research Agency with Reference TIN2016-80644-P; BIOinPAD: **2020** Geneva, Switzerland,

[10] Anjos, A.; Komulainen, J.; Marcel, S.; Hadid, A.; Pietikäinen, M. Face anti-spoofing: Visual approach. In *Handbook of Biometric Anti-Spoofing*; Springer: Berlin, Germany, **2014**; 65–82.

[11] Galbally, J.; Marcel, S.; Fiérrez, J. Biometric Antispoofing Methods: A Survey in Face Recognition. *IEEE Access* **2014**, *2*, 1530–1552.

[12] International Organization for Standardization. *Information technology—Biometric Presentation Attack Detection—Part 1: Framework*, **2016** ISO: Geneva, Switzerland.

[13] The Management of Operational Cooperation at the External Borders of the Member States of the European Union; Fergusson, J. *Twelve Seconds to Decide: In Search of Excellence: Frontex and the Principile of Best Practice*; **2014**, Publications Office of the European Union: Brussels, Belgium,.

[14] Wen, D.; Han, H.; Jain, A. Face Spoof Detection with Image Distortion Analysis. *IEEE Trans. Inf. Forensic Secur.* **2015**, *10*, 746–761

[15] De Freitas Pereira, T.; Anjos, A.; De Martino, J.; Marcel, S. *LBP-TOP Based Countermeasure Against Face Spoofing Attacks*; Springer: Berlin, Germany, **2013**, *7728*, 121–132.

[16] Liu, Y.; Jourabloo, A.; Liu, X. Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision. *Proceedings of the IEEE Computer Vision and Pattern Recognition*, **2018**. Salt Lake City, UT, USA.

[17] Górriz, J.M.; Ramírez, J.; Ortíz, A.; Martínez-Murcia, F.J.; Segovia, F.; Suckling, J.; Leming, M.; Zhang, Y.D.; Álvarez Sánchez, J.R.; Bologna, G.; et al. Artificial intelligence within the interplay between natural and artificial computation: Advances in data science, trends and applications. *Neurocomputing* **2020**, *410*, 237–270.

[18] Gomez-Barrero, M.; Busch, C. Multi-Spectral Convolutional Neural Networks for Biometric Presentation Attack Detection. *NISK J.* **2019**, *12*, 209528032.

[19] Tolosana, R.; Gomez-Barrero, M.; Busch, C.; Ortega-Garcia, J. Biometric Presentation Attack Detection: Beyond the Visible Spectrum. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 1261–1275.

[20] Rathgeb, C.; Drozdowski, P.; Fischer, D.; Busch, C. Vulnerability Assessment and Detection of Makeup Presentation Attacks. *Proceedings of the 2020 8th International Workshop on Biometrics and Forensics (IWBF)*, **2020**, 1–6, Porto, Portugal.

[21] Ortega-Delcampo, D.; Conde, C.; Palacios-Alonso, D.; Cabello, E. Border Control Morphing Attack Detection With a Convolutional Neural Network De-Morphing Approach. *IEEE Access* **2020**, *8*, 92301–92313.

[22] Ferrara, M.; Franco, A.; Maltoni, D. Face demorphing in the presence of facial appearance variations. *Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO)*, **2018**, 2365–2369, Rome, Italy ISSN 2076-1465.

[23] Gao, W.; Cao, B.; Shan, S.; Chen, X.; Zhou, D.; Zhang, X.; Zhao, D. The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2007**, *38*, 149–161.

[24] Akhtar, Z.; Kale, S. Security Analysis of Multimodal Biometric Systems against Spoof Attacks. *Proceedings of the Advances in Computing and Communications: First International Conference*, **2011**, *191*, 604–611, Kochi, India.

[25] Kotwal, K.; Bhattacharjee, S.; Marcel, S. Multispectral Deep Embeddings as a Countermeasure to Custom Silicone Mask Presentation Attacks. *IEEE Trans. Biom. Behav. Identity Sci.* **2019**, *1*, 238–251.

[26] George, A. Biometric Face Presentation Attack Detection with Multi-Channel Convolutional Neural Network. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 42–56.

[27] Lai, C.; Tai, C. A smart spoofing face detector by display features analysis. *Sensors* **2016**, *16*, 1136.

[28] Albakri, G.; Alghowinem, S. The effectiveness of depth data in liveness face authentication using 3D sensor cameras. *Sensors* **2019**, *19*, 1928.

[29] Yi, D.; Lei, Z.; Zhang, Z.; Li, S.Z. Face Anti-spoofing: Multi-spectral Approach. *Handbook of Biometric Anti-Spoofing*; Advances in Computer Vision and Pattern Recognition; Springer: Berlin, Germany, **2014**, 83–102.

[30] Zhang, Z.; Yi, D.; Lei, Z.; Li, S.Z. Face liveness detection by learning multispectral reflectance distributions. *Proceedings of the 2011 IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, **2011**, 436–441,Santa Barbara, CA, USA.

[31] Hou, Y.L.; Hao, X.; Wang, Y.; Guo, C. Multispectral face liveness detection method based on gradient features. *Opt. Eng.* **2013**, *52*, 113102.

[32] Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231.

[33] Simard, P.Y.; Steinkraus, D.; Platt, J.C. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. In Proceedings of the ICDAR, **2003**; *3*, 958–962, Edinburgh, UK.

[34] Cireşan, D.; Meier, U.; Masci, J.; Gambardella, L.; Schmidhuber, J. Flexible, High Performance Convolutional Neural Networks for Image Classification. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, **2011**, 1237–1242, Barcelona, Spain.

[35] Yang, J.; Lei, Z.; Li, S.Z. Learn Convolutional Neural Network for Face Anti-Spoofing. *arXiv* **2014**, arXiv:1408.5601.

[36] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*; **2012**; 1097–1105 ACM: New York, NY, USA.

[37] Xu, Z.; Li, S.; Deng, W. Learning temporal features using LSTM-CNN architecture for face anti-spoofing. *Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition*, **2015**; 141–145. ACPR, Kuala Lumpur, Malaysia

[38] Lucena, O.; Junior, A.; Moia, V.; Souza, R.; Valle, E.; Lotufo, R. Transfer learning using convolutional neural networks for face anti-spoofing. *International Conference Image Analysis and Recognition*; **2017**, 27–34, Springer: Berlin, Germany.

[39] Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.

[40] Ramachandra, R.; Busch, C. Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey. *ACM Comput. Surv.* **2017**, *50*, 1–37.

[41] Zhang, Z.; Yan, J.; Liu, S.; Lei, Z.; Yi, D.; Li, S.Z. Una base de datos antispoofing de cara con diversos ataques. *Proceedings of the 2012 5th IAPR International Conference on Biometrics*, **2012**, 26–31, (ICB), Nueva Delhi, India.

[42] Chingovska, I.; Anjos, A.; Marcel, S. On the Effectiveness of Local Binary Patterns in Face Anti-Spoofing. *Proceedings of the International Conference of Biometrics Special Interest Group* 2012 (BIOSIG), Darmstadt, Germany.

[43] Doc, I. *Documentos de viaje legibles por máquina, parte*; OACI: Montreal, QC, Canadá, **2006**.

[44] Equipo, T.D. Theano: Un marco de Python para el cómputo rápido de expresiones matemáticas. *arXiv* **2016**, arXiv:1605.02688.

[45] Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; **2009** Springer Science &Business Media: New York, NY, USA.

[46] Organización Internacional de Normalización. *Tecnología de la información— Detección de ataques de presentación biométrica— Parte 3: Pruebas e informes;* **2016**, ISO: Ginebra, Suiza.

[47] Zhang, S.; Liu, A.; Wan, J.; Liang, Y.; Guo, G.; Escalera, S.; Escalante, H.J.; Li, S.Z. Casia-surf: Un punto de referencia multimodal a gran escala para la lucha contra la suplantación de identidad. *Ieee Trans. Behav. Identidad Sci.* **2020**, *2*, 182–193.

Bibliografía Capítulo 5

[1] Leden, L. Pedestrian risk decrease with pedestrian flow. A case study based on data from signalised intersections, *Accid. Anal. And Prev.* **2002**, *34*, 457-464.

[2] Várhelyi, A. Drivers speed behaviour at a zebra crossing: a case study, *Accid. Anal. and Prev.* **1998**, *30*, 731-743,.

[3] Zeedyk M.S; L. Kelly, L. Behavioural observations of adult-child pairs at pedestrian crossings, *Accid. Anal. and Prev.* **2003**, *35*, 771-776.

[4] eSafety, Final Report of the eSafety Working Group on Road Safety, Final Report, European Commission, **2002**.

[5] Hydén, C. The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique, *Bulletin 70, Institute fr Trafikteknik*, **1987**, LTH, Lund,.

[6] Schneider, RJ.; Ryznar, R.M.; Khattak, A.J. An accident waiting to happen: a spatial approach to proactive pedestrian planning, *Accid. Anal. and Prev.*, **2004**, 36, 193-211.

[7] Lajunen, T.; Parker, D.; Summala, H. The Manchester driver behaviour questionnaire: a cross-cultural study, *Accid. Anal. And Prev.* **2004**, *36*, 231-238.

[8] Older, S.J.; and Spicer, B.R. Traffic conflicts: a development in accident research, *Human Factors*, **1976**, *18*, 335- 350.

[9] Perkins, S.R.; Harris, J.I. Traffic conflict characteristics: Accident potential at intersections, *Highway Research Record*, **1968**, 225, 45-143, Highway Research Board, Washington DC.

[10] Tiwari, G.; Mohan, D.; Fazio, J. Conflict analysis for prediction of fatal crash locations in mixed traffic streams, *Accid. Anal. and Prev.***1998**, *30*, 207-215.

[11] Lord, D. Analysis of pedestrian conflicts with left-turning traffic Transportation Research Record 1538, University of Toronto, **1996**.

[12] Triggs, B. Autocalibration from planar scenes, *Proc. 5th European Conf. on Computer Vision*, **1998**, *1*, 89-105, Freiburg.

[13] Khan, S.; Javed, O.; Shah, M. Tracking in Uncalibrated Cameras with Overlapping Field of View **2001,** PETS, Kauai, Hawaii

[14] Papanilolopoulos, N. Pedestrian Control al Intersections, *Intelligent Transportation Systems Institud*, **2000**, University of Minnesota,.

[15] Obolensky, N.; Erdogmus, D.; Principe J.C. An Time-Varing Kalman Filter to Moving Target Tracking, *Proc. of CONTROLO'02*, **2002**, 418-422.

[16] Jain, R.; Kasturi, R.; and Schunck, B.G. Machine Vision, **1995**, McGraw-Hill.

[17] Gonzalez, R.C.; Woods, E.E. Digital Image Processing. **1993**, Addison-Wesley.

[18] Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems, *Trans. of the ASME-Journal of Basic Engineering*, **1960**, *82*, Series D, 35-45.

[19] Hargrave, P.J. A tutorial Introduction to Kalman Filtering, *IEEE Colloquium on Kalman Filters: Introduction, Applications and Future Developments*, 1989, (Digest No.27).

[20] Almost an Accident, University of Lund (Sweden), Department for Traffic Planning and Engineering, Scenic Television.