UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR

PROYECTO FIN DE CARRERA

MEJORA DE ALGORITMOS DE RECONOCIMIENTO DE HUELLAS DACTILARES EN ENTORNOS FORENSES

Fátima García Donday

Abril 2014

Mejora de Algoritmos de Reconocimiento de Huellas Dactilares en Entornos Forenses

AUTOR: Fátima García Donday
TUTOR: Ram Prasad Krishnamoorty
PONENTE: Daniel Ramos

ATVS Grupo de Reconocimiento Biométrico
Dpto. de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid

Resumen

Este proyecto se centra en el uso de huellas dactilares como elemento de reconocimiento de personas. Basándose en su gran poder de discriminación como rasgo biométrico, en este PFC se desarrollarán mejoras sobre un sistema biométrico de extracción automática de características.

Para la realización de pruebas y experimentos sobre las mejoras realizadas en el sistema se ha adquirido una base de datos de casos forenses reales en colaboración con la Dirección General de la Guardia Civil como parte de una beca otorgada por el laboratorio de investigación biométrica ATVS. Dicha base de datos, existente sobre papel en el laboratorio de Lofoscopia del departamento de Criminalística, ha sido digitalizada y convertida a un formato conveniente para su posterior utilización.

Adicionalmente, se han desarrollado, en colaboración con otros PFC's, dos herramientas para el marcado automático de minucias sobre las imágenes de las huellas y el cálculo de relaciones de verosimilitud (LR). Se han realizado sobre estas herramientas pruebas para detección y corrección de errores. Ambas han sido utilizadas en casos reales por la DGGC.

La mejora realizada sobre el sistema consiste en la eliminación de falsas minucias detectadas fuera de la región de la huella o región de interés (ROI). Para su eliminación se ha procedido a la segmentación de las imágenes que componen la base de datos para separar la huella del fondo. La segmentación está basada en Filtros de Gabor. Tras la identificación de la ROI, se eliminan las falsas minucias.

El plan de pruebas y experimentos desarrollado hace uso de un software basado en algoritmos de códigos cilíndricos. En ellos se analizaran los resultados del sistema biométrico antes y después de la extracción de la ROI que demuestran una mejora notable del en el proceso de reconocimiento.

Palabras clave

Biometría, reconocimiento biométrico, huella dactilar, sistemas biométricos, segmentación, ROI.

Abstract

This Project focuses on the use of the fingerprint as a recognition tool for people recognition. Based on its great power of discrimination as a biometric trait, this PFC develops improvements on a biometric automatic feature extraction tool.

For the experiments that will prove the improvement made on the system, a data base has been collected. This data base is composed of real forensic images from real cases. The work of collecting this database was part of a grant from the ATVS biometric laboratory in collaboration with the Guardia Civil. The database, which existed on paper at the GC laboratory, was digitalized and converted into a proper format for later use.

Additionally, two tools have been developed in collaboration with several students. This tools allow the location of minutiae on the fingerprint images and the calculation of the likelihood ratio (LR) between a pair of fingerprints. Errors on this tools have been detected and corrected. Both of them have been used in real cases by the Guardia Civil.

The improvement developed on the system consists on the location and removal of false minutiae detected outside the region of interest (ROI) of the fingerprint image. For its removal, the fingerprint images have been applied with segmentation techniques to separate the ROI from the background of the image. This segmentation techniques are based on Gabor Filters. After the ROI identification, false minutiae have been removed from the fingerprints.

The developed test and experiments plan uses a software based on minutia cylinder code algorithms to give the scores of the comparisons made between the database users. This experiments will analyze the biometric system results before and after the ROI extraction, which show a notorious improvement on the recognition process.

Key words

Biometric, biometric recognition, fingerprint, biometric system, segmentation, region of interest.

Agradecimientos

Gracias al laboratorio ATVS, por ofrecerme la oportunidad de formar parte de su equipo. En especial a Ram por tu paciencia y apoyo, y a Daniel Ramos, por tu confianza.

Gracias al laboratorio de Lofoscopia de la Dirección General de la Guardia Civil, quienes han hecho posible este proyecto y me enseñaron tantas cosas.

Gracias a mis compañeros de universidad, entre los que no puedo dejar de mencionar a Pencho, Juanma, Raúl, Ricardo, Jaime, Eva y Berta. Gracias por tantas horas de laboratorios, clases, prácticas, comidas y alguna que otra noche de fiesta. Cada minuto ha merecido la pena por haberlo compartido con todos vosotros.

Gracias a Sandra por ser un ejemplo de esfuerzo y trabajo diario; a Sara por no pasar ni un día sin hacerme reír; a María, por tu grandísima amistad; y a Rodrigo, por tu apoyo y comprensión durante todo este tiempo. Sin cada uno de vosotros no estaría hoy aquí. Gracias por hacer de estos años sin duda los mejores.

Gracias a mis tíos, primos y en especial a mis abuelos, por todas esas vacaciones compartidas que me cargan de energías.

Gracias a mi padre por exigirme siempre el máximo y ayudarme a sacarlo; a mi madre, por tu humildad, generosidad y cariño cada día. Gracias a mi hermana Celia por tu constante esfuerzo y sacrificio; y a María por tu paciencia y dedicación a los demás. Sois un ejemplo cada día.

Gracias Luis. Por acompañarme, por hacerme ver la importancia del trabajo y motivarme a ser cada día mejor.

GRACIAS

Fátima García Donday

Abril de 2015

ÍNDICE DE CONTENIDOS

INDICE DE FIGURAS

1 Introducción

La identificación es un proceso fundamental en nuestras vidas que realizamos constantemente incluso de forma inconsciente. Por ejemplo, al identificar a nuestra familia, amigos y conocidos. Identificamos a las personas por medio de sus caras (al verles en persona o en una imagen), por la voz (al hablar por teléfono), por su forma de actuar o de moverse, etc.

¿Por qué somos capaces de llevar a cabo esta tarea? A pesar de no ser conscientes de ello, nuestro cerebro realiza una serie de operaciones de forma automática tras las cuales identifica a una persona que previamente conocíamos y teníamos almacenada en nuestro cerebro.

La identificación se usa para restringir a las personas el acceso a determinadas acciones (entrada en una instalación, acceso a una información, etc.). Esta identificación se ha realizado mediante distintas técnicas: tarjetas personales, llaves, claves… en resumen objetos que asumen que quien los posee es la persona autorizada para su uso. Por desgracia esto no siempre es así, lo cual los convierte en sistemas muy vulnerables frente a ataques. Por tanto surge la necesidad de encontrar técnicas de identificación inmunes a este tipo de fraude. De esta necesidad nace la idea de la biometría.

¿Cómo hacer que una persona posea su permiso de acceso a un área restringida, y sea intransferible e imposible de robar? Utilizando las características inherentes de una persona. Esa operación que lleva a cabo nuestro cerebro de forma involuntaria, es lo que pretende realizar la identificación biométrica. Los rasgos a través de los cuales se lleva a cabo la identificación biométrica se denominan "rasgos biométricos".

Por tanto el objetivo de la biometría es encontrar rasgos biométricos que idealmente sean: únicos para cada persona e imposibles de falsificar.

El rápido avance de las tecnologías ha hecho que se produzcan grandes logros en el

ámbito de la identificación biométrica. Sin embargo, al mismo tiempo, también se producen nuevos avances en las técnicas de falsificación por lo que es necesario asumir la existencia de casos de impostores y tratar de detectarlos lo antes posible para desarrollar las técnicas necesarias para su eliminación.

De entre los rasgos biométricos más utilizados en la actualidad, la huella dactilar cobra especial importancia gracias a su alta eficiencia y poder de discriminación como método identificativo, su reducido tamaño, lo que permite que los sistemas de reconocimiento basados en huella sean fácilmente integrables, su bajo coste, su relativamente sencillo funcionamiento y su probada eficacia. La huella dactilar es, de hecho, el rasgo biométrico con mayor ocupación de mercado en la actualidad.

1.1 Motivación

El uso práctico de huellas dactilares como método de identificación de personas ha sido utilizado desde finales del siglo XIX cuando Sir Francis Galton [1] definió los axiomas básicos del reconocimiento dactilar donde se identificaban algunos de los puntos o características desde las cuales las huellas dactilares podían ser identificadas. La huella es un rasgo biométrico altamente discriminativo, y este hecho, a pesar de ser un dato puramente empírico [2], ha sido ampliamente aceptado en todo el mundo [3].

Con el surgimiento de los ordenadores a finales de los años 60, la identificación por huella dactilar comienza su transición a la automatización, momento en el que se crean los sistemas AFIS (Automatic Fingerprint Identification System) [4]. Este proceso viene motivado por el crecimiento de las bases de datos forenses que hizo que la indexación y la comparación manual de huellas fuese cada vez más complicadas dado su volumen. El uso de minucias como medio de identificación ha sido utilizado para desarrollar la tecnología de reconocimiento automatizado de huellas dactilares. Estos sistemas son capaces de clasificar bases de datos de millones de huellas y realizar comparaciones en muy poco tiempo. Su éxito ha sido tan rotundo que actualmente la mayoría de los países poseen un AFIS [5].

En los últimos años se ha abordado desde el mundo del reconocimiento biométrico el problema de comparar dos imágenes de huellas dactilares cuando una de ellas (latente o anónima) está recogida de una escena de un crimen. Se trata de un problema muy complejo, para el que actualmente no se cuenta con una solución satisfactoria de cara al uso de estas tecnologías automáticas en escenarios de aplicación reales.

En este proyecto se pretende abordar la problemática del reconocimiento biométrico de huellas dactilares en entornos forenses, como un trabajo incremental sobre el ya realizado por el grupo de investigación ATVS. Su realización se lleva cabo en colaboración con el departamento de Identificación del Servicio de Criminalística de la Dirección General de la Guardia Civil. De trata de proporcionar a los sistemas forenses actuales, en primer lugar la agilización en el proceso de adquisición de las evidencias y cálculo de relación de verosimilitud (LR) entre huellas dactilares, y en segundo, se pretende adaptar algoritmos ya existentes en problemas controlados al entorno forense.

1.2 Objetivos

A partir de las motivaciones expuestas en la sección anterior, durante el desarrollo del proyecto se plantean alcanzar los siguientes objetivos:

- Estudio del estado del arte en biometría forense, especialmente en sistemas automáticos de reconocimiento dactilar.

- Evaluación del funcionamiento del sistema de identificación dactilar utilizando casuísticas y escenarios adaptados al trabajo diario del especialista forense, y con diversas herramientas de evaluación diferentes.

- Agilización en el proceso de adquisición de las evidencias mediante la creación de herramientas que faciliten la mecánica de la extracción de características en las huellas dactilares, así como el cálculo de LR.

- Mejora de algoritmos para la extracción de minucias mediante la aplicación de Filtros de Gabor a las imágenes de las huellas para la diferenciación de la región de interés.

- Realización de pruebas para evaluar los algoritmos de reconocimiento de huellas dactilares en condiciones forenses.

1.3 Metodología y plan de trabajo

1.3.1 Documentación previa

En primer lugar es necesario realizar un estudio del estado del arte en biometría, sistemas biométricos, reconocimiento de huella dactilar, identificación en el ámbito forense y capacidad de discriminación de los sistemas biométricos.

1.3.2 Desarrollo

Una vez están claros los conceptos previos, se comienza con la implementación de las mejoras que se van a realizar en el sistema y la posterior realización de pruebas que constaten la mejora del rendimiento del sistema.

El desarrollo al completo del proyecto incluye varias fases:

I. Creación de una base de datos de huellas dactilares forense para la realización de pruebas y experimentos.

   La adquisición de la base de datos se lleva a cabo en colaboración con la Dirección General de la Guardia Civil. La idea principal era digitalizar la base de la que disponían originalmente en su laboratorio en papel. Cada individuo de la base de datos disponía de un documento que recoge una comparación entre dos huellas (latente e impresión) pertenecientes a un mismo dedo.

   Esta base de datos está compuesta de dos tipos de imágenes de huellas:

   - Indubitadas (impresión o fingerprint): huellas obtenidas en un entorno controlado, con el usuario identificado.
   - Dubitadas (latent o fingermark): huellas sin identificar obtenidas en la escena del crimen.

   De cada individuo se tienen por tanto dos imágenes emparejadas: una de la huella dubitada, y otra de la indubitada.

   A su vez, cada par de imágenes tiene asociados los siguientes tipos de sets de minucias:

   - Ideal: set compuesto por todas las minucias que se pueden encontrar en la huella.
   - Emparejadas (matched): set de minucias numeradas donde se hace una correspondencia entre las minucias de la huella latente y las de su huella indubitada equivalente.

   La base de datos obtenida de la Guardia Civil es del tipo emparejadas, y las minucias estaban marcadas directamente en la imagen de la huella.

II. Digitalización de la base de datos (generar archivos digitales de los sets de minucias).

III. Familiarización, uso y mejora de las herramientas de adquisición de minucias utilizadas por los peritos forenses.

IV. Modificación de los archivos que componen la base de datos para su adaptación al formato necesario para la realización de mejoras en el sistema.

V. Mejora del sistema mediante la selección de minucias en la región de interés de la huella:

   - Segmentación de la imagen: debido a la baja calidad de las imágenes de las huellas indubitadas de la Base de Datos de la Guardia Civil, en ocasiones, los sistemas biométricos generan falsas minucias fuera de la región de la huella. Para solucionar este problema, para las huellas indubitadas, generaremos nuestro propio algoritmo de segmentación basado en los Filtros de Gabor para eliminar estas falsas minucias que se encuentran fuera de la región de interés, como una etapa de post procesado.

VI. Experimentos para comprobación de los resultados obtenidos: usando el extractor de minucias automático y considerando únicamente minucias con alta frecuencia de aparición para las huellas latentes y las impresiones, con los sets de minucia ideales, calcular el EER (Equal Error Rate) y el rango de exactitud de identificación tanto para los sets de minucias ideales obtenidos mediante VeriFinger, como para los obtenidos tras el procesado con Filtros de Gabor.

   El objetivo final de las pruebas es evaluar según distintos criterios qué métodos se comportan mejor para los distintos sistemas propuestos.

1.3.3 Organización de la memoria

Tras el desarrollo práctico será necesaria la creación de una memoria que contemple todos los aspectos contenidos en el proyecto.

Esta memoria consta de los siguientes capítulos:

1. **Introducción**:
   Este capítulo presenta la motivación para la realización de este proyecto y los objetivos que se persiguen durante el desarrollo del mismo.

2. **Estado del Arte:**
   Este apartado empieza con una introducción a la biometría donde se revisa la literatura al respecto, haciendo hincapié en la identificación dactilar en el ámbito forense. A continuación habla de sistemas biométricos y sus tipos, especialmente aquellos dedicados al reconocimiento de huella dactilar.

Finalmente se explican los métodos de evaluación del rendimiento de los sistemas biométricos.

3. **Diseño y Desarrollo**:

   3.1. **Herramientas de Adquisición de una Base de Datos de Casos Forenses Reales**:
   Esta sección recoge todo el trabajo realizado durante el periodo de colaboración con el departamento de lofoscopia de la DGGC. Tanto la creación de una base de datos de huellas dactilares digital a partir de los documentos existentes, como el desarrollo de mejoras para las herramientas de extracción de minucias y cálculo de relación de verosimilitud.

   3.2. **Preparación de la base de datos**:
   En esta sección se especifica el formato de cada uno de los archivos que componen la base de datos. Se pueden encontrar tres tipos diferentes de archivos, dependiendo de la información que contienen y la manera en la que han sido obtenidas estas minucias (mediante herramientas de extracción automática o de forma manual por los peritos de la DGGC).

   3.3. **Extracción de la Región de Interés**:
   Este apartado desarrolla el proceso de mejora del sistema de reconocimiento dactilar mediante la extracción de la región de interés dentro de la huella y la eliminación de falsas minucias fuera de la misma.

4. **Integración, pruebas y resultados**:
   Obtención de las puntuaciones al realizar comparaciones entre los distintos tipos de archivos de la base de datos, antes y después de las modificaciones llevadas a cabo en el sistema, de manera que se puede contrastar la mejora.

5. **Conclusiones y trabajo futuro:**
   Resumen de las principales conclusiones extraídas de la realización del proyecto en su totalidad y posibles rutas de investigación para incrementar el trabajo realizado en un futuro.

2
Estado del arte

2.1 Biometría

De igual manera que ocurre con las personas cuando identificamos a un conocido por sus rasgos faciales, forma de caminar o de hablar, la biometría trata de realizar este mismo ejercicio de reconocer a un individuo y automatizar este proceso.

Más formalmente, la biometría es la ciencia que estudia el reconocimiento de personas por sus rasgos fisiológicos o de comportamiento. Estas características se conocen como rasgos biométricos.

2.1.1 Características de los rasgos biométricos

Los rasgos biométricos se definen como aquellas características intrínsecas a una persona que la hacen única y la diferencian del resto. Pueden ser de dos tipos:

- Anatómicos: son aquellos que provienen de la biología del ser humano, genéticos y heredados.
- Rasgos de comportamiento: son aquellos que han sido adquiridos o aprendidos con el tiempo por la realización de un mismo comportamiento de manera habitual.

Los rasgos biométricos pueden ser medidos y utilizados por un sistema de reconocimiento automático para distinguir a un individuo de los demás.

Para que una determinada característica de un ser humano pueda ser considerada como rasgo biométrico, debe cumplir las siguientes premisas de forma estricta:

1. **Universalidad:** toda persona debe poseer dicho rasgo biométrico.
2. **Unicidad:** personas distintas deben poseer rasgos diferenciados.

3. **Permanencia:** el rasgo debe ser invariante con el tiempo a corto plazo.

4. **Perennidad:** el rasgo debe ser perpetuo, es decir, invariante con el tiempo a largo plazo.

5. **Mensurabilidad:** el rasgo debe poder ser caracterizado cuantitativamente.

Es evidente que no todos los rasgos biométricos cumplen de igual manera todos estos requisitos. Por ejemplo, la escritura no es un rasgo universal ya que no todo el mundo sabe escribir; la cara no es un rasgo perpetuo porque varía significativamente lo largo de la vida de una persona. Por ello, a la hora de elegir un rasgo biométrico para su utilización en reconocimiento de personas, debe evaluarse si cumple las características requeridas en función de la finalidad para la que utilice el sistema biométrico.

Adicionalmente a estas premisas, existen otras de tipo opcional:

1. **Aceptabilidad**: nivel de invasión al cual el individuo es sometido para extraer el rasgo biométrico. Debe ser muy bajo, para que presente la máxima colaboración posible por parte del individuo.

2. **Rendimiento**: precisión, fiabilidad, eficacia y velocidad de adquisición y evaluación de los rasgos.

3. **Fraude o ataques**: capacidad para ser falseado. Debe ser mínima o existir un método de veracidad del individuo.

2.1.2 Clasificación de los rasgos biométricos

Como se ha dicho anteriormente, dentro de los rasgos biométricos se diferencian dos subgrupos: los rasgos fisiológicos y los de comportamiento.

A continuación se presenta una breve introducción a los rasgos biométricos más comunes y sus principales características.

Ejemplos de rasgos fisiológicos:

Figura 1 – Ejemplo de rasgo biométrico

Huella dactilar: las crestas dactilares de los dedos y las palmas de manos y pies se forman en el séptimo mes de gestación y permanecen invariantes a lo largo de toda la vida. Esto hace de las huellas dactilares un rasgo biométrico muy atractivo para sistemas de reconocimiento. Su alto grado de aceptación hace que su uso esté muy extendido en aplicaciones comerciales, pero también en el ámbito forense, en el que se trata de identificar criminales que dejan sus huellas en la escena de un crimen. La unicidad de las huellas

dactilares está totalmente asumida pese a ser un hecho concebido a partir de **datos empíricos.**

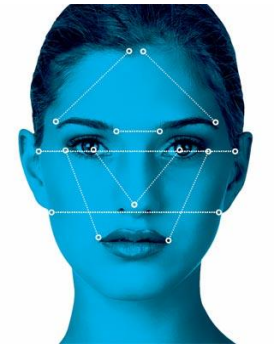

Figura 2 – Ejemplo de rasgo biométrico

Cara: es uno de los rasgos biométricos más aceptados, ya que es el más empleado por los humanos para reconocernos entre nosotros de manera visual. Además, el método empleado en la adquisición de imágenes de la cara es un método no intrusivo, lo que le permite ser una técnica con buena aceptación entre los usuarios. El reto para este tipo de aplicaciones es lograr desarrollar técnicas de reconocimiento capaces de tolerar cambios por los efectos de la edad, las expresiones faciales, ligeras variaciones en las condiciones de iluminación del entrono y variaciones en la posición de la cara con respecto a la cámara que esté captando la imagen.

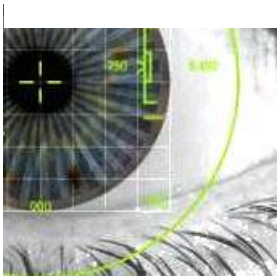

Figura 3 – Ejemplo de rasgo biométrico

Iris: el patrón de textura de cada iris es único para el individuo y se crea durante el desarrollo embrionario, manteniéndose invariante a lo largo de toda la vida. Su captura se realiza mediante imágenes, donde la iluminación y la cooperación del usuario son determinantes, por lo que es bastante sensible a las condiciones del entorno. Por todo ello se considera un método intrusivo pero con un alto potencial debido a la rapidez de los sistemas y al alto poder de discriminación que ofrece.

Ejemplos de rasgos conductuales:

Figura 5 – Ejemplo de rasgo biométrico

Firma manuscrita: el modo en que firma una persona es un rasgo de comportamiento característico de cada individuo. A lo largo de la historia, la firma ha sido el medio de identificación más común. Sin embargo, este rasgo presenta una gran variabilidad a corto y largo plazo, y un alto riesgo de falsificación por parte de otros individuos, lo cual lo convierte en un rasgo difícil para un reconocimiento automático fiable. Por otro lado al tratarse de un proceso de adquisición no intrusivo tiene un alto grado de aceptación.

Figura 4 – Ejemplo de rasgo biométrico

Forma de caminar: este no es un rasgo biométrico especialmente distintivo, pero sí lo suficientemente característico como para permitir la verificación en algunas aplicaciones de baja seguridad. Al ser un rasgo conductual puede no ser invariante, especialmente a largo plazo, por culpa de fluctuaciones importantes en el peso, o lesiones en las articulaciones o el cerebro. La adquisición de este rasgo es similar a la de fotografías faciales, por lo que puede considerarse un rasgo biométrico aceptable. No obstante,

debido a que los sistemas biométricos basados en este rasgo utilizan secuencias de video de la persona caminado para medir los diferentes movimientos de cada punto de articulación, su carga computacional es más bien alta.

Figura 6 – Ejemplo de rasgo biométrico

Voz: el poder de discriminación de la voz es una característica ampliamente reconocida. Su captura se realiza mediante un proceso no invasivo, lo que la convierte en un rasgo biométrico muy atractivo. No se considera, sin embargo, que sea lo suficientemente distintivo como para permitir la identificación de un individuo en una gran base de datos. Por otro lado, la señal de voz que está disponible para el reconocimiento de un individuo ha sufrido degradación de calidad por el micrófono, el canal de comunicación y la digitalización. La voz se ve afectada, además por la salud de la persona, su estado de ánimo y emociones. Este rasgo es en verdad una combinación de características físicas (fisionomía del tracto vocal) y de comportamiento (ritmo, acento). Estas últimas no son invariantes a lo largo del tiempo.

Además de los rasgos mencionados, existen otros estudiados en menor medida como pueden ser: la retina, la oreja, el termograma, la distribución de las venas en la mano e, incluso, rasgos que se pueden considerar menos distintivos como el olor, la forma de teclear, etc. Son los denominados "soft biometrics". Únicamente pueden proporcionar información adicional a otros rasgos y no pueden ser discriminativos.

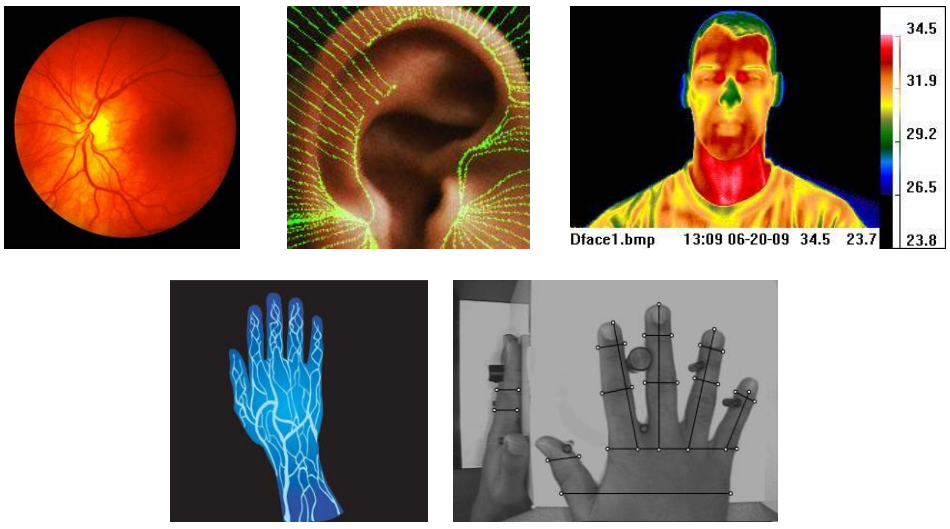

Figura 7 – Ejemplos de otros rasgos biométricos

2.1.3 La biometría en el ámbito forense

En la actualidad, la biometría se encuentra presente en la gran mayoría de las situaciones en las que se requiere la identificación de personas. Empieza a ser bastante habitual encontrarse un lector de huellas dactilares para el control de acceso a una zona restringida, encender un ordenador, etc. También nos vamos acostumbrando al manejo de dispositivos electrónicos con la voz, que aunque no siempre tiene una aplicación de identificación, está estrechamente relacionado.

A parte de la infinidad de aplicaciones comerciales que la biometría ofrece, en este caso nos centraremos en su utilidad en el ámbito forense, para lo cual empezaremos introduciendo el término de ciencia forense.

La ciencia forense se refiere a la aplicación de la ciencia o la tecnología en la investigación de actividades criminales y al establecimiento de los hechos o evidencias en un tribunal [5] [6].

La biometría es una parte muy importante de las ciencias forenses ya que permite la asociación de un individuo con una cierta información sobre una persona, o lo que es lo mismo, la identificación de una persona. Gracias a esto, los científicos forenses son capaces de recopilar información sobre la identidad de personas implicadas en crímenes [6], a través de sus rasgos físicos y de comportamiento.

Como ya se ha dicho, la biometría consiste en la extracción automática de estos rasgos característicos de cada individuo [7]. Sin embargo, el principal objetivo de la biometría forense a día de hoy no es la identificación automática de personas a partir de las pruebas de la escena del crimen, sino un filtrado automático de fuentes potenciales de estas pruebas, a partir del cual se realizarán las pruebas pertinentes de forma más exhaustiva por parte de los expertos que corresponda [5]. Es decir, la automatización de los procesos de adquisición e identificación.

En lo que se refiere a los procesos de adquisición se pretende mejorar la extracción de características, creando herramientas que agilicen el proceso además de intentar minimizar el error humano. Una parte importante de este proyecto consiste en el desarrollo de una herramienta para la extracción automática de características de las huellas dactilares.

Con respecto a la identificación, resulta útil la automatización que se puede realizar con un programa para poder reducir el tiempo de búsqueda en una base de datos. Normalmente se reduce significativamente el número de candidatos utilizando este tipo de herramientas para posteriormente realizar la verificación de forma manual. Se aumenta la fiabilidad en los resultados ya que la biometría es capaz de aportar métodos de cálculo que han sido entrenados previamente para verificar su correcto funcionamiento.

La principal diferencia entre la biometría comercial y la biometría forense radica en la

calidad de las muestras. En general, las condiciones de adquisición y la colaboración de los usuarios en un sistema comercial son mucho mejores y hacen que la variabilidad entre las marcas dejadas por un mismo individuo sea mucho menor que en los casos de biometría forense. En un caso de este último tipo, es habitual que las marcas sean dejadas accidentalmente en la escena del crimen, por lo que muchas veces son incompletas o deterioradas. En biometría forense la calidad de las marcas biométricas es mucho peor, por lo que el poder de discriminación disminuye considerablemente.

En general, el uso de la biometría forense se resume en cuatro tipos:

1. Identificación de una persona, viva o muerta.

2. Identificación de marcas de rasgos biométricos encontrados en la escena de un crimen.

3. Asociación de dos o más marcas biométricas como pertenecientes a un mismo individuo sin identificar.

4. Evaluación del apoyo que da la comparación de dos huellas a dos hipótesis sobre la culpabilidad o inocencia del sospechoso. Esta evaluación se realiza mediante el cálculo de una razón de verosimilitud, o LR.

En las dos primeras situaciones se comparan muestras conocidas con muestras provenientes de una fuente desconocida y su objetivo es asignarles una identidad. Mientras que en el tercer caso no se conoce la fuente de ninguna de las muestras, pero el objetivo es determinar si pertenecen o no a la misma persona aunque esta permanezca sin identificar.

Son muchos los rasgos biométricos utilizados en la identificación de las marcas encontradas:

- Notas manuscritas
- Huellas dactilares o palmares
- Huellas de los pies
- Manchas de sangre, saliva y otros fluidos corporales de los que se puedan extraer muestras de ADN.
- Marcas de la oreja
- Impresiones de cara en airbags de coches accidentados
- Imágenes y vídeos de individuos
- Grabaciones de voz

Además de la biometría, existen otras muchas ramas dentro del ámbito de la identificación forense, tales como la balística, que trata de identificar el arma con el

que una bala ha sido disparada, o el reconocimiento de fibras, cristales, y otros muchos materiales. En todas ellas se estudia la evidencia forense.

La evidencia forense es la relación existente entre una prueba extraída de la escena de un crimen (por ejemplo una muestra de ADN de sangre encontrada) cuya fuente se desconoce, y un material con información similar originado por una fuente identificada (muestra de sangre obtenida con el consentimiento de la persona), ambos relacionados en una misma situación.

- El primer tipo es conocido como **material recuperado, muestra o marca**, y es transferido a la escena del crimen desde una persona implicada o viceversa (por ejemplo restos de cristal en una prenda de ropa).

- El segundo tipo se conoce como **material de control**, tratándose en este caso de muestras cuyo origen o fuente es conocido. Puede ser material recuperado de la escena del crimen o directamente de un sospechoso (impresiones dactilares).

La labor del científico forense consiste en examinar ambos materiales y establecer una decisión entre dos posibles hipótesis contrarias: *¿pertenecen el material recuperado y el material de control a la misma fuente?, ¿o pertenecen a fuentes distintas?*

2.2 Identificación dactilar en ámbito forense

La identificación dactilar, o dactiloscopia, ha sido el método principalmente utilizado durante las últimas décadas para la identificación de personas no solo con fines civiles sino también policiales. El estudio comparativo de las impresiones digitales (aquellas tomadas de forma voluntaria, por personal y con material idóneos, en el departamento de policía o registro civil) y marcas de huellas (dejadas involuntariamente en el lugar del crimen) han llevado a la resolución concluyente de casos judiciales donde tales rastros fueron evidencia innegable de la presencia de un sujeto determinado en la escena del delito.

Hace ya más de 100 años que esta idea fue concebida por Alphonse Bertillon: utilizar rasgos físicos individuales con el objetivo de resolver crímenes [3][6]. En 1893 se aceptó en Reino Unido la idea de que no existen dos individuos que tengan huellas dactilares iguales. Se comenzaron a recopilar entonces huellas dactilares de criminales detenidos, para utilizarlas en su identificación y solucionar el problema de los cambios de identidad que estos realizaban continuamente con sus nombres. Además, con la comparación de estos registros con huellas anónimas encontradas en casos de delitos, las fuerzas de seguridad podían identificar al culpable si este había sido arrestado previamente. Así fue como las huellas dactilares encontraron una aplicación forense.

Desde entonces hasta ahora, la identificación dactilar forense ha sufrido cambios muy importantes. Sin embargo, gracias a los avances de la tecnología, la dactiloscopia es, a día de hoy, una de las principales técnicas de identificación forense que ayuda en la resolución de casos reales.

2.2.1 La huella dactilar: formación y clasificación

La huella dactilar se considera formada y con capacidad discriminatoria a partir del sexto mes fetal. Este patrón de valles y crestas que se ha formado permanece invariable hasta la descomposición que se produce tras el fallecimiento de la persona, a excepción de cortes, quemaduras, etc. (sin olvidar la capacidad regenerativa de la piel).

La figura formada por las crestas y valles tiene tal variedad que resultan características de cada individuo.

El patrón de la huella dactilar, o dactilograma, puede ser analizado desde tres niveles:

Nivel 1: determina la forma general del dactilograma. Para ello se localiza el núcleo y la delta. El núcleo es el punto que se encuentra más al norte de la cresta más interna de la huella [8]. La delta se corresponde con una estructura de tipo triangular, formada por tres orientaciones de crestas que divergen en un punto. Se produce por la intersección de las tres zonas de la huella dactilar: la zona basilar, la zona nuclear y la zona marginal.

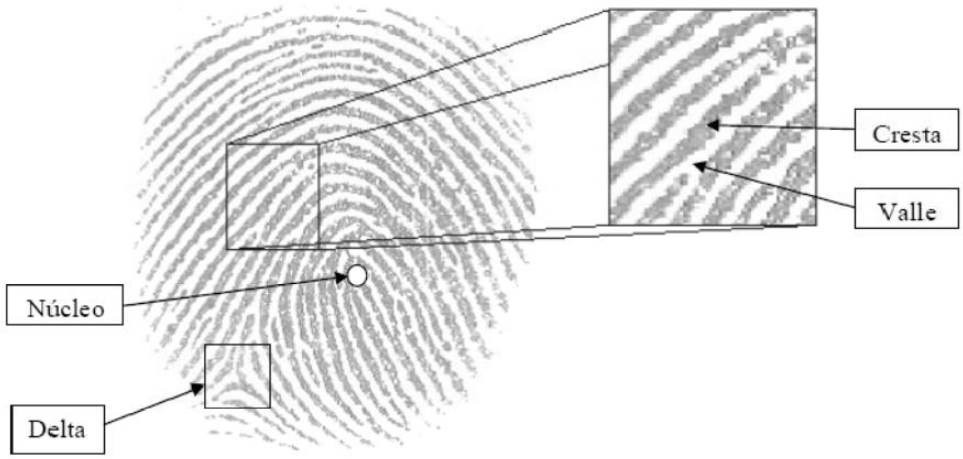

Figura 8 – Patrón de huella dactilar [25]

Según la aparición y distribución de núcleo y deltas, se tienen distintos tipos de dactilograma: monodelto (una única delta), bidelto (dos deltas), adelto (no contiene deltas), etc.

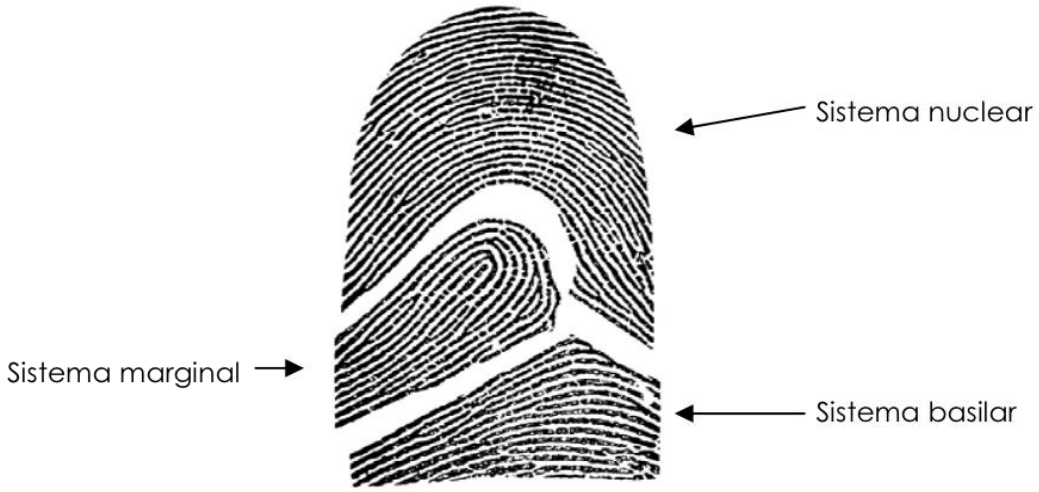

Figura 9 – Patrón de la huella NIVEL 1 [26]

El tamaño y forma de la huella y la orientación del flujo de crestas se incluyen también como características pertenecientes a este nivel.

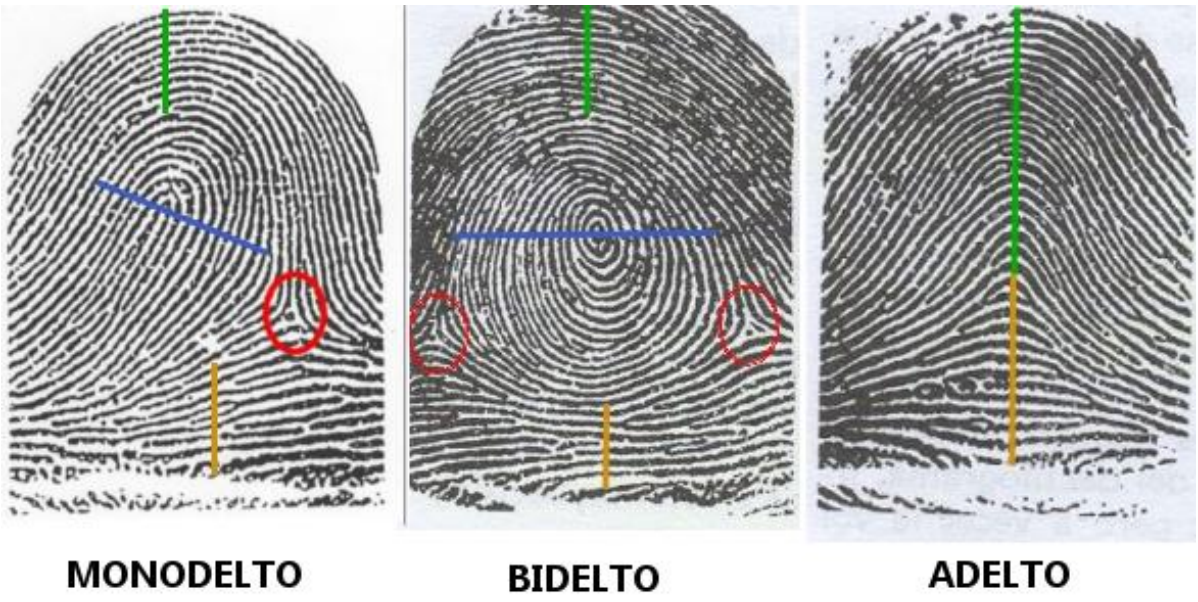

Figura 10 – Ejemplos de huellas monodelto, bidelto y adelto [24]

Nivel 2: existe un tipo de singularidades locales en las huellas denominadas minucias (minutiae, en inglés). Los tipos de minucias con mayor frecuencia de aparición son: bifurcación o convergencia (punto donde se divide una cresta en dos) y terminación abrupta (fin de una cresta). La localización de estos puntos característicos y su distribución y orientación son la clave de la unicidad de las huellas dactilares.

Figura 11 – Minucia tipo terminación abrupta

Figura 12 – Minucia tipo convergencia/bifurcación

Cada minucia se caracteriza por sus coordenadas x e y, su ángulo θ que forma la recta tangente a la cresta con el eje horizontal, y el tipo de minucia.

Nivel 3: es el más detallista y hace uso de las características internas de cada cresta como pueden ser el grosor, la ubicación de los poros de la piel dentro de estas, la forma que tienen , etc.

Figura 13 – Detalle de las características internas de las crestas

También existen distintas formas de analizar el dactilograma dependiendo del formato y la situación en el que se presente. Se puede observar directamente sobre el dedo, al natural, o se pueden plasmar por medio de distintos métodos sobre una superficie, ya sea en un entorno controlado, de forma voluntaria e intencionada para marcar la huella (impresiones dactilares) o de forma menos visible y normalmente accidental en una escena que esté siendo analizada (latente). Ambas se explican a continuación en más detalle.

2.2.2 Impresiones dactilares

Es una reproducción gráfica de la huella natural y normalmente se obtiene en un entorno controlado. Las impresiones dactilares son generalmente el material de control utilizado en ámbito forense.

Por ejemplo, cuando una persona es detenida, se le crea una reseña decadactilar que contiene todas sus huellas. Estas reseñas han sido hasta la fecha tomadas sobre una cartulina haciendo uso de tinta, aunque cada vez se hace más uso de las nuevas tecnologías de adquisición de huellas como escáneres (ya utilizado para crear el DNI por ejemplo).

Estas capturas pueden ser de dos tipos:

- Impresiones rodadas: se toma la huella por el rodamiento de cada dedo por separado sobre el papel o escáner.
- Impresiones posadas: se posa el dedo sobre la superficie de impresión. En el caso de estas se realizan los dedos índice, corazón, anular y auricular (o meñique) por un lado, y el pulgar por separado.

En ocasiones además se incluye información adicional de:

- Marca del escritor: impresión posada del canto exterior de la mano.
- Reseña palmar: impresión posada de la palma de la mano.

En cualquiera de las dos opciones de adquisición el resultado final es un archivo digital.

Tanto si es adquirida con escáner como si es plasmada en papel que posteriormente se digitaliza para pasar a formar parte de la base de reseñas.

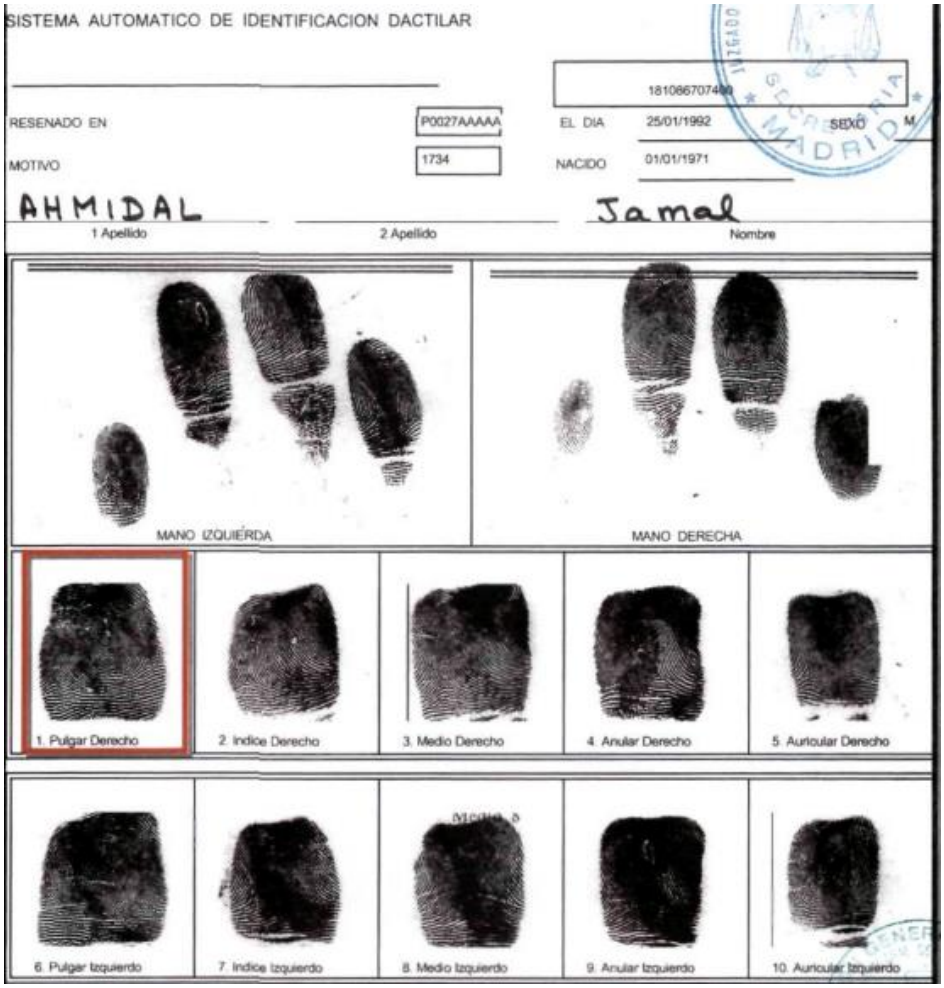

Figura 14 – Ejemplo de reseña decadactilar [29]

A pesar de que la adquisición de las impresiones dactilares se realiza en un entorno controlado debe ser supervisado por un especialista en todo momento ya que es habitual que el individuo sea un criminal por lo que su colaboración y honestidad no es asumible. Sin embargo, y gracias a este control exhaustivo, las impresiones dactilares son generalmente imágenes de una gran calidad y alta resolución. Todo esto, junto con el hecho de tener todas huellas de una misma persona por duplicado, aumenta la probabilidad de discriminación e identificación enormemente.

2.2.3 Huellas latentes

Se utiliza el término latente para las huellas anónimas encontradas en la escena del crimen y otros escenarios de interés forense. Es la impresión producida por el contacto

con una superficie, donde se queda plasmada la huella gracias a las secreciones cutáneas que se liberan por el contacto. Este tipo de huellas, que son invisibles al ojo humano, son extraídas con productos químicos y una luz especial y posteriormente reveladas para su análisis e identificación.

Son imágenes de calidad mucho peor normalmente incompletas y con artefactos, cuya fuente es desconocida. Por todo esto la identificación de la fuente de las huellas latentes es un proceso mucho más complicado que en el caso de las impresiones dactilares.

Figura 15 – Ejemplos de huellas latentes [28]

2.2.4 Identificación dactilar en ámbito forense

En el campo de la dactiloscopia, la palabra identificación es sinónimo de individualización y representa la certeza de que una marca particular fue hecha por las crestas papilares de la piel de un determinado individuo [9]. La identificación de dicha muestra se realiza mediante el análisis de las características extraídas. Por lo que el proceso de identificación es posterior al de extracción de características de la muestra, en nuestro caso, la huella dactilar.

Para poder afirmar la identificación de una huella es necesario establecer unos criterios previos que definan el protocolo de actuación. Este protocolo debe recoger un convenio común para emparejar huellas anónimas con huellas identificadas. El objetivo es conseguir que la identificación de los autores de las muestras sea justa, imparcial y ante todo correcta.

El criminalista Edmond Locard enunció la primera regla que establecía un número mínimo de minucias coincidentes necesarias para la identificación de una huella anónima. En 1911 inició un debate para crear un estándar numérico para la identificación forense de huellas dactilares, que concluyo con las siguientes reglas:

1. Si se encuentran más de 12 minucias coincidentes y la huella anónima es nítida, entonces hay identificación (en ausencia de diferencias significativas).

2. Si hay entre 8 y 12 puntos coincidentes, la confirmación de la identidad depende de:

- La nitidez de la marca
- La rareza de la huella
- La presencia de núcleo y deltas
- La presencia de poros
- El parecido entre la marca y la impresión en cuanto a la anchura de las crestas y valles, su orientación y el valor angular de las bifurcaciones.

3. Si existen menos de 8 minucias coincidentes, no se puede identificar la huella, por lo que se clasifica como no concluyente.

Estas reglas fueron ampliamente aceptadas por la comunidad dactiloscópica forense, aunque lamentablemente la tercera regla fue bastante ignorada [9].

Hoy en día el proceso de identificación de las huellas dactilares ha evolucionado bastante y se ha formalizado en un proceso de 4 pasos conocido como ACEV (Análisis–Comparación–Evaluación-Verificación). Sin embargo, el reglamento puede variar entre continentes y países. En general, el paso de evaluación puede seguir dos vertientes: vertiente del umbral cualitativo o del umbral cuantitativo.

El umbral **cualitativo** es la corriente seguida en EEUU. Esta defiende la postura de que cada proceso de identificación representa un conjunto único de circunstancias y no puede reducirse todo el problema de individualización a un simple número fijado de características coincidentes [9]. Por lo que el concepto de identificación no puede reducirse a contar minucias en las huellas.

Por otro lado, el umbral **cuantitativo** es la tendencia más común en la mayoría de los países europeos y sudamericanos. Consiste en fijar un número mínimo de minucias coincidentes entre ambas huellas para la identificación, tal y como dictan las reglas de Locard. Siguiendo este criterio, aún así hay variaciones entre el número de minucias fijado en cada país, variando entre las 7 de Rusia y las 16 de Italia, siendo 12 el umbral en la mayoría de países, incluido España.

Dado que este proyecto se ha realizado en colaboración con la DGGC el protocolo seguido para la extracción de características es el mismo que ellos siguen. Se aplica el método de reconocimiento dictado por Locard basado en minucias y se establece que debe existir un número mínimo de 12 minucias coincidentes entre huellas para que la identificación sea válida y pueda llevarse a cabo. De esta manera se excluyen aquellas huellas en las que no se localice el número mínimo de minucias de la evidencia debido a la mala calidad de la imagen o porque el fragmento que se posee sea pequeño, existiendo ciertas excepciones mencionadas anteriormente en el método.

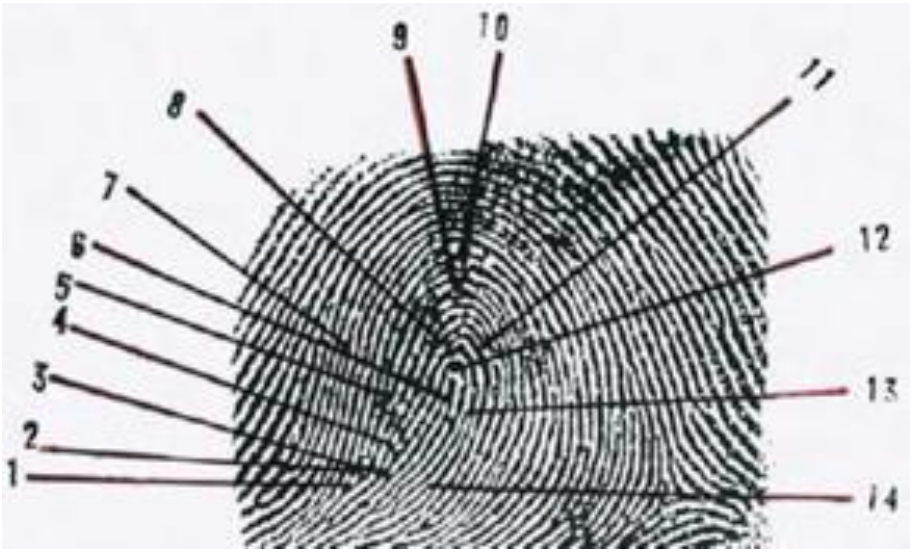

Figura 16 – Localización de minucias en una huella [27]

Existe también, como alternativa a la identificación mediante las tendencias mencionadas, la metodología LR para evaluar la evidencia forense. Los LR representan el peso de la evidencia forense en un marco probabilístico, tal y como se hace en el ámbito del análisis de ADN. Además, la metodología LR se adapta a las nuevas necesidades de la identificación forense, aplicando procedimientos científicos, repetibles, transparentes y objetivos. A pesar de estar fuera del alcance de este PFC, cabe mencionar que la metodología del cálculo de LR se encuentra actualmente en auge aunque no se trata de una forma estándar de trabajar.

2.3 *Sistemas biométricos*

Los sistemas biométricos son procedimientos que se utilizan para la identificación automática de personas mediante el uso de características físicas del individuo o de su comportamiento [7].

Por norma general, un sistema biométrico incluye las siguientes fases:

5. **Adquisición de datos:** es el proceso por el cual se recogen las muestras con la información y se digitalizan. Este paso tiene especial importancia, ya que de la forma en que se realice depende la calidad de la información y por tanto el posterior rendimiento del sistema.

6. **Pre-Procesamiento:** fase de acondicionamiento de la señal digital obtenida como muestra. Se selecciona la parte importante (dependiendo del tipo de muestra que sea, será una imagen, una grabación de voz, de video, etc.) y se eliminan ruido y otros factores no deseados.

7. **Extracción de características:** según el rasgo biométrico que se esté utilizando se extrae la información que vaya a caracterizar la muestra para representar al individuo y diferenciarlo del resto.

8. **Cálculo de similitud:** con la plantilla de características extraídas del usuario se compara el o los modelos almacenados en la base de datos y se calcula un score que cuantifica el parecido entre ambas muestras.

9. **Toma de decisión:** según el valor del score obtenido con respecto a un umbral fijado previamente, se clasifica la muestra como coincidente o no coincidente (en sistemas de verificación), o lo que es lo mismo se declara al individuo como impostor o genuino.

2.3.1 Tipos de sistemas biométricos

Los sistemas biométricos pueden ser de dos tipos según su modo de funcionamiento: de verificación o de identificación. Además de estos, existe otro modo de operación que es complementario y común a ambos, llamado modo de registro.

A continuación se explica brevemente en qué consiste cada uno de ellos.

Modo de registro: es el modo de funcionamiento en el cual el usuario es dado de alta en el sistema. Para ello es necesario que introduzca su identidad y su rasgo biométrico. A continuación se extraerán las características que se asociarán a esta identidad y se almacenarán en la base de datos del sistema. En ocasiones se puede solicitar al usuario la aportación de su rasgo biométrico más de una vez para tener en cuenta la variabilidad, lo que hace el sistema más robusto frente a posibles ataques.

Figura 17 – Funcionamiento de un sistema biométrico en modo Registro

Modo verificación: en este modo de funcionamiento el usuario introduce su nombre (o identificación en el sistema) a través de una tarjeta de identificación o similar. Tras esto los rasgos biométricos se comparan únicamente con los de un patrón previamente guardado. Este proceso implica conocer presuntamente la identidad del individuo a identificar, por lo tanto, dicho individuo ha presentado algún tipo de credencial, que después del proceso de autentificación biométrica será validada o no. Es decir, un sistema biométrico de verificación se encarga de confirmar a partir de un rasgo biométrico si un usuario es realmente quien dice ser. Por norma general, esta decisión se basa en si el resultado cuantitativo de la comparación supera o no un umbral de decisión. Como el comparador del sistema realiza una única comparación, se dice que es una comparación "uno a uno".

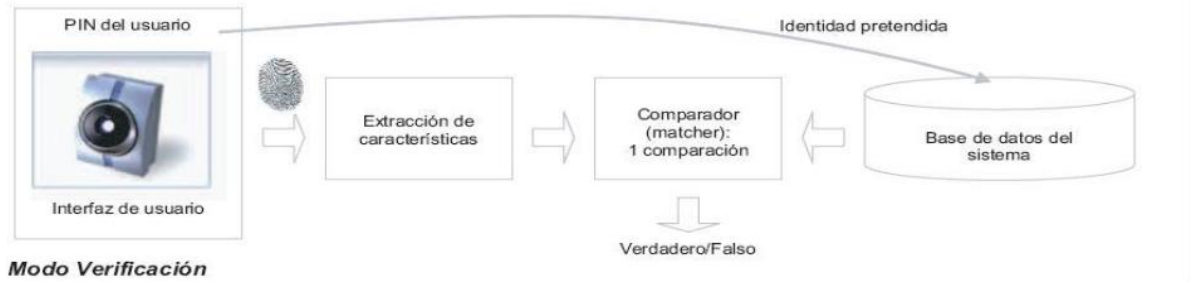

Figura 18 – Funcionamiento de un sistema biométrico en modo Verificación

Modo identificación: en este modo de operación, el usuario no reclama ninguna identidad, sino que se trata de averiguar si el individuo que solicita el acceso se encuentra o no en una base de datos previamente introducida en el sistema. A diferencia de en el modo verificación, en el que sólo se realizaba una comparación, en este caso se realizan múltiples comparaciones, ya que se debe encontrar el usuario dentro de la base de datos. Esto conlleva un gran coste computacional, sin embargo,

este modo es necesario en sistema en los que el usuario cuya identidad se busca no va a aportar información sobre su identidad, porque su objetivo precisamente es no ser identificado. Este es el caso más habitual de los sistemas forenses de reconocimiento biométrico. Este modo de trabajo se denomina "uno a muchos". La salida de un sistema de identificación puede ser o bien determinista (el usuario se encuentra o no en la base de datos) o puede ser una lista de candidatos ordenados de mayor a menor score (puntuación de similitud).

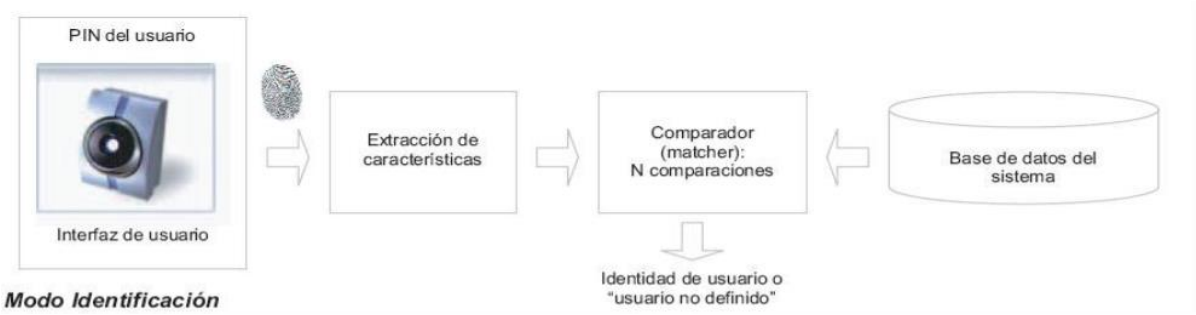

Figura 19 – Funcionamiento de un sistema biométrico en modo Identificación

2.3.2 Sistemas forenses de identificación dactilar

Como se ha explicado anteriormente, los sistemas biométricos utilizados en el ámbito forense son en su gran mayoría sistemas de identificación. La diferencia que tiene con los sistemas comerciales, es que en este caso el individuo que aporta el rasgo biométrico no desea ser identificado, por lo que no hay un nombre asociado a la muestra.

Sin embargo, la única diferencia que existe con los sistemas de verificación es el número de comparaciones que es necesario realizar antes de extraer un resultado. En el caso de los sistemas de verificación, al disponer de una identidad asociada a la muestra, sólo es necesario realizar una comparación entre las dos muestras que supuestamente pertenecen al mismo individuo para verificar que efectivamente es así. De este tipo de sistemas se obtiene un resultado de confirmación o rechazo. Pero en el caso de los sistemas de identificación es distinto, se realizan tantas comparaciones como muestras de las que se disponga en la base de datos. El resultado será una lista ordenada con el valor del score entre la huella de la que se desea obtener la identidad y el resto de muestras de la base de datos. Es decir, esta lista mostrará de mayor a menor, cuales son las identidades a las que más se parece nuestra huella; las que tienen mayor probabilidad de acierto.

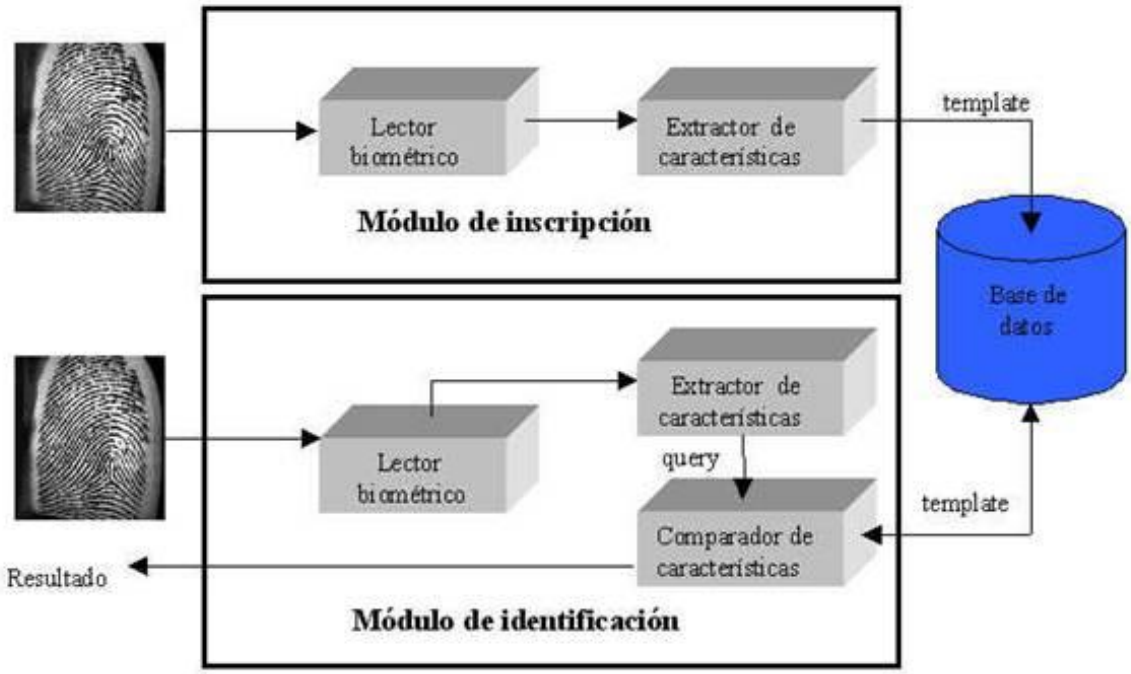

Figura 20 – Diagrama de funcionamiento de un sistema de identificación biométrico [23]

Para poder extraer un resultado al comparar dos huellas, tanto en los sistemas de verificación como en los de identificación, es necesario llevar a cabo un determinado proceso que conlleva varias tareas. En general, un sistema de reconocimiento de huella dactilar tiene dos partes diferenciadas: el extractor de características y el comparador.

Extracción de características:

La imagen de una huella dactilar es un mapa de crestas y valles papilares de la piel. Un sistema de reconocimiento dactilar compara dos huellas mediante un examen de características de las crestas y los valles para decidir si pertenecen o no a la misma fuente [10]. La extracción de minucias se realiza de la siguiente manera:

1. **Segmentación:** consiste en diferenciar dentro de la imagen el fondo de la región de interés (área de la imagen que incluye las crestas y valles de la huella). Existen distintas técnicas de segmentación como por ejemplo algunas de ellas basadas en la gran diferencia de nivel de gris que existe entre el fondo de la imagen y la huella dactilar.

2. **Estimación de la orientación de las crestas:** la orientación es calculada para cada pixel de la imagen como la dirección del flujo de las crestas alrededor de ese píxel. Esta orientación viene entonces determinada por el ángulo que forman las crestas con la horizontal y se calcula en bloques, haciendo uso de ventanas deslizantes.

3. **Extracción de crestas: mejora y binarización**. En esta fase se pretende mejorar la calidad de la imagen para resaltar las crestas de manera que sea más fácil la extracción posterior de las minucias. Se mejora la claridad de crestas y valles y finalmente se binariza la imagen para asignar dos tipos de valores: o cresta (negro) o valle (blanco).

4. **Adelgazamiento**: se reduce la anchura de las crestas a un solo píxel en la imagen, ayudando a eliminar ruido y artefactos no deseados.

5. **Extracción de minucias**: por último se seleccionan todos los puntos automáticamente en los que existe una terminación o bifurcación de las crestas. Se comprueban los pixeles negros (crestas) de forma que si sólo tienen otro píxel negro a su alrededor será una terminación abrupta, y si tiene tres, será una bifurcación.

Figura 21 – Diagrama del proceso de extracción de minucias [22]

En realidad, las características a extraer en esta fase del proceso no tienen por qué ser necesariamente minucias ya que hay numerosos tipos de características que pueden ser utilizados en el proceso de identificación.

Comparador:

Tras la extracción de las características viene la etapa de comparación o *matching*. Esta es una de las fases más críticas en el funcionamiento de un sistema de reconocimiento de huellas dactilares, y en general en cualquier sistema biométrico. La principal dificultad que tiene es la gran variabilidad que hay en la captura de la muestra. El rasgo biométrico en sí permanece invariante (tal y como indica su definición) pero la forma de capturarlo provoca que cada muestra presente ciertas diferencias. En especial en las imágenes de las huellas se puede variar el grosor de las crestas, ya que se ve alterado por la presión ejercida sobre la superficie, la orientación, desplazamiento, curvatura de la superficie, estado de la piel, y otros muchos factores.

A pesar de esto, las técnicas de comparación de los sistemas automáticos en el estado del arte tienen un alto rendimiento, dando lugar a unas tasas de error muy bajas. Sin embargo, en general dependen bastante de la calidad de las huellas. Por ello, la identificación automática con huellas latentes sigue siendo un tema que requiere mucha investigación.

En general, los sistemas de huella dactilar se dividen en tres tipos de comparadores:

I. **Comparador basado en minucias**: consiste en la comparación de patrones de minucias que han sido almacenados tras la extracción. Al hacer la comparación entre huellas, son estos patrones los que representan a cada huella y con los que realmente se realiza la comparación, dando lugar a una medida cuantitativa de similitud conocida como score.
   Antes de la fase de comparación, suelen alinearse los patrones de minucias a partir de una minucia en común para tomar un punto de referencia en ambas huellas a partir del cual empezar la comparación.
   Existen varios métodos de comparación de patrones de minucias [4][11] como por ejemplo pasar las minucias a coordenadas polares, tomando una minucia como punto de referencia para posteriormente, según un criterio fijado previamente, ordenarlas en cadenas que se van comparando entre los dos patrones.

II. **Comparador basado en textura**: en este método se utiliza el patrón del campo de orientación y de la frecuencia espacial de la imagen de la huella [10][12]. La principal ventaja de este método es su robustez frente al ruido en imágenes de baja calidad, en las que la extracción de minucias puede resultar más complicada. Sin embargo, para imágenes de buena calidad, este proceso tiene mayor tasa de error.

III. **Comparador basado en correlación**: se calcula la correlación entre las imágenes de ambas huellas. Para ello se superponen las dos imágenes y se hace la comparación de cada pareja de píxeles correspondiente. Cuando la correlación supera un umbral, se considera que ambas huellas comparadas pertenecen a la misma fuente. También existen otras técnicas para el cálculo de la correlación entre ambas imágenes como la multiplicación en el dominio de la frecuencia , aunque su coste computacional es mucho mayor debido a la necesidad de

conversión al dominio espectral. Otra opción es dividir la imagen en partes y calcular la correlación por sectores. En general son técnicas que presentan muchos problemas en cuanto las huellas no están alineadas o cuando existe deformación no lineal [10].

Hoy en día los sistemas usan combinaciones de varias de estas técnicas para mostrar una mayor robustez frente a posibles ataques.

Por tanto, un sistema forense de reconocimiento dactilar será un sistema biométrico de identificación que recibe una imagen de una huella sin identificar y devuelve, tras la extracción de características y comparación con una base, una lista de los candidatos que más puntuación han obtenido y que serán analizados posteriormente por un experto humano. Estos sistemas se conocen como AFIS (Automated Fingerprint Identification System).

En este proyecto se utiliza un comparador basado en minucias utilizando sus coordenadas cilíndricas. Por lo que la base de datos de la que partimos se centra en la localización de minucias dentro de la huella.

2.3.3 Extracción de características del S.A.I.D.

El S.A.I.D. o Sistema Automático de Identificación Dactilar, es la herramienta utilizada por el departamento de Dactiloscopia del Cuerpo Nacional de Policía para la identificación de impresiones dactilares y huellas anónimas en España. [13] Este sistema utiliza un método de reconocimiento basado en un comparador de minucias en el que se localizan las minucias de cada huella para obtener una plantilla de datos de cada muestra.

Dado que el trabajo que se ha realizado en este PFC ha sido estrechamente relacionado y orientado hacia el sistema de identificación dactilar utilizado en nuestro país, cabe destacar brevemente las peculiaridades de dicho proceso.

Para que dos huellas se consideren coincidentes, es decir, pertenecientes a una misma fuente, es necesario que sus minucias coincidan en ubicación, tipo, y cantidad. Para poder realizar esta comprobación los peritos extraen tres datos de cada minucia localizada con el fin de obtener una plantilla que englobe todos los datos extraídos.

- **Ubicación**: la ubicación de la minucia se asocia a las coordenadas donde se localiza el punto característico. Todas las minucias de una huella se extraen manteniendo un sistema de referencia para posteriormente poder compararlas entre sí. Es importante que la ubicación dentro de la huella coincida en ambas muestras, pero también es importante la distancia y posición entre minucias de una misma huella.

- **Orientación**: para cada minucia se almacena la orientación de la cresta que la contiene. La orientación identifica como el ángulo que forma la recta tangente a la cresta en el punto donde se localiza la minucia, con respecto a la horizontal. Gracias a la orientación se puede estimar la composición que adquieren las

crestas dentro de la huella.

- **Tipo**: por convenio se numeran todos los tipos de minucia conocidos. De esta manera se facilita la tarea de reconocimiento mediante el uso de un mismo código identificativo.

Actualmente existen métodos y algoritmos que se encargan de la extracción de características de forma automática, no sólo de minucias sino también de otros tipos de características propias de las huellas. Sin embargo, este proyecto no se focaliza en el proceso de automatización de la extracción, ya que los peritos todavía no se muestran muy familiarizados y convencidos de utilizar estos sistemas y por ello se sigue necesitando de la participación de un humano para estas tareas.

2.3.4 Poder de discriminación de un sistema biométrico

Se denomina poder de discriminación a la capacidad de un sistema biométrico de diferenciar comparaciones de muestras de tipo genuino, de las de impostores [14][15].

Se entiende por comparación genuina aquella que se realiza entre dos muestras biométricas pertenecientes a una misma fuente, mientras que la comparación impostora es en la que las muestras pertenecen a individuos distintos. El poder de discriminación determina la fiabilidad del sistema cuando se trabaja en modo verificación. No es otra cosa que establecer un umbral de decisión para el valor del score que resulta de la comparación, para afirmar que la muestra es genuina o por el contrario pertenece a un impostor.

Para poder valorar el poder de discriminación de un sistema se hace uso de distintos métodos para representar los resultados obtenidos con dicho sistema.

Curvas de Falsa Aceptación y Falso Rechazo

Este tipo de curvas representan la función de distribución acumulada de los scores obtenidos. Consiste en la aplicación de un umbral de decisión a partir del cual se considera que el valor del score obtenido de una comparación hace que la muestra sea genuina. Las muestras con un valor de score por debajo del umbral son impostoras. Se representa en el eje vertical la tasa de aceptación o rechazo en función del valor del umbral, representado en el eje horizontal.

- **Tasa de Falsa Aceptación (FA)**: representa la probabilidad de que una muestra perteneciente a una fuente impostora obtenga un score por encima del valor del umbral. Es decir, de que un impostor sea aceptado. Esta curva es siempre decreciente ya que cuanto mayor sea el umbral, menos impostores obtendrán un score superior.

- **Tasa de Falso Rechazo (FR):** representa la probabilidad de que una muestra genuina sea rechazada y confundida con un impostor. Esto se produce cuando

el umbral de aceptación supera el valor del score obtenido. Esta curva es creciente ya que cuanto mayor sea el umbral, más exigente es el sistema y más probabilidad hay de rechazar una muestra genuina.

Una manera de determinar el valor del umbral óptimo de un sistema es mediante el uso de estas curvas. El punto de intersección entre ambas localiza el EER (Equal Error Rate), valor del umbral para el cual se tiene el mismo valor de tasa de falsa aceptación y de falso rechazo, por lo que el sistema estaría equilibrado. El EER es una manera de resumir en un solo punto el poder de discriminación de un sistema biométrico.

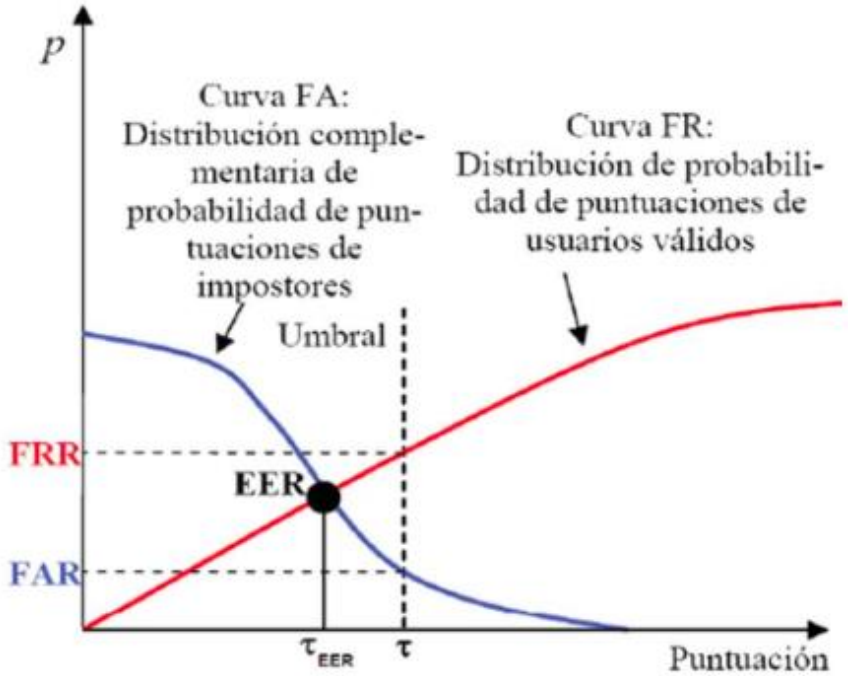

Figura 22 – Ejemplo de curvas de Falsa Aceptación y Falso Rechazo

Curva DET

También es habitual representar la densidad de probabilidad de las puntuaciones para usuarios genuinos e impostores. Fijado un umbral, el área bajo la curva de impostores que quede por encima del mismo, coincide con la probabilidad de que un impostor sea aceptado (FAR); así como el área bajo la curva de usuarios genuinos por debajo del umbral, coincide con la probabilidad de que el sistema rechace a un usuario válido (FRR).

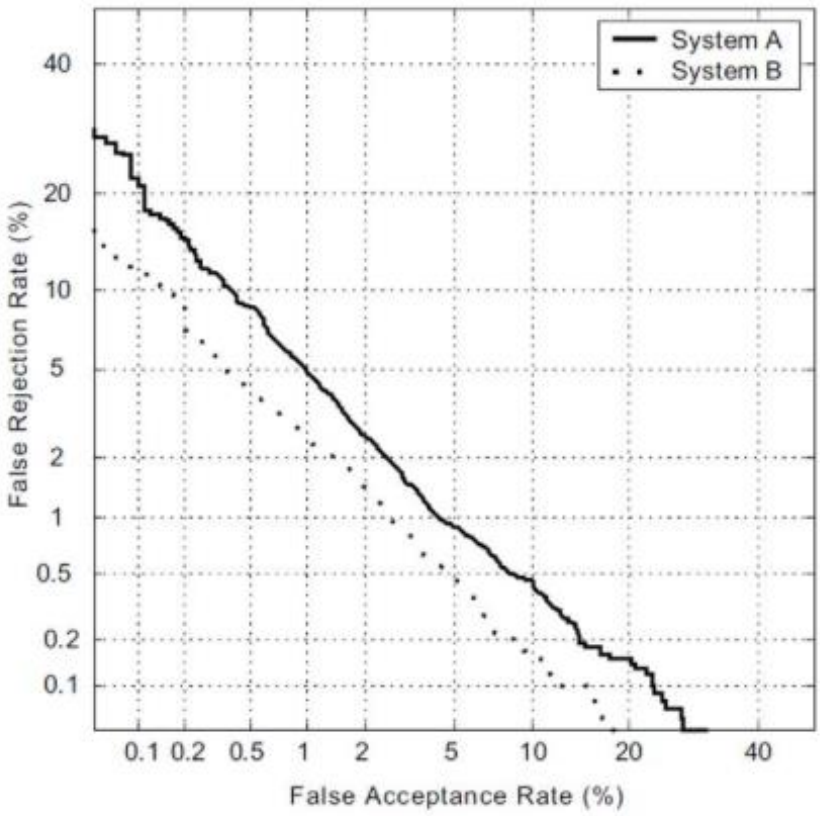

Figura 23 - Ejemplo de Curva DET

En algunas ocasiones, el punto de trabajo se fija estableciendo unos límites máximos de error de FA y FR. La representación en forma de curvas DET (Detection Error Tradeof) resulta muy útil en estos casos. En una curva DET se presenta un error frente a otro en un eje normalizado, originándose una única curva definida para todos los posibles puntos de trabajo del sistema. El EER coincide con el punto donde la curva DET corta con la bisectriz de la gráfica. Una gran ventaja de este tipo de representación es que permite comparar a simple vista distintos tipos de sistemas en cualquier punto de trabajo. Cuanto mejor sea el sistema, más se acercará su curva DET al origen (menor porcentaje de errores FA y FR).

Curva CMC

Un sistema biométrico que trabaja en modo de identificación hace una comparación uno a muchos y devuelve el modelo que mayor puntuación haya obtenido. En este caso, se mide la frecuencia con la que el modelo del usuario genuino consigue la mayor puntuación. Otras veces el sistema devuelve una lista de N candidatos que superan un cierto nivel de similitud, y se trabaja con curvas CMC (Cumulative Match Characteristic). Estas curvas indican la probabilidad con la que el candidato genuino aparece en cada posición de la lista devuelta por el sistema. En el eje horizontal se

representa el número N de candidatos de la lista que se han comprobado. En el vertical se representa el porcentaje de candidatos del sistema en los que el individuo genuino se encuentra entre los N primeros puestos de la lista.

Por ejemplo, la probabilidad de que el individuo que ha obtenido un mayor score sea realmente el genuino para el sistema A es del 84%. La probabilidad de que el genuino se encuentre entre los 8 primeros candidatos de la lista en los sistemas A y C es del 100%, es decir, el sistema puede asegurar que el candidato es el correcto reduciendo la lista a 8 individuos.

Idealmente, la curva CMC de un sistema alcanzaría el 100% de probabilidad de acierto en el 1. Lo cual indicaría que siempre encuentra el candidato genuino a la primera.

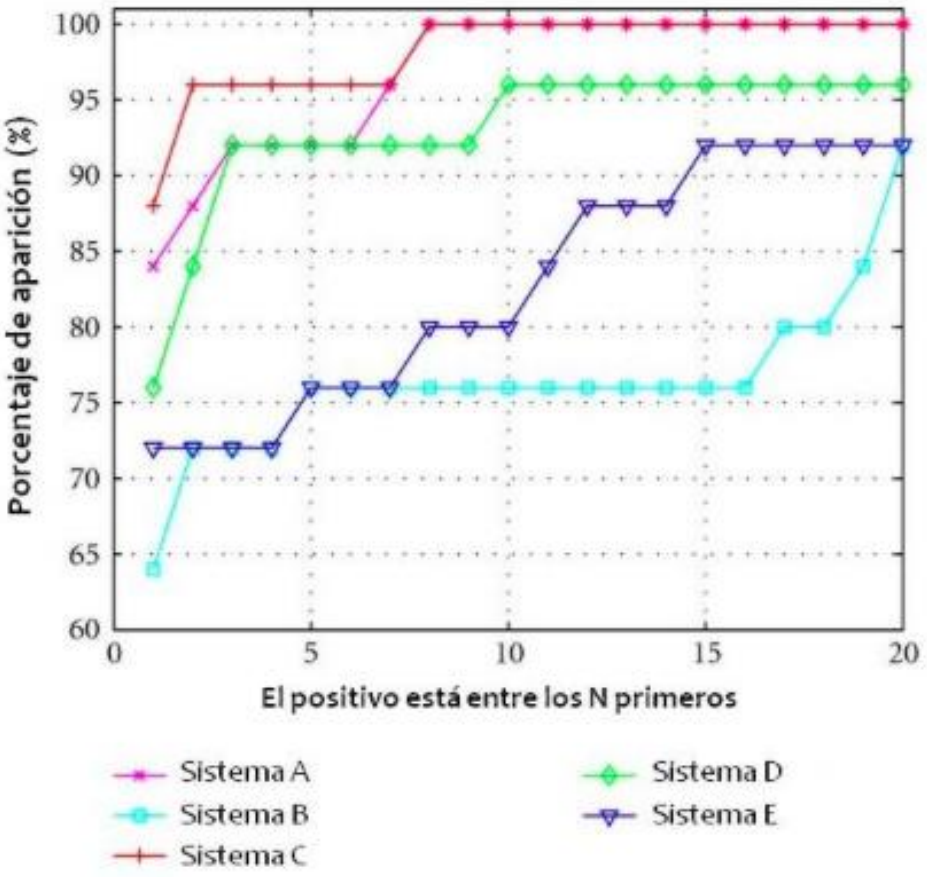

Figura 24 – Ejemplo de Curva CMC

3

Diseño y desarrollo

En este capítulo se pretende abordar dos temas:

1. En primera lugar, el proceso adquisición de una base de datos de huellas dactilares, incluyendo tanto las herramientas necesarias para su adquisición como los problemas a los que hay que enfrentarse a la hora de digitalizar una base ya disponible en papel.

2. En segundo lugar el proceso de mejora de algoritmos de identificación diseñado en este PFC y su aplicación a un sistema de comparación de patrones de minucias basado en códigos cilíndricos.

3.1 Herramientas de Adquisición de una Base de Datos de Casos Forenses Reales

Desde hace varios años, la colaboración entre la Universidad Autónoma de Madrid y la Dirección General de la Guardia Civil (en adelante DGGC), hace posible el desarrollo de nuevas tecnologías y métodos de trabajo en el entorno de criminalística. En concreto, el inicio de este proyecto se produjo gracias a una de estas colaboraciones.

La primera fase de este proyecto consistió en la adquisición de una base de datos de huellas dactilares en colaboración con el laboratorio de Lofoscopia de la DGGC. Esta idea surge por varias necesidades:

1. Disponer de una fuente de información de huella dactilar fiable para la investigación y pruebas que realiza el laboratorio ATVS de la UAM en este entorno.

2. Conversión de la información ya disponible en la DGGC a un formato digital y su

unificación.

3. Mejorar la metodología de trabajo de los peritos de la DGGC en la medida de lo posible para facilitar la extracción y comparación de minucias de la huella.

Como consecuencia de estas necesidades fue preciso en primer lugar familiarizarse con el entorno de trabajo de la DGGC para detectar las posibles mejoras. Tras este análisis se procedió a la creación de dos herramientas para la adquisición y procesado de las minucias de las huellas.

El proceso de programación de estas herramientas se solapó en el tiempo con el inicio de la adquisición de la base de datos. Al empezar a usarse, se pudieron detectar errores y hacer propuestas de mejoras que se realizaron durante todo el tiempo que duró la adquisición de la base de datos. Finalmente se consiguió disponer de dos herramientas fiables, cómodas y que en resumen, respondían a las necesidades planteadas.

A continuación se explican con mayor detalle el proceso de adquisición de la base de datos y el contenido de cada una de las herramientas. Para poder entender mejor estos apartados, se expone previamente una breve introducción con la información previa necesaria.

3.1.1 Información previa requerida

Como ya se ha explicado previamente en el apartado "Estado del Arte", las huellas dactilares están compuestas por crestas y surcos que forman un patrón irrepetible y permanente en cada individuo. Según el dibujo de este patrón, los protocolos establecidos indican cómo tomar decisiones sobre si las huellas son coincidentes o no. La manera de determinar la similitud entre dos huellas se basa en la comparación de minucias. En este proyecto utilizaremos la comparación basada en minucias.

Las minucias son las posibles singularidades que pueden suceder en una cresta. Existen dos eventos básicos (también llamados minucias de alta frecuencia de aparición), cuyas distintas combinaciones dan lugar a otros muchos tipos de eventos (minucias de baja frecuencia de aparición). Estas son:

- **Terminación abrupta**: fin de una cresta

Figura 25 – Minucia tipo Terminación Abrupta

- **Convergencia/bifurcación**: unión de dos crestas en una única o división de una cresta en dos.

Figura 26 – Minucia tipo Convergencia/Bifurcación

En este PFC se ha considerado la siguiente clasificación de tipos de minucia, sin embargo, existen más tipos de minucias no contempladas aquí:

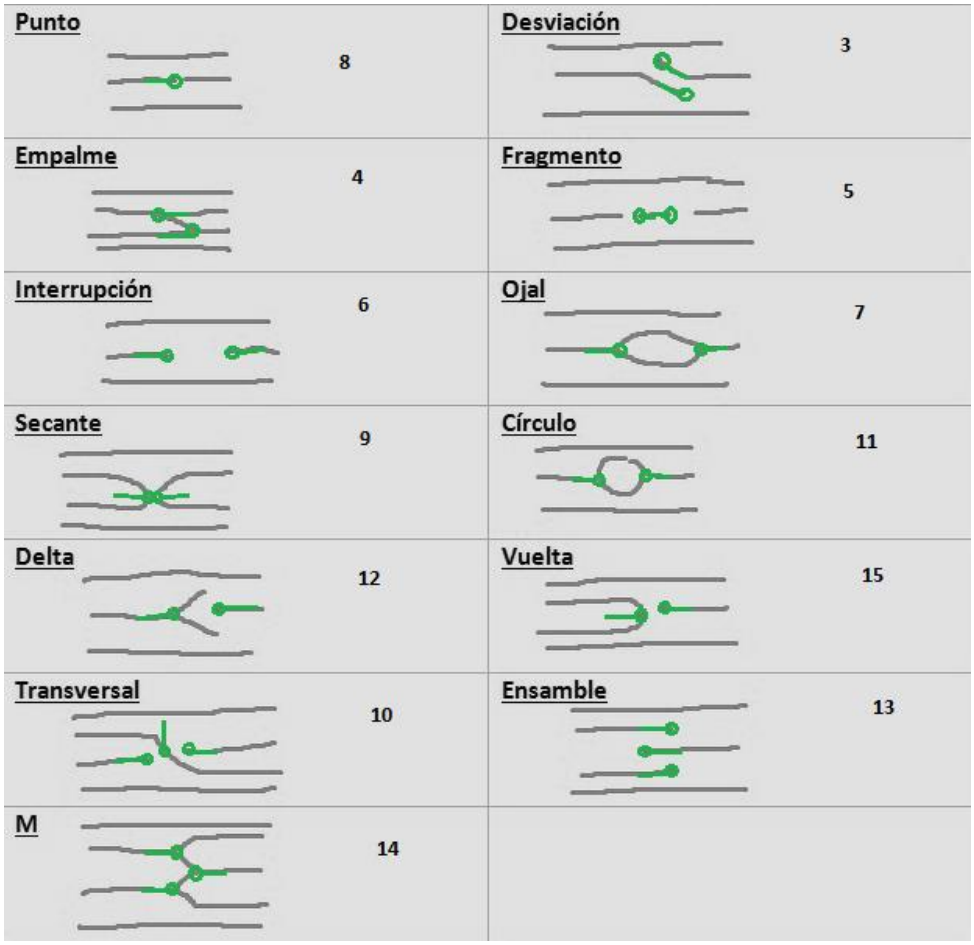

Figura 27 – Clasificación de minucias

En total se contemplan 15 tipos de minucias distintas.

La información requerida en cada huella es la localización de las minucias que la componen. Como se puede ver en la anterior imagen, en ocasiones, las minucias están compuestas por varias minucias de baja frecuencia de aparición. Por ejemplo, el ensamble, se compone de tres terminaciones abruptas localizadas de una manera muy concreta. En estos casos es necesario localizar cada una de las minucias que la

componen.

El sistema legal español determina que el número de minucias coincidentes necesario para identificar dos huellas como pertenecientes a un mismo individuo es de 12, por lo que éste será el número de minucias marcadas en las huellas de esta base de datos.

3.1.2 Adquisición de la base de datos

La adquisición de la base de datos fue un proceso que duró un curso entero con una dedicación a tiempo parcial de aproximadamente 12 horas a la semana, incluido como parte de una beca de trabajo en colaboración con la DGGC otorgada por el laboratorio de investigación biométrica ATVS. La adquisición se realizaba en el laboratorio de lofoscopia de la DGGC trabajando con los peritos.

Como ya se ha explicado, la base de datos adquirida es en realidad la digitalización de una base de datos ya existente en papel en la DGGC. Esta base de datos está compuesta por documentos llamados "cotejos". Cada uno de ellos contiene la siguiente información:

- Información administrativa para la identificación del cotejo (número de huella, fecha de adquisición, nombre/número de individuo, nombre del perito encargado de la comparación de minucias, etc.)

- Imagen huella indubitada: huella procedente de la base de datos del SAID. Estas huellas son adquiridas en un entorno controlado con el consentimiento del individuo al que pertenecen. Normalmente es una imagen en blanco y negro de calidad considerable y que presenta la totalidad de la superficie de la huella.

- Imagen huella dubitada: huella procedente de la escena del crimen tomada con una cámara. En esta imagen aparece tanto la huella como un testigo métrico para tener una referencia de su tamaño. Es una imagen adquirida en un entorno mucho más precario por lo que su calidad se ve significativamente reducida. No se asegura que aparezca la totalidad de la huella. Los colores pueden variar dependiendo de la situación y del método seguido y los productos utilizados para obtener la huella.

- Marcado de minucias: sobre cada una de estas dos imágenes el perito es el encargado de marcar las minucias de forma manual. Buscará en cada caso un par de minucias coincidentes a las que asignará el mismo número. Finalmente se tendrán al menos 12 parejas de minucias marcadas entre las dos imágenes.

- Número de minucias emparejadas encontradas.

- Resultado: se concluye si ambas huellas pertenecen al mismo individuo o no.

A parte de estos cotejos, también se disponía de las imágenes de las huellas en formato digital:

- Imagen huella indubitada: se presenta en dos posibles resoluciones, 500ppp (píxel por pulgada) o 1000ppp (originalmente las imágenes del SAID se obtenían a una resolución de 500ppp, más adelante se aumento la resolución a 1000ppp).

- Imagen huella dubitada: la resolución de estas imágenes es variable, debido a la inestabilidad del entorno en el que se adquieren.

Partiendo de esta información el proceso de conversión a un formato digital se realizó tal y como se indica en los siguientes apartados.

3.1.2.1 Pre procesado de la imagen

Previamente al proceso de localización de las minucias fue necesaria la unificación de las imágenes de la base de datos existente. Es decir, obtener parejas de imágenes lo más similares entre sí posible en lo que se refiere a su formato, así como en el conjunto de todas ellas.

Para esto se realizó un reescalado de manera que todas estuvieran en una resolución de 500ppp. El motivo de la elección de esta resolución es:

- Que la resolución de las huellas indubitadas es casi siempre de 500 o 1000ppp y al no disponer de testigo métrico es imposible reescalar estas imágenes teniendo consciencia de "cuánto" se están reescalando excepto en el caso de reducirla a la mitad cuando se encontraban a 1000ppp. En caso de tener imágenes en una resolución distinta a 500 o 1000ppp, la imagen es descartada.

- En el caso de las imágenes dubitadas, al disponer de testigo métrico, sí que es posible realizar el reescalado a cualquier resolución. Pero en cualquier caso no se aconseja aumentar la resolución de la imagen ya que se estarían añadiendo píxeles de relleno.

Por estos motivos se llegó a la conclusión de que 500ppp era la forma más óptima de unificar la base de datos de imágenes descartando el menor número de imágenes.

Este concepto de reescalado es un proceso un poco más complejo. Para entenderlo adecuadamente es necesario estar familiarizado con los siguientes términos:

- Tamaño de la imagen: viene indicado por el ancho y alto de la imagen en píxeles.

- Resolución: definida como número de píxeles por pulgada real (1 pulgada = 2.54 cm)

- Testigo métrico: requisito obligatorio en todas las imágenes de huellas dubitadas. Sin él resulta imposible conocer el tamaño real de la huella y poder

reescalar la imagen.

El reescalado es en definitiva cambiar el tamaño de la imagen en píxeles para conseguir la resolución deseada de 500 pixeles por pulgada real (por lo que es indispensable el testigo métrico).

Operación realizada:

$$T_i\{pulgadas\} = \frac{T_i\{píxeles\}}{R_i\left\{\frac{píxeles}{pulgada}\right\}}$$

$$T_f\{píxeles\} = 500\left\{\frac{píxeles}{pulgada}\right\} \cdot T_i\{pulgadas\}$$

Donde, T es tamaño, R resolución, i inicial y f final.

Esta operación se realiza tanto para el eje de ordenadas como abscisas. Y una vez calculado el tamaño final en pixeles de la imagen se reescala con dicho tamaño y ya se puede proceder a trabajar con las imágenes.

3.1.2.2 Formato final deseado

Por otro lado es importante conocer el formato al que se deseaba llegar. De cada caso se recopilaba la siguiente información:

- Imagen de la huella indubitada a 500ppp en formato ".png"
- Imagen de la huella dubitada a 500ppp en formato ".png"
- Dos archivos de texto en formato ".txt" con la información sobre las minucias contenidas en dicha huella (uno para la dubitada y otro para la indubitada). La información era almacenada por columnas separadas por espacios:

```
1 174 339 30 1
2 193 307 45 1
3 182 217 338 1
4 84 173 45 5
4 100 153 237 5
5 183 170 206 2
6 171 92 19 1
7 210 232 293 1
8 213 262 295 5
8 226 282 129 5
9 307 267 133 1
10 239 310 345 2
11 305 305 344 2
12 292 333 172 1
```

Figura 28 – Ejemplo de archivo con el formato deseado

Primera columna: número de minucia (con rango de valores de 1 a 12).

Segunda columna: coordenada en el eje de abscisas de la minucia dentro de la imagen de la huella (se considera como origen de coordenadas el vértice superior izquierdo de la imagen, con avance positivo hacia abajo y hacia la derecha, y un rango de 0 a 500 píxeles).

Tercera columna: coordenada en el eje de ordenadas de la minucia (mismo origen que en el caso anterior).

Cuarta columna: orientación de la cresta en la que está contenida la minucia (ángulo en grados, con un rango de 0 a 360).

Quinta columna: tipo de minucia (en la tabla mostrada en la imagen 27 se muestra un índice de minucias con la numeración asignada a cada tipo).

3.1.2.3 Huellas dactilares y palmares

La base de datos está compuesta en su gran mayoría por huellas dactilares, es decir, pertenecientes al área de los dedos de la mano. Tras finalizar el trabajo de conversión a formato digital de los cotejos dactilares disponibles en la DGGC, se continuó con las huellas palmares. El trabajo se realiza de forma muy similar, sin embargo, estas huellas presentan algunas peculiaridades.

De igual manera que con las dactilares, la información a extraer de cada huella son las minucias que en esta aparecen. Los tipos de minucia son los mismos. Lo que diferencia las huellas dactilares de las palmares es el área dentro de la mano con la que se trabaja.

La huella palmar tiene dos regiones de análisis: la palma y el hipotenar. Se considera la palma la región de la mano que queda plasmada al apoyar la mano boca abajo sobre

una superficie lisa, y el hipotenar situando la mano de forma perpendicular, apoyando el lateral exterior.

No existe un número fijo establecido de minucias en una huella palmar. Dado que el área de trabajo es mayor que en la huella dactilar se permite que aparezcan más de 12 minucias. Partiendo de esta premisa, el objetivo es encontrar el máximo número de minucias en común entre ambas huellas que sea posible.

El preprocesado de estas imágenes varía ligeramente con respecto a las dactilares. En el caso de las huellas dubitadas el procedimiento es exactamente igual, por lo que sigue siendo imprescindible la presencia del testigo métrico dentro de la imagen para el reescalado. La diferencia está en que para las huellas indubitadas también es necesario realizar un reescalado. Se parte de un cotejo en formato PDF en el que se presentan las huellas de ambas palmas, izquierda y derecha, y sus hipotenares. Dentro del PDF se muestra también un testigo métrico. De manera similar a las huellas dubitadas se realiza el reescalado de la imagen con la ayuda del testigo métrico. A partir de este punto todo el proceso de extracción de minucias se realiza exactamente igual que para las huellas dactilares.

Figura 29 – Ejemplo de huella palmar

3.1.3 Las herramientas de trabajo

Una vez realizado todo el trabajo de preparación de las imágenes y con los conocimientos previos necesarios, se puede proceder a la verdadera adquisición de la base de datos.

En un principio, comenzó de manera manual, sin las herramientas, ya que se encontraban en la fase final de su programación. Mientras tanto se utilizó el software gratuito GIMP. Un editor de imágenes que permitía obtener los mismos resultados aunque de manera más lenta y tediosa. Una vez terminada la herramienta de

adquisición el trabajo se hizo mucho más efectivo aumentando la producción hasta en un 120%.

3.1.3.1 Herramienta Minucia

Esta herramienta fue diseñada para el marcado de minucias en las huellas y su principal objetivo era facilitar y agilizar el trabajo de adquisición de la base de datos.

A pesar de que la programación original de la herramienta no fue una tarea perteneciente a este proyecto (ya que fue realizada por otro estudiante), sí que se utilizó para la extracción de minucias durante la adquisición de la base de datos y se realizaron mejoras en su modo operar, por lo que se considera lo suficientemente importante como para dedicarle un breve espacio a la explicación de su funcionamiento dada la presencia y gran ayuda que supuso en este PFC.

Se trata de una interfaz gráfica creada en el entorno de programación Matlab, que ofrece la posibilidad de gestionar las imágenes de la base de datos y realizar tanto el preprocesado como el marcado de minucias. Permite además exportar los ficheros ".txt" con la información correctamente almacenada en el formato explicado anteriormente.

Al iniciar la herramienta, la pantalla inicial ofrece una serie de opciones, a partir de las cuales se habilitan otras tantas.

Figura 30 – Menú de opciones de la herramienta Minucia

- **Abrir imagen**:
  Esta opción permite escoger una imagen del ordenador para cargarla en el cuadro principal. Al cargar la imagen aparecerá un cuadro de diálogo en el que se ofrece la posibilidad de reescalarla.

- **Reescalar imagen**:
  En caso de seleccionar esta opción el cursor se transformará en una cruz con la que se debe marcar el principio y el fin de una medida en la imagen elegida previamente (1cm o 1mm). Una vez hecho esto, el programa reajustará la imagen automáticamente.

- **Dibujar minucia**:
  Sirve para marcar una minucia. El cursor se transformará de nuevo en una cruz con la que se deben marcar varios puntos. El primero, la localización de la minucia (el fin o la unión de la cresta o crestas). El segundo, el punto que indica la orientación de la cresta a la que pertenece la minucia teniendo en cuenta que el ángulo almacenado será el de la línea imaginaria formada por la unión de estos dos puntos con la horizontal. Esta acción se repetirá tantas veces como

puntos posea el tipo de minucia que se esté marcando. Al finalizar el marcado de puntos, el programa hará una serie de preguntas para asignar número y tipo a la minucia. Además se puede elegir el color con el que dibujar la minucia sobre la imagen (entre azul o rojo), lo cual resulta muy útil cuando los tonos de la huella son muy similares a uno de estos dos.

- **Exportar fichero**:
  Una vez marcados todos los puntos que se deseen es posible exportar el fichero de minucias correspondiente a la imagen escogiendo el nombre y la ubicación que se quiere dar.

- **Importar fichero de minucias**:
  De igual manera, también es posible importar un fichero de minucias ya creado y mostrar las minucias dibujadas en la imagen.

- **Editar minucia**:
  Es posible, mientras se está en el proceso de marcado de minucias, editar una minucia que se quiera corregir. En este caso se seleccionará la minucia, y aparecerá un panel de botones con los que es posible cambiar tanto las coordenadas como la orientación de la minucia. También es posible eliminar la minucia por completo.

Además la herramienta cuenta con un cuadro de instrucciones en el que se va indicando en todo momento los pasos a seguir para su correcta utilización.

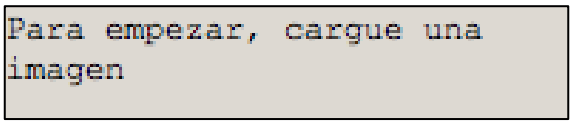

Figura 31 – Ejemplo de cuadro de instrucciones de Minucia

3.1.3.2 Aportaciones a Minucia

Como ya se ha comentado anteriormente, la programación de la herramienta Minucia no forma parte de las competencias de este proyecto, sin embargo, se pudieron ir detectando pequeños errores y posibles mejoras que podrían ayudar a su perfeccionamiento. Por esta razón una vez adquirida la base de datos, se realizó un replanteo de la herramienta y se realizaron las siguientes mejoras:

- Inclusión del cuadro de instrucciones, que indica cómo manejar la herramienta según se va utilizando, lo cual resulta muy útil ya que no es necesario conocer la herramienta antes de usarla por primera vez.

- Redistribución de los elementos dentro de la pantalla principal. Se decidió recolocar algunos de los elementos para optimizar el espacio disponible. El principal objetivo era obtener el mayor espacio posible para la imagen de la

huella con la que se trabaja.

- Opción de color en marcado de minucias. Muy útil si los colores de la imagen de la huella hacen que se camufle la minucia marcada.

- Control del correcto marcado de minucias. Impide marcar minucias fuera de los límites de la imagen de la huella, pulsar opciones no disponibles en un momento inapropiado, etc.

- Edición de minucias. Permite variar la posición y ángulo de una minucia ya guardada sin tener que eliminarla por completo y volver a dibujarla.

Todas estas aportaciones hicieron que a lo largo del tiempo la herramienta se convirtiese cada vez más en un instrumento de trabajo útil y productivo.

Figura 32 – Pantalla inicial de Minucia

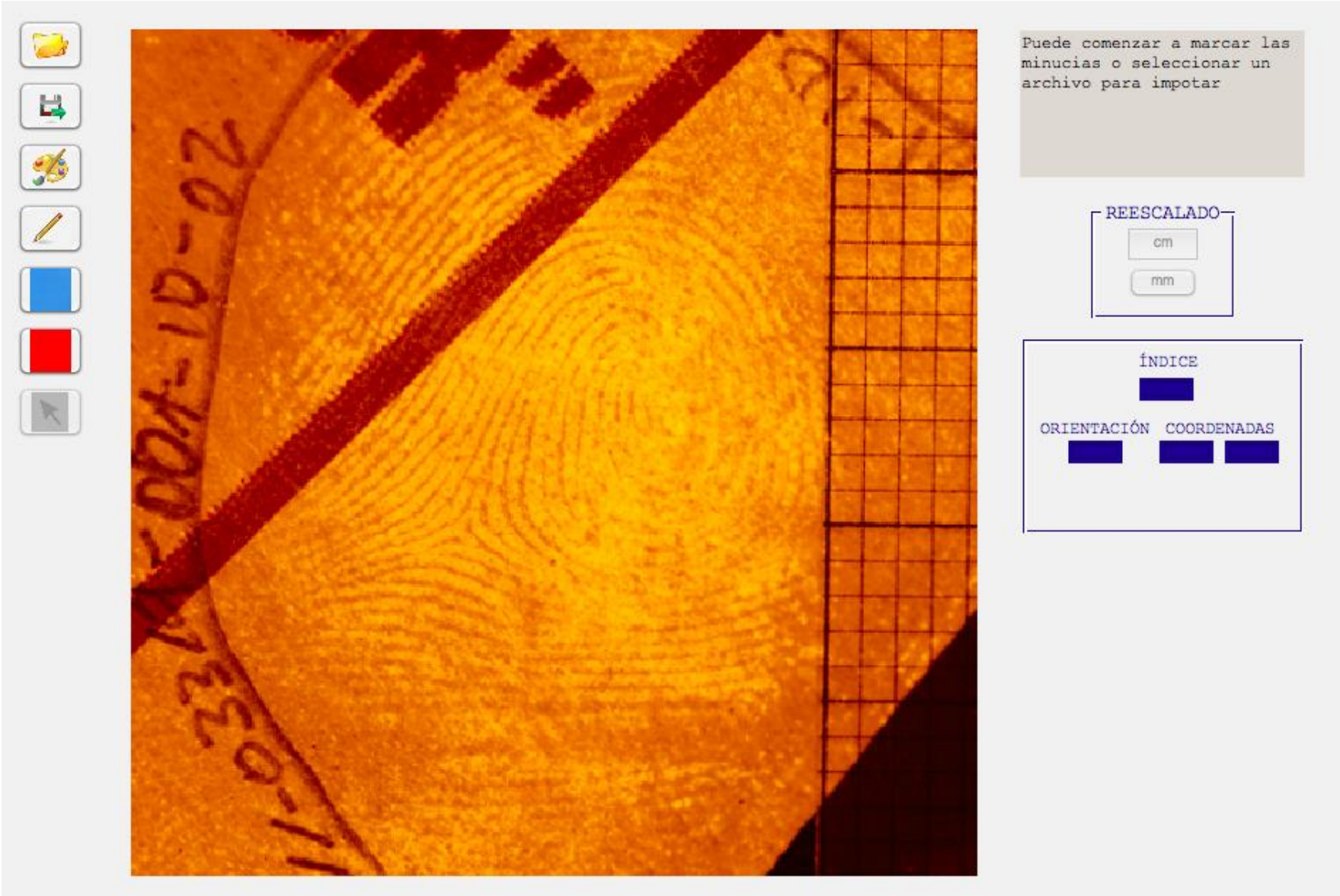

Figura 33 – Ejemplo de pantalla del proceso de extracción de minucias

3.1.3.3 Herramienta LR-Lofoscopia

La segunda herramienta, muy similar a la primera, no tuvo apenas presencia en el trabajo contemplado dentro de este proyecto. Surge como un añadido a las funciones que ya realizaba "Minucia". Dispone de dos modos de trabajo. Además de proporcionar la opción de marcado de minucias en una imagen, sirve para calcular la relación de verosimilitud de una pareja de huellas (la huella dubitada y su correspondiente indubitada).

El modo de trabajo para cálculo de LR parte de la situación de tener una pareja de huellas cargadas en el programa con sus correspondientes ficheros de minucias importados. A partir de aquí se puede acceder a la opción de cálculo de LR que devolverá una puntuación como resultado del cálculo interno que realiza.

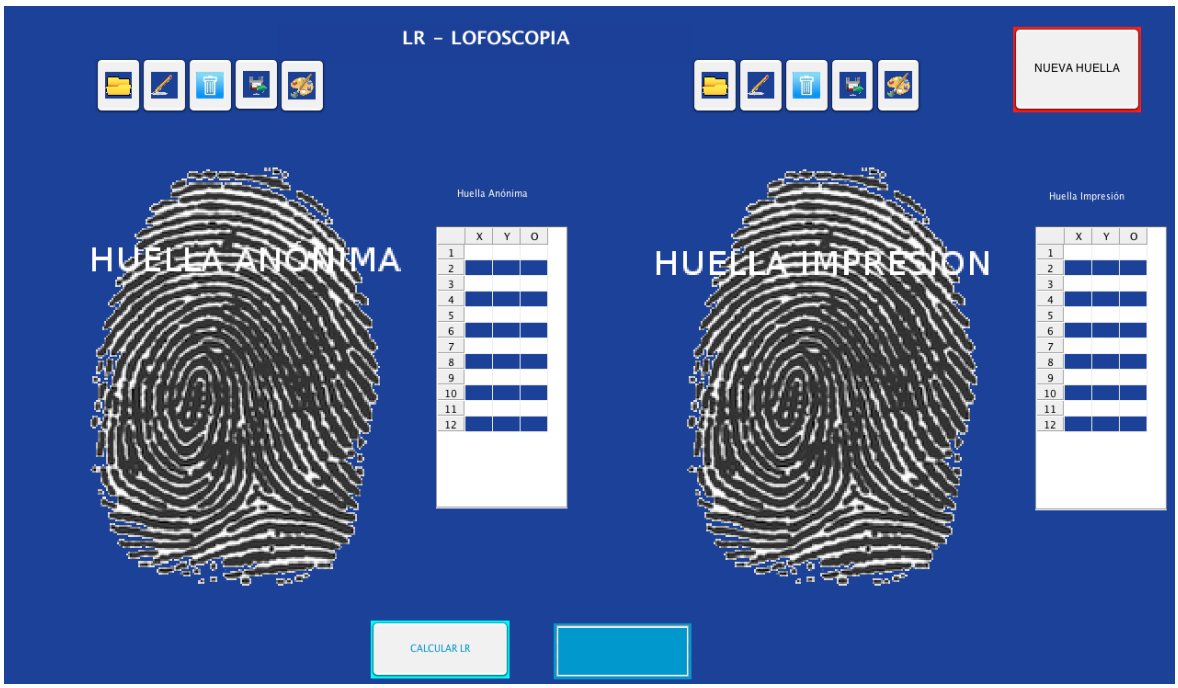

Figura 34 – Pantalla inicial de LR-Lofoscopia

3.1.4 Conclusiones y resultados finales

A continuación se resumen los resultados obtenidos de la recopilación de la base de datos:

- 258 pares de huellas dactilares de 248 individuos distintos implicados (de algunos individuos se tienen muestras de varios dedos, concretamente entre 1 y 4 pares de huellas por individuo)

- 20 pares de huellas palmares de 14 individuos distintos (entre 1 y 3 pares de huellas por individuo)

- Cada set de huellas posee sus correspondientes ficheros de minucias en formato de archivo de texto ".txt".

Se considera que se han alcanzado los objetivos fijados al comienzo de la recopilación de esta base de datos:

- ✓ Disponer de una fuente de información de huella dactilar fiable para la investigación y pruebas que realiza el laboratorio ATVS de la UAM en este entorno.

- ✓ Conversión de la información ya disponible en la DGGC a un formato electrónico y su unificación.

- ✓ Mejorar la metodología de trabajo de los peritos de la DGGC en la medida de lo posible para facilitar la extracción y comparación de minucias de la huella.

Más allá de los resultados puramente teóricos obtenidos, la realización de este trabajo supuso por encima de todo un enriquecimiento personal gracias a la experiencia laboral junto con la DGGC así como la oportunidad de conocer de primera mano las verdaderas aplicaciones de la investigación realizada en el laboratorio ATVS.

3.2 Preparación de la base de datos

El objetivo de esta fase del proyecto es conseguir una nueva base de datos ideal. Lo que esto significa es que deseamos tener el mayor número de minucias posibles para cada huella, ya que cuántas más minucias se tengan, mayor es la probabilidad de identificación.

Para poder llevar a cabo esta tarea se hará uso de dos tipos de datos: los obtenidos en la fase de digitalización de la base de datos de DGGC (a partir de ahora llamados archivos *min*) y los que se obtienen al utilizar un software de extracción automática de minucias (a partir de ahora llamados archivos *veri*). La combinación de estos dos sets de minucias, siguiendo una serie de requisitos, dará como resultado los archivos que componen nuestra base de datos "ideal" (a partir de ahora llamados archivos *ideal*).

Para esta fase únicamente se trabaja con las huellas indubitadas, y en concreto de momento, sólo con los archivos que contienen la localización de las minucias. No será necesaria la utilización de las imágenes hasta más adelante salvo para la comprobaciones intermedias del proceso en caso de que sea necesario representar las minucias sobre la huella.

A continuación se detallan el formato y contenido de los archivos *min* y *veri*, a partir de los cuales se obtendrán los *ideal* como combinación de ambos tras eliminar las minucias duplicadas.

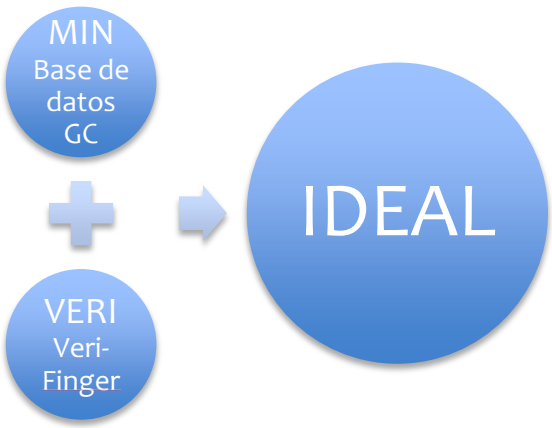

Figura 35 – Esquema de tipos de archivos de la base de datos

3.2.1 Archivo *min*

Se trata de los archivos creados anteriormente en el proceso de digitalización de la base de datos de la DGGC de casos reales forenses.

Como ya se explicó son archivos en formato '.txt' que contienen información acerca de la localización, orientación y tipo de las minucias. Estos sólo contienen 12 minucias (algunas pueden estar formadas por varios puntos) que tienen a su vez sus 12 minucias emparejadas en la huella dubitada de esta misma base de datos por lo que es necesario que las minucias vayan numeradas en ambos casos para poder saber la correspondencia entre huella dubitada e indubitada. Estas minucias pueden ser de cualquiera de los tipos de minucias definidos en la tabla, tanto de alta como de baja frecuencia de aparición.

El formato es el siguiente: se organizan por columnas, correspondiendo cada línea del archivo a un punto de una minucia. La separación entre columnas se hace con un espacio en blanco. Cada columna contiene la siguiente información:

1. Número de minucia
2. Coordenada horizontal
3. Coordenada vertical
4. Orientación de la cresta en grados
5. Número de identificación del tipo de minucia

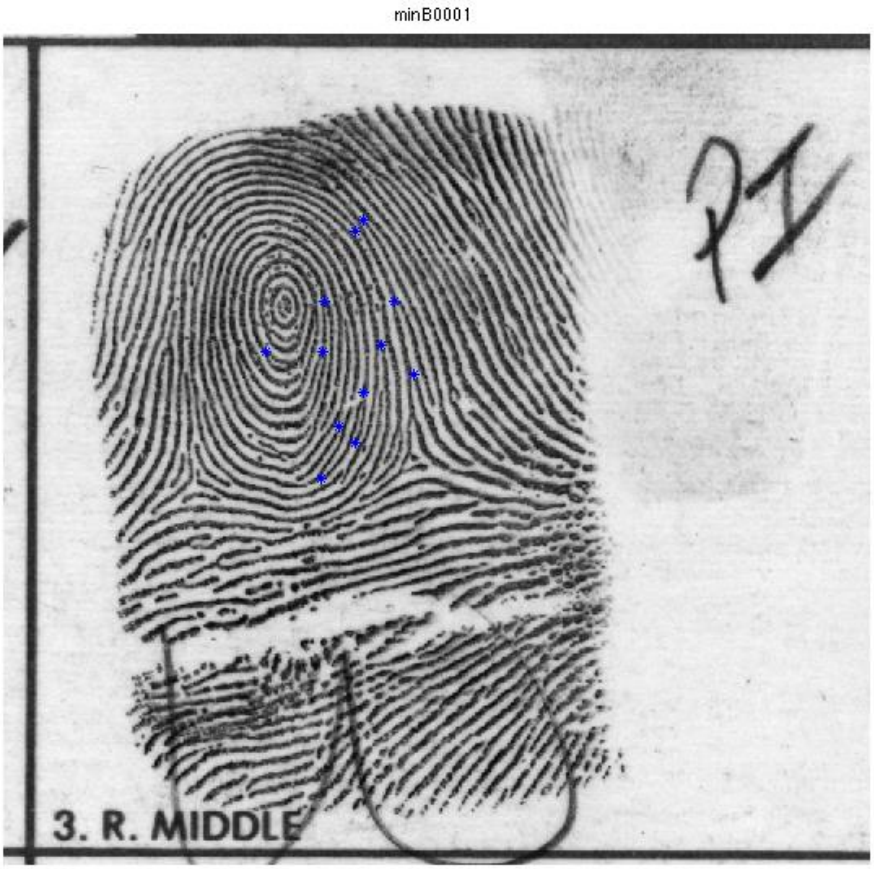

Figura 36 – Minucias pertenecientes a un archivo tipo *min*, representadas sobre su huella

```
1 368 462 132 1
2 375 412 117 2
3 356 384 61 2
4 380 357 334 2
5 357 337 232 1
6 407 326 315 2
7 347 294 213 1
8 436 332 131 1
9 391 338 159 1
10 421 389 294 1
11 398 415 293 7
11 382 396 109 7
12 404 458 140 1
```

Figura 37 – Ejemplo de archivo de minucias tipo *min*

3.2.2 Archivo veri

Estos archivos, al igual que los anteriores, contienen información acerca de la localización, orientación y tipo de minucia. Sin embargo existen algunas diferencias entre ellos y la principal es el método de obtención de las minucias.

Haciendo uso de un software de extracción de minucias automática (VeriFinger) disponible de forma gratuita, generamos un set de minucias para las huellas indubitadas. Este set está compuesto por todas las minucias posibles que se puedan extraer sin importar si están emparejadas o no con las de la huella dubitada.

Al tratarse de una herramienta automática, tiene sus limitaciones. Una de ellas es que únicamente se obtienen minucias de alta frecuencia de aparición por lo que en los archivos sólo aparecerán minucias del tipo 1 y 2.

Otro problema de utilizar este tipo de herramientas es que depende enormemente de la calidad de la imagen, por lo que, en imágenes con baja calidad, es posible que se identifiquen minucias fuera de la huella. Este problema se abordará más adelante.

Por el momento nos centraremos en el archivo que se obtiene al ejecutar el software. El formato es idéntico al de los archivos *min*. La información se almacena en ficheros ".txt" en 5 columnas. La única diferencia es que en este caso el número de minucias en cada archivo no está limitado a 12 sino que habrá tantas como se puedan encontrar en la huella (suele variar entre 100 y 200 minucias por huella), aunque sólo podrán ser terminaciones abruptas o convergencia/bifurcación.

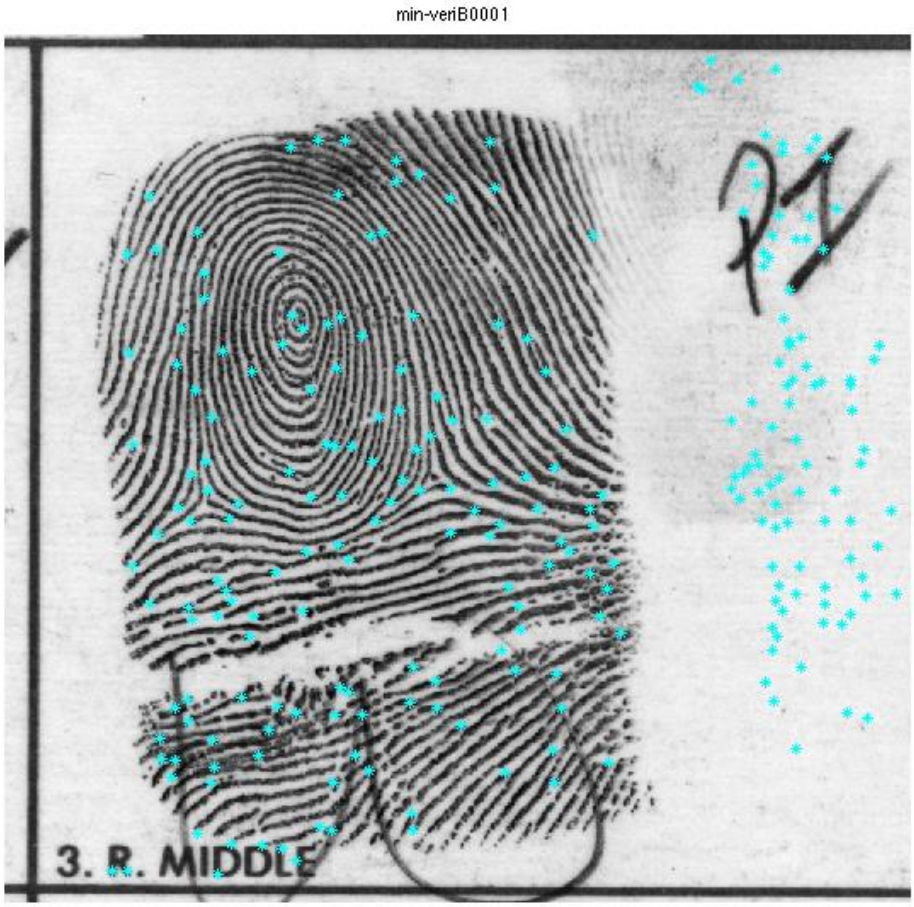

Figura 38 – Minucias pertenecientes a un archivo tipo *veri*, representadas sobre su huella

```
116 208 531 244 1
117 300 531 114 2
118 260 539 114 2
119 555 539 128 2
120 390 550 76 1
121 556 551 126 1
122 330 554 102 2
123 496 554 132 2
124 291 565 101 2
125 512 569 254 1
126 243 570 242 1
127 397 572 228 1
128 318 573 215 1
129 430 573 130 1
130 578 574 120 1
131 299 575 224 1
132 210 576 240 1
```

Figura 39 – Ejemplo de archivo de minucias tipo *veri*

3.2.2.1 Software de extracción automática de minucias VeriFinger SDK

El software utilizado para la extracción de minucias automática es VeriFinger SDK [16]. Se trata de un software de identificación de huellas dactilares diseñado para los desarrolladores e integradores de sistemas biométricos. Su tecnología asegura un rendimiento rápido y fiable para huellas digitales. Funciona tanto en modo verificación (comparación uno a uno) como en modo identificación (comparación uno a varios).

El algoritmo del que hace uso este sistema fue desarrollado en 1998 por Neurotechnology [17]. Desde entonces se ha actualizado en más de 10 versiones posteriores constituyendo uno de los algoritmos de identificación dactilar más potentes hasta la fecha. Su última versión (VeriFinger 7.0) está basada en el identificador de huellas dactilares MegaMatcher, que ha sido reconocido por el NIST como adecuado para su uso en aplicaciones de programa de verificación de identidad personal (en inglés personal identity verification, PIV).

El algoritmo de identificación VeriFinger sigue el esquema de identificación de huellas dactilares comúnmente aceptado, que utiliza un conjunto de puntos de huellas dactilares específicas (minucias), junto con una serie de soluciones algorítmicas patentadas que mejoran el rendimiento y fiabilidad del sistema.

3.2.3 Archivo *ideal*

Por último el archivo ideal contendrá la información final que se desea obtener, la combinación de los dos archivos anteriores, min y veri.

El objetivo principal es obtener el mayor número de minucias de cada huella para lo cual se combinarán las obtenidas de forma automática con VeriFinger y las obtenidas manualmente por la DGGC. Este proceso puede parecer sencillo a simple vista pero es necesario tener en cuenta varios detalles para realizarlo correctamente:

- El formato final del fichero será exactamente igual al de los anteriores: información ordenada en 5 columnas separadas por espacios. Cada línea corresponde a un punto de una minucia.

- Como es lógico, al unir ambos ficheros, es posible que haya minucias repetidas que se encuentren en ambos. Será necesario identificar estas minucias y eliminar los duplicados. Una vez identificados se eliminará la minucia correspondiente al archivo *veri*, y se dejará la de *min*, que fue extraída de forma manual.

3.2.3.1 Conceptos previos

A continuación se explica el detalle del proceso de obtención de estos ficheros para lo cual se exponen brevemente una serie de conceptos previos con los que es necesario estar familiarizado para su correcto entendimiento.

DISTANCIA EUCLÍDEA

En matemáticas, la distancia euclídea se define como la distancia "ordinaria" entre dos puntos de un espacio euclídeo, la cual se puede deducir a partir del teorema de Pitágoras.

Por ejemplo, en un espacio bidimensional, la distancia euclídea entre dos puntos P_1 y P_2, de coordenadas cartesianas (x_1, y_1) y (x_2, y_2), respectivamente es:

$$d_e(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

En nuestro caso se utilizará este concepto para el cálculo de la distancia entre minucias. Para lo cual se hará uso de sus coordenadas, que se encuentran en las 2ª y 3ª columnas de los archivos *min* y *veri*.

3.2.3.2 Detalle del proceso de obtención de los archivos ideal

Para realizar esta tarea se ha desarrollado un pequeño programa que recibe los dos ficheros de minucias *veri* y *min*, y devuelve como salida el fichero *ideal*. El uso de este programa convierte esta tarea en un proceso mucho más rápido y efectivo teniendo en cuenta el tamaño de la base de datos.

1. El primer paso consiste en identificar dentro de *min* las minucias de tipo terminación abrupta y bifurcación/convergencia. La razón es que *veri* sólo posee minucias de este tipo por lo que no es necesario buscar minucias repetidas de otros tipos ya que nunca las habrá.
   Para esto se hace una primera pasada por el fichero de minucias *min* identificando las líneas en las que la última columna sea distinta de 1 o 2 y se ponen todos sus campos a 0. Después se hará una segunda pasada para eliminar estas líneas. Y como resultado tendremos un fichero únicamente con minucias de tipo 1 y 2. Por último es necesario reenumerar las minucias para sustituir los números que hayan sido eliminados.

1	368	462	132	1
2	375	412	117	2
3	412	384	61	2
4	117	357	334	2
5	357	337	232	1
6	407	326	315	2
7	347	294	213	1
8	436	332	131	1
9	391	338	159	1
10	421	389	294	1
11	398	415	293	7
11	382	396	109	7
12	404	458	140	1

1	368	462	132	1
2	375	412	117	2
3	412	384	61	2
4	117	357	334	2
5	357	337	232	1
6	407	326	315	2
7	347	294	213	1
8	436	332	131	1
9	391	338	159	1
10	421	389	294	1
11	404	458	140	1

Figura 40 – Archivo *min*, antes y después de la eliminación de las minucias

2. A continuación se define dentro de la totalidad del área de la huella, la zona en la que se encuentran localizadas el conjunto de minucias *min*. Esto permite acotar el área de búsqueda de minucias repetidas dentro de la huella. Para esto se busca entre las minucias min resultantes del paso anterior (sólo las de tipo 1 y 2) las coordenadas máxima y mínima tanto horizontal como vertical. Estas coordenadas, añadiendo un margen de 10 pixeles en cada dirección, definen el área rectangular de búsqueda de minucias repetidas.

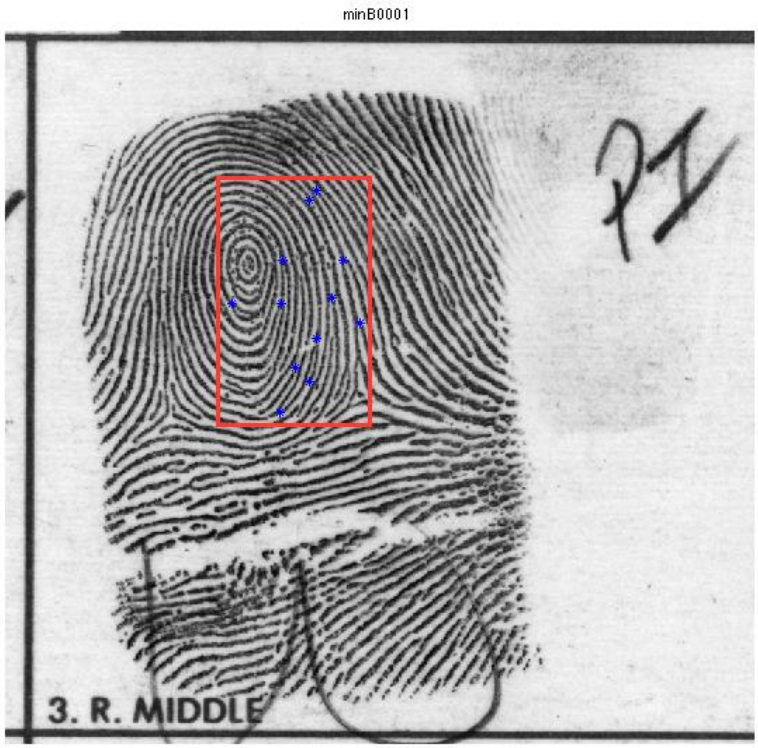

Figura 41 – Diagrama de localización del área de *min* dentro de la huella

3. Una vez definida el área de búsqueda el siguiente paso es identificar las minucias de *veri* repetidas en *min* y eliminarlas.

   I. La primera comprobación es saber si la minucia de *veri* se encuentra dentro del área de búsqueda con una simple comparación de coordenadas. En caso de no estarlo, se pasa directamente a la siguiente minucia, dejando ésta intacta. En caso de sí estar contenida en el área de búsqueda, es necesario hacer una segunda comprobación.

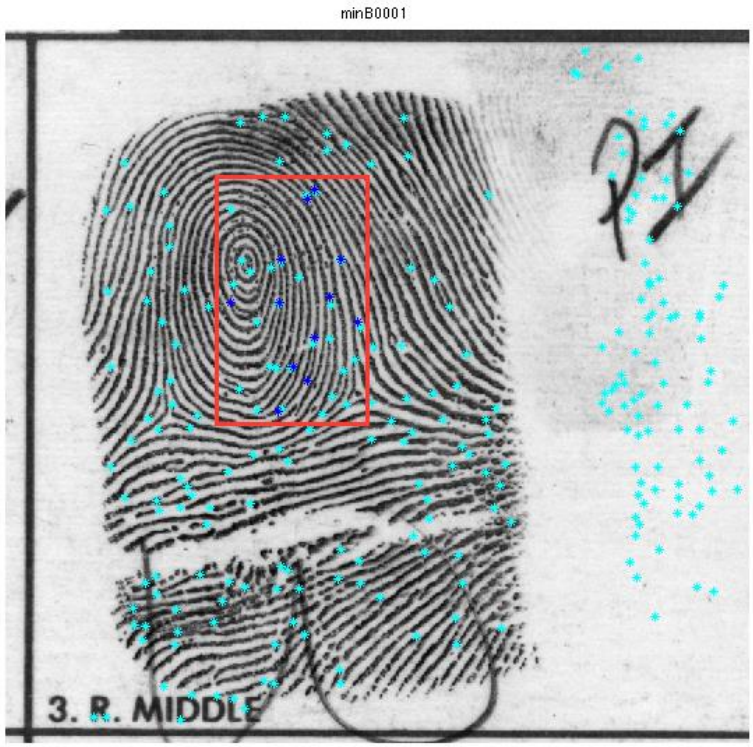

Figura 42 – Huella que contiene tanto las minucias tipo *min* como *veri*

   II. Se compara esta minucia con cada una de las minucias *min* dentro del área de búsqueda y se calcula la distancia euclídea entre los dos puntos. Se asume que si la distancia entre ambas es menor a 10 píxeles la minucia es la misma. Por tanto si la **distancia** es menor que 10, se elimina la minucia de *veri*. La eliminación de minucias se hace igual que en el caso de *min*. Se localiza la minucia, se ponen a 0 todos sus campos, y posteriormente se elimina la línea.

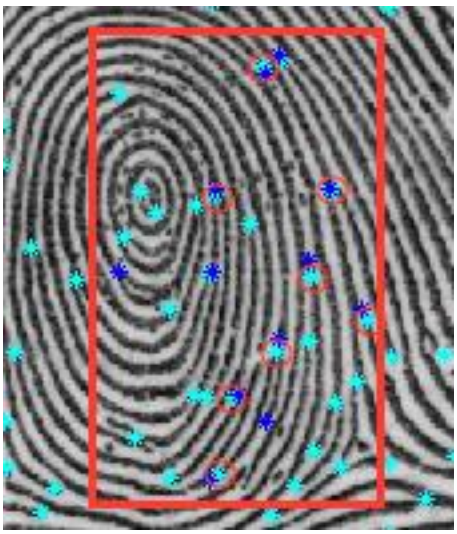

Figura 43 – Minucias repetidas en *min* y *veri*

4. Ya tenemos definido el nuevo fichero *veri*. El último paso es unirlo con el *min* original en uno, que será el archivo *ideal*. Es importante destacar que el archivo *min* utilizado para crear el *ideal* es el original que contiene las 12 minucias (que pueden ser de cualquiera de los tipos). La unión se hace situando al principio las 12 minucias de *min*, seguidas de todas las que compongan el nuevo *veri*. En este caso también Será necesario reenumerar esta nueva lista de minucias.

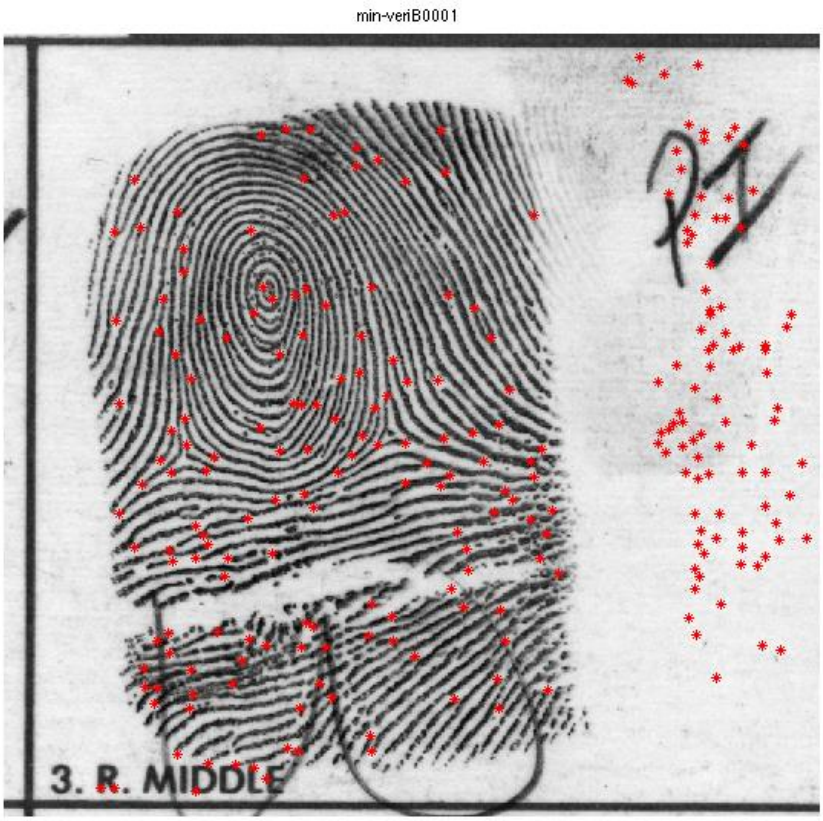

Figura 44 - Minucias pertenecientes a un archivo tipo *ideal*, representadas sobre su huella

Los ficheros *ideal* se almacenarán en su carpeta correspondiente con el nombre "idealXXXX.txt" donde XXXX será la etiqueta del número de huella a la que pertenecen. Y por tanto cada una de ellas tendrá sus correspondientes archivos *min*, *veri* y su imagen de la huella todos ellos identificados con este mismo número.

3.3 Extracción de la región de interés (ROI)

3.3.1 Objetivo

Se define como la región de interés, o en inglés "region of interest" (de aquí en adelante ROI), el área de la imagen de la huella dactilar que contiene las crestas. Es decir, la parte de la imagen que no es el fondo.

Debido a la baja calidad de las imágenes de las huellas indubitadas de la Base de Datos de la Guardia Civil, VeriFinger genera falsas minucias fuera de la ROI de la huella. Obviamente, estas minucias son información falsa acerca de la huella, que interfiere en el proceso de identificación. Por ello previamente se elimina la región del fondo, que no es huella.

La segmentación, en el ámbito del procesamiento de imágenes, es el proceso de dividir una imagen digital en varias partes u objetos. El objetivo de la segmentación es simplificar y/o cambiar la representación de una imagen en otra más significativa y más fácil de analizar. La segmentación se usa tanto para localizar objetos como para encontrar los límites de estos dentro de una imagen.

Para solucionar el problema de las falsas minucias localizadas fuera de la ROI, se ha generado en este PFC un algoritmo propio de segmentación para localizar la región de interés, basado los Filtros de Gabor, como una etapa de postprocesado de las imágenes.

Al igual que en la fase anterior, esta sólo se aplicará a las huellas indubitadas de la base de datos digitalizada de la DGGC. Para las huellas dubitadas, se utilizarán las minucias extraídas a mano como set de minucias ideal ya que todas están localizadas dentro de la ROI.

Figura 45 – Huella antes de comenzar la extracción de la ROI

3.3.2 Detalle del proceso

A continuación se detalla paso a paso el proceso de segmentación (extracción de la región de interés) de la imagen de la huella. Cada una de estas operaciones se realiza para todas las imágenes de la base de datos indubitada.

El proceso sigue este esquema:

Figura 46 – Diagrama resumen del proceso de extracción de la ROI

3.3.2.1 Imagen en escala de grises

En primer lugar, para tratar las imágenes es necesario asegurarse de que todas se encuentran en el mismo formato. La herramienta que se va a utilizar para todo el proceso de segmentación es Matlab por lo que es necesario familiarizarse con los formatos con los que esta herramienta trabaja.

Las imágenes de las huellas son imágenes en blanco y negro, todas ellas con iguales dimensiones, 500x500 píxeles (esta comprobación ya se realizó en la fase de digitalización). Disponer de las imágenes en un formato unificado es importante para agilizar el proceso y realizarlo lo más rápidamente posible. Para ello se ha decidido trabajar con éstas en escala de grises.

El formato de escala de grises consiste en asignar un valor a cada píxel de la imagen que corresponde a su graduación de gris. El valor del píxel puede variar entre 256 valores (desde 0 a 255) siendo 0 el equivalente al negro, y el 255 el blanco). Por tanto al

operar con la imagen en Matlab estaremos tratando una matriz de 500x500 con valores comprendidos entre 0 y 255.

El otro formato en el que se encuentran alguna de las imágenes es RGB, que al contrario que el anterior almacena para cada píxel 3 valores, uno para cada tonalidad (red, green, blue). En este formato se almacena como un array de datos con 3 matrices. Cada una de ellas del tamaño de la imagen, donde cada valor se corresponde a un píxel de la imagen. La combinación de las tres componentes básicas de color guardadas en cada plano define el propio color a representar en el píxel correspondiente. Cada componente de color es de 8 bits (256 valores).

Para realizar la comprobación y convertir las imágenes al formato deseado simplemente se utiliza la función de Matlab "rgb2gray" que transforma las imágenes en color a escala de grises.

3.3.2.2 Media y varianza de una imagen: normalización

El siguiente paso es normalizar todas las imágenes a los mismos valores de media y varianza. Es decir, que el valor medio de gris utilizado en todas las imágenes sea el mismo y que varíen aproximadamente entre los mismos tonos de gris. Traducido en la imagen esto significa que no haya una imagen que tenga mucho contraste entre blancos y negros y otra que no tenga apenas.

Para entender este proceso correctamente es necesario conocer los conceptos de media y varianza. A continuación se explican ambos y se detalla su cálculo.

MEDIA

Se define como el valor medio de los valores de todos los píxeles de la imagen. Cálculo:

$$m = \frac{\sum valor\ píxel}{n}$$

Donde *n* es el número total de píxeles de la imagen:

$$n = num\ píxeles_{vertical} \cdot num\ píxeles_{horizontal}$$

VARIANZA

Se define como la variabilidad de los valores de los píxeles alrededor de la media. Cálculo:

$$v = \frac{\sum(valor\ píxel^2)}{n} - m^2$$

Los valores de media y varianza son distintos para cada imagen.

Para la normalización de las imágenes a unos mismos valores de media y varianza lo primero es elegir estos valores deseados, que en este caso han sido los siguientes:

$$media\ deseada: m_0 = 100$$

$$varianza\ deseada: v_0 = 128$$

El proceso de normalización consiste en reasignar valores a cada píxel de la imagen a partir de una determinada operación. Para lo cual primero se calculan sus valores de media (m) y varianza (v) para cada imagen y a continuación se realiza la siguiente operación para cada píxel:

- Si el valor del píxel se encuentra por encima de su media: $p_i > m$

$$p_f = m_0 + \sqrt{(p_i - m)^2 \cdot \frac{v_0}{v}}$$

- Si el valor del píxel se encuentra por debajo de su media: $p_i < m$

$$p_f = m_0 - \sqrt{(p_i - m)^2 \cdot \frac{v_0}{v}}$$

Donde p_i y p_f son los valores del píxel antes y después de la normalización respectivamente.

A continuación se muestra la imagen antes y después de la normalización de sus valores de media y varianza:

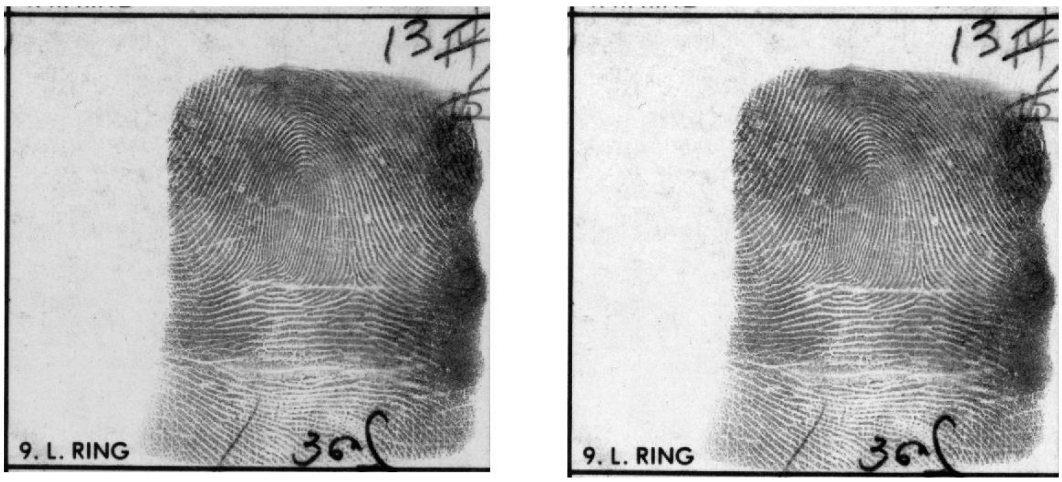

Figura 47 – Huella antes y después de la normalización

3.3.2.3 Aplicación del Filtro de Gabor

En esta fase entramos de lleno en la segmentación de la imagen. Y para ello haremos uso de los Filtros de Gabor.

El Filtro de Gabor es un filtro lineal cuya respuesta al impulso es una función sinusoidal multiplicada por una función gaussiana. Son funciones casi paso banda. La principal ventaja que se obtiene al introducir la envolvente gaussiana es que las funciones de Gabor están localizadas tanto en el dominio espacial como en el de frecuencia, a diferencia de lo que ocurre con las funciones sinusoidales, que están perfectamente localizadas en el dominio frecuencial y completamente deslocalizadas en el espacial.

La transformada de Fourier de un filtro de Gabor son gaussianas centradas en la frecuencia de la función sinusoidal. Se puede llegar a este resultado empleando la propiedad de convolución de la transformada de Fourier, que transforma los productos en convoluciones.

El filtrado de una imagen con funciones de Gabor está relacionado con los procesos en la corteza visual. Además, los filtros se han empleado en el procesamiento digital de imágenes, donde se han mostrado eficientes a la hora de realizar diferentes tareas, tales como segmentación de texturas, compresión, etc.

En este caso se han utilizado filtros de Gabor para 8 direcciones distintas, separadas 45° entre sí, para abarcar toda la circunferencia. La dimensión que se ha especificado es de áreas de 25x25 píxeles.

Implementación del filtro de Gabor:

$$filter = e^{\left(-\frac{1}{2}t\right)\cdot\cos(2\pi\cdot f\cdot\theta_x)}$$

Donde:

$$t = \frac{\theta_x^2}{\sigma_x^2} + \frac{\theta_y^2}{\sigma_y^2}$$

$$\theta_x = \sin(\theta)\cdot x + \cos(\theta)\cdot y$$

$$\theta_y = -\cos(\theta)\cdot x + \sin(\theta)\cdot y$$

f (frecuencia) = $1/8$

σ_x, σ_y (sigma) = 4

θ (Theta): ángulo de orientación del filtro

Y donde x e y son las coordenadas de cada píxel dentro del filtro (que varían de -12 a 12).

Este es un ejemplo de los filtros de Gabor aplicados en las 8 direcciones:

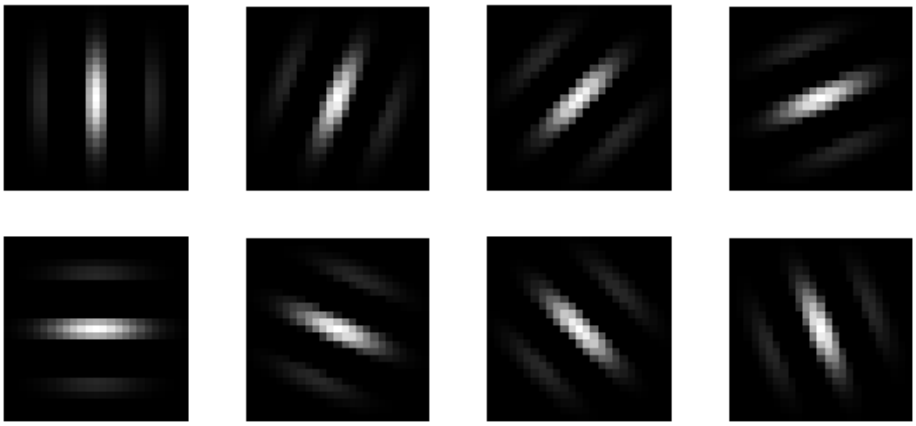

Figura 48 – Filtro de Gabor representado en 8 direcciones

Para aplicar el filtrado a la imagen se utiliza la función *imfilter* de Matlab que convoluciona la matriz de la imagen con la matriz del filtro definida anteriormente. Se obtiene como resultado las 8 matrices correspondientes a la convolución con los filtros en cada una de las 8 orientaciones seleccionadas con el ángulo θ.

A continuación se muestran ejemplos del resultado de la convolución para 4 de los 8 filtros de Gabor aplicados a la imagen original. Como se puede observar, cada una de las orientaciones del filtro, resalta las crestas que siguen su misma dirección.

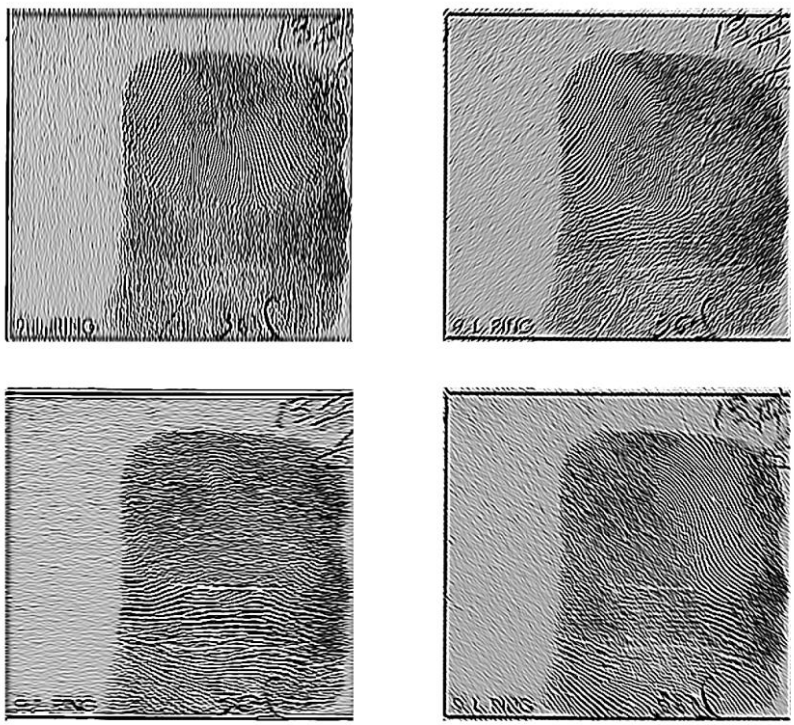

Figura 49 – Resultado de la convolución de la imagen de la huella con el Filtro de Gabor en 4 de las 8 direcciones

Por último se calcula la media de estas 8 imágenes para obtener una única. El cálculo de la media se realiza con la definición de media aritmética, ya que todas las imágenes tienen las mismas dimensiones:

$$Filtered\ Image_{mean} = \frac{\sum_{i=1}^{8} Filtered\ Image_i}{8}$$

Figura 50 – Resultado tras calcular la media de las 8 imágenes resultantes de aplicar Filtro de Gabor

3.3.2.4 Umbralización mediante el gradiente

Tras probar distintos métodos finalmente se llegó a la conclusión de que la manera de obtener mejores resultados en la extracción de la ROI era mediante el uso del gradiente.

En cálculo vectorial, el gradiente ∇f de un campo escalar f es un campo vectorial. El vector gradiente de f evaluado en un punto genérico *x* del dominio de *f*, $\nabla f(x)$, indica la dirección en la cual el campo *f* varía más rápidamente y su módulo representa el ritmo de variación de *f* en la dirección de dicho vector gradiente.

Esto se traduce en la imagen como la identificación de las zonas en las que hay más variabilidad en los píxeles para una dirección en concreto. En este caso el área que queremos identificar dentro de la imagen es la zona donde se encuentran las crestas y surcos, que es la que tiene mayor variabilidad, mientras que el fondo tiende a ser más uniforme.

Para realizar esta tarea se desarrolló un código cuyo funcionamiento se explica a continuación.

La función recibe como argumentos la imagen obtenida como resultado de aplicar los filtros de Gabor (la imagen media de las 8 resultantes), y un factor de umbralización.

En primer lugar, se calcula el gradiente de cada píxel, tanto en la dirección x como en la y, de la imagen de forma automática gracias a la función "gradient" de Matlab. Por lo que se almacenarán dos valores para cada píxel. Cuando el valor del gradiente se encuentra por encima del valor de umbralización en ambas direcciones se etiqueta el píxel correspondiente como huella o fondo.

El valor elegido como óptimo de umbralización es de 0.15.

Esta acción se lleva a cabo para toda la imagen por lo que al final, todos sus píxeles quedan clasificados como huella o fondo. La representación se realiza mediante un píxel blanco o negro, aunque realmente lo que se almacena de cada huella es una matriz con unos (píxel blanco) y ceros (píxel negro). Gracias a esto posteriormente la selección de las minucias dentro de la ROI es casi trivial.

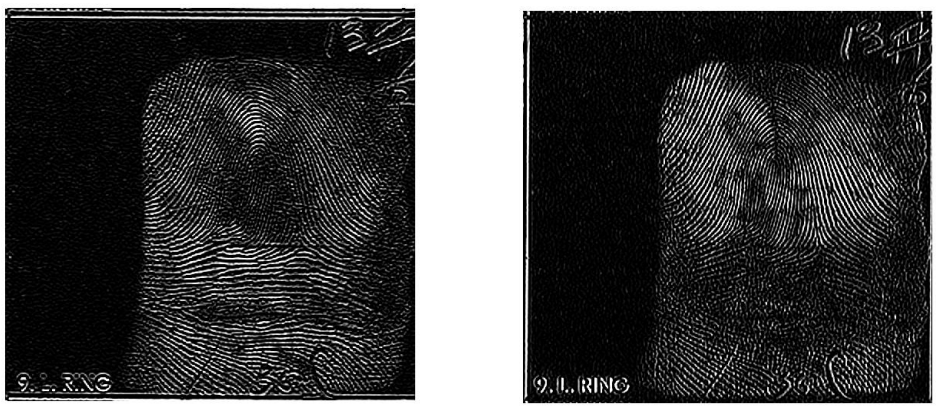

Figura 51 – Resultado del cálculo del gradiente en ambas direcciones

Las imágenes anteriores muestran el resultado de la aplicación del cálculo del gradiente a cada píxel de la imagen. Una para el eje vertical y otra para el horizontal. Y tras la umbralización:

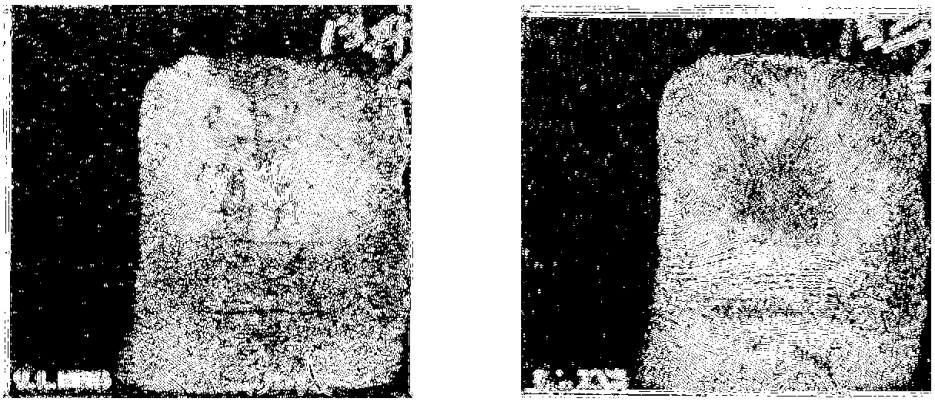

Figura 52 – Resultado tras umbralizar con un valor de 0.15

Por último la combinación de ambas:

Figura 53 – Resultado de la unión de ambas imágenes

3.3.2.5 Eliminación de imperfecciones mediante dilatación/erosión

Gracias al proceso anterior hemos obtenido casi perfectamente la ROI. Sin embargo, en algunos casos pueden surgir imperfecciones y el área no quedar definida correctamente. Esto se debe a que para agilizar el proceso se ha elegido un mismo factor de umbralización para la base de datos al completo cuando en realidad se podría adaptar mínimamente a cada imagen en particular para obtener un resultado más estricto.

Para solventar estas imperfecciones que puedan surgir del hecho de tener un único nivel de umbralización añadimos una etapa posterior en la que utilizaremos las herramientas de dilatación y erosión sobre las máscaras anteriormente definidas.

La dilatación y erosión son las operaciones básicas en el procesamiento de imágenes morfológico, en las cuales se basan el resto de operaciones morfológicas. Fue definida originalmente para imágenes binarias, más tarde se extendió a imágenes en escala de grises y posteriormente a retículos completos. En nuestro caso lo aplicaremos sobre imágenes binarias.

La idea básica en la morfología binaria es probar como una imagen con una forma predefinida simple (por ejemplo un cuadrado o un rectángulo), encaja o no en las formas de la imagen. Esta forma simple se llama elemento estructurante y es también en sí misma otra imagen binaria.

Gracias a la elección del elemento estructurante adecuado, es posible ensalzar o eliminar por completo formas determinadas de la imagen (por ejemplo, líneas horizontales o verticales o pequeños puntos aislados).

Si analizamos las imágenes obtenidas en el apartado anterior, observamos que las irregularidades obtenidas en la mayoría de ellas son muy similares: áreas marcadas en los bordes de la imagen, puntos aislados, etc. Cuando en general lo que se desea obtener es la forma redondeada del contorno de la huella.

Para tratar de obtener esta figura, se han aplicado las siguientes operaciones, todas ellas programadas de nuevo en un script de Matlab:

1. Eliminación de los huecos entre crestas (unificar el área dentro de la huella): como la huella está marcada con píxeles blancos la operación a realizar para eliminar los negros es una dilatación seguida de una erosión, ambas aplicadas con un elemento estructurante cuadrado de 8x8.

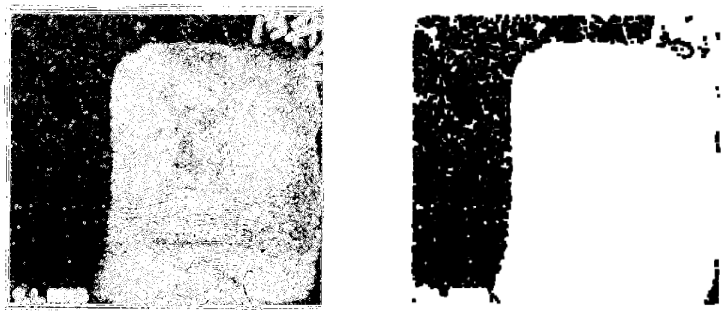

Figura 54 – Imagen de la huella antes y después de la fase 1

2. Eliminación de los huecos en el fondo (unificar el área fuera de la huella): como el fondo está marcado con píxeles negros, la operación a realizar para eliminar los blancos es una erosión seguida de una dilatación, ambas aplicadas con un elemento estructurante cuadrado de 16x16. La razón de que se utilice un elemento estructurante mayor es que en el fondo las imperfecciones son mayores.

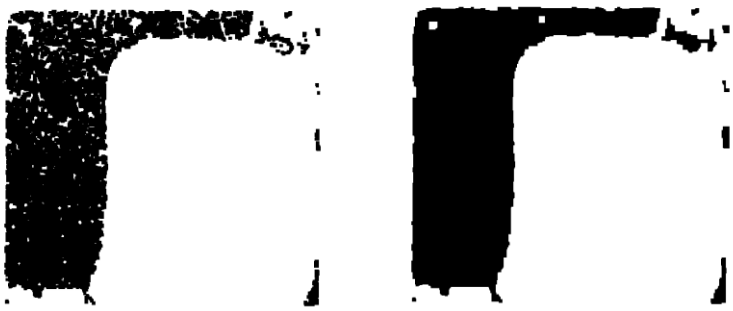

Figura 55 – Imagen de la huella antes y después de la fase 2

3. Tras estos dos primeros pasos aún se pueden observar en algunos casos agujeros grandes dentro de la ROI. Esto ocurre cuando los huecos entre crestas son más grandes que el elemento estructurante. Para eliminarlos se aplica de nuevo la operación de dilatación + erosión pero en este caso se aumenta el tamaño del elemento estructurante y se utiliza uno de 16x16.

Figura 56 – Imagen de la huella antes y después de la fase 3

4. Eliminación de líneas verticales: para ello se define un elemento estructurante de 1x70 y se aplica la operación de erosión seguida de dilatación (ya que lo que se desea eliminar son los píxeles blancos del fondo), ambas con el mismo elemento estructurante.

Figura 57 – Imagen de la huella antes y después de la fase 4

5. Eliminación de líneas horizontales: de forma similar a la anterior pero utilizando un elemento estructurante de 70x1.

Figura 58 – Imagen de la huella antes y después de la fase 5

6. Por último se reduce el área de la imagen un poco para eliminar las falsas minucias que se detectan en los bordes de la huella como terminaciones abruptas. Se utiliza un elemento estructurante de 16x16 y en este caso únicamente se erosiona la imagen ya que sólo queremos reducir y no ampliar el área de la ROI.

Figura 59 – imagen de la huella antes y después de la fase 6

Una vez hecho esto ya tendríamos definidas correctamente las ROI de la base de datos al completo.

Como se puede apreciar en el ejemplo anterior, no se eliminan del todo los elementos fuera de la ROI. Por motivos de seguridad el ejemplo mostrado no corresponde a una huella de la base de datos con la que se ha trabajado realmente, sino a una huella de la base de datos NIST. A continuación se muestran algunos ejemplos reales de las máscaras finales que resultaron salir del la segmentación:

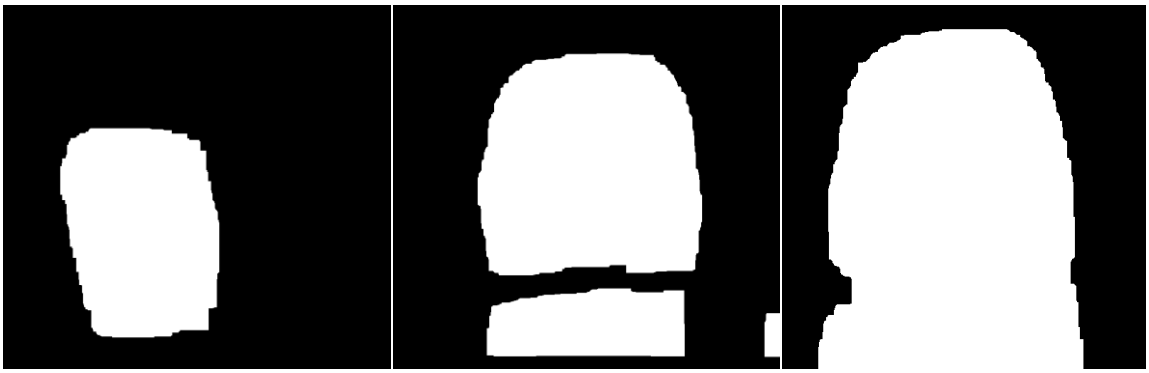

Figura 60 – Ejemplos de extracción de la ROI en huellas pertenecientes a la base de datos

En estas imágenes se puede corroborar los buenos resultados obtenidos de la segmentación, aunque es cierto que no se llegaban a eliminar el 100% de las zonas fuera de la ROI, sí en su mayoría.

3.3.2.6 Archivos *post*: eliminación de minucias fuera de la ROI

Por último se añade un último tipo de archivo a la base de datos, el cual seguirá el mismo formato que los archivos *min* y *veri*, y contendrá únicamente las minucias que se encuentran dentro de la región de interés de la huella. Denominaremos a estos archivos con el nombre *post*.

Su obtención se lleva a cabo mediante el uso de las máscaras con la ROI que hemos estado obteniendo en todo el proceso anterior. Ya que, tal y como se ha indicado anteriormente, la información almacenada de cada máscara es una matriz de unos y ceros, la obtención de los archivos post es muy rápida y prácticamente trivial.

Figura 61 – Diagrama del proceso de obtención de los archivos *post*

El proceso consiste en recorrer el archivo ideal creado anteriormente como combinación de min y veri, y comprobar cuales de las minucias se encuentran dentro de la ROI gracias a la matriz de la máscara de su imagen correspondiente. Para comprobarlo utilizaremos las coordenadas de cada minucia y accederemos con ellas a la matriz de la máscara. En caso de ser un 1 significa que se encuentra dentro de la ROI, y en caso de ser un 0, significa que se encuentra fuera. En caso de estar dentro, no se hace nada con la minucia. En caso de estar fuera se elimina. La eliminación de minucias de ideal se hace de igual forma que en el resto de archivos. En una primera pasada se ponen todos los campos a 0 y una vez hecho esto para todas las minucias de la huella se hace una segunda pasada en la que se eliminan todas las minucias con todos sus campos a 0.

4

Integración, pruebas y resultados

En este apartado se desarrollarán los experimentos necesarios para comprobar que con los cambios efectuados en el sistema se obtiene una mejora en su rendimiento.

Para ello se desarrolla un plan de experimentos realizados mediante un software basado en algoritmos de comparación de minucias. El plan de experimentos recoge una serie de pruebas en las que se compara el comportamiento del sistema antes y después de la selección de minucias dentro de la región de interés de la huella.

Además, se explica en detalle el funcionamiento del software utilizado para la realización de dichos experimentos.

4.1 Software MCCSdk v1.4

La herramienta utilizada para realizar la fase de pruebas y experimentos es el Software MCCSdk. Se trata de una librería .Net DLL que permite el desarrollo de aplicaciones para identificación dactilar mediante la utilización de algoritmos de tipo "Minutia Cyinder-Code" (MCC). [19][20][21]

El software permite comparar archivos de minucias y obtener una lista ordenada de candidatos según los scores obtenidos en la comparación. Se pueden hacer comparaciones de los distintos tipos de archivos de minucias de la base de datos (*min*, *veri, ideal* y *post*) al completo y extraer conclusiones.

Para realizar las comparaciones entre archivos de minucias se ejecutan varios pasos que adaptan los archivos al formato necesario del programa y posteriormente comparan las minucias.

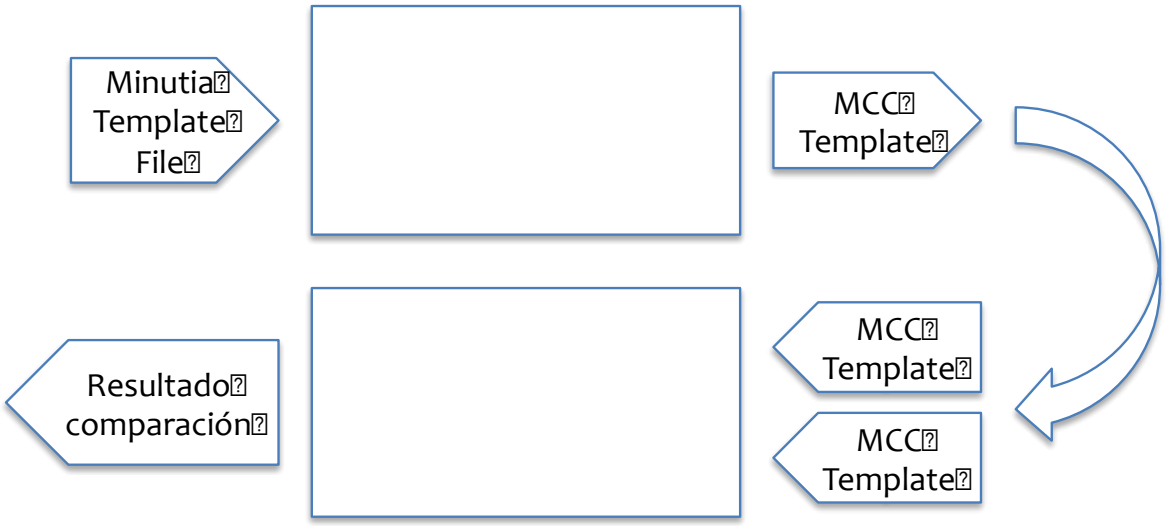

Figura 62 – Diagrama de funcionamiento de MCCSdk v1.4

El primero de ellos, el ejecutable "MccEnroller" recibe archivos con formato de tipo "Minutia Template File Text Format" definido en las instrucciones de uso del programa. Para usar el programa y crear comparaciones entre huellas, los archivos que se usen como entrada deben estar almacenados en un archivo de texto con el siguiente formato:

1. La primera línea contiene el ancho de la imagen de la huella (en píxeles)
2. La segunda línea contiene el alto de la imagen de la huella (en píxeles)
3. La tercera línea contiene la resolución de la imagen (en pixel por pulgada)
4. La cuarta línea contiene el número total de minucias (n)
5. n líneas que contienen las coordenadas X e Y (en píxeles) y el ángulo θ (en radianes) de cada minucia.

```
imageWidth
imageHeight
imageResolution
n
x(1) y(1) θ(1)
x(2) y(2) θ(2)
…
…
…
x(n) y(n) θ(n)
```

Figura 63 – Minutia Template File Text Format

A continuación se explica el funcionamiento de los ejecutables utilizados y las tareas que realizan.

Para convertir los archivos existentes a este formato se ha desarrollado un código que hacía la operación de forma automática.

4.1.1 MccEnroller

MccEnroller es una aplicación tipo .exe (ejecutable) que se usa para registrar los sets de minucias y crear un archivo de tipo MccTemplate. Estos archivos tienen su propio formato predefinido para el uso de este software:

1. La primera línea contiene el ancho de la imagen de la huella (en píxeles)
2. La segunda línea contiene el alto de la imagen de la huella (en píxeles)
3. La tercera línea contiene la resolución de la imagen (en pixel por pulgada)
4. La cuarta línea contiene el número total de minucias (n)
5. n líneas que contienen las coordenadas X e Y (en píxeles) y el ángulo θ (en radianes) de cada minucia.
6. Un línea que contiene el número total de coordenadas cilíndricas (m, normalmente n=m)
7. m líneas que contienen la información de cada coordenada cilíndrica: en la primera columna se valida la coordenada (True o False), si la coordenada es válida, sigue una secuencia de 1 y 0 que representan la validez de cada pixel de la base cilíndrica (0=inválido, 1=válido); finalmente le sigue la secuencia de valores de las coordenadas cilíndricas.

```
imageWidth
imageHeight
imageResolution
n
x(1) y(1) θ(1)
x(2) y(2) θ(2)
…
…
…
x(n) y(n) θ(n)
m
CylinderValidity(1) v(1)(1) v(1)(2) … v(1)(N_S · N_S) c(1)(1) c(1)(2) … c(1)(N_S · N_S · N_D)
CylinderValidity(2) v(2)(1) v(2)(2) … v(2)(N_S · N_S) c(2)(1) c(2)(2) … c(2)(N_S · N_S · N_D)
…
…
…
CylinderValidity(m) v(m)(1) v(m)(2) … v(m)(N_S · N_S) c(m)(1) c(m)(2) … c(m)(N_S · N_S · N_D)
```

Figura 64 – Mcc Template

Para ejecutar el programa se escribe la instrucción:

MccEnroller <MinutiaeTemplateFile> <MccTemplateFile> <MccEnrollParametersFile> <OutputFile>

Donde,

- MinutiaTemplateFile: ruta del set de minucias de entrada en formato de texto (.txt)
- MccTemplateFile: el archivo de salida en el que guardar el MCC Template en formato de texto editable (.txt) predeterminado por el programa (especificada en el párrafo anterior)
- MccEnrollParametersFile: ruta del archivo con los parámetros de registro (Enroll Parameters) guardado en formato XML. El archivo de parámetros utilizado es el MccSdk1.4OptimalEnrollParameters.xml proporcionado junto con el código del programa.
- OutputFile: el archivo de salida en formato de texto editable (.txt), donde se almacena una cadena de registro con el estado resultante de la operación realizada en cada archivo de salida de tipo MccTemplateFile. Se representa en una columna el nombre del archivo de salida y a su lado el resultado. El resultado será OK si el registro se ha realizado de forma correctamente, o FAIL si la imagen de entrada no ha podido ser procesada por el algoritmo.

4.1.2 MccMatcher

MccMatcher, al igual que MccEnroller, es una aplicación tipo .exe que permite realizar comparaciones entre dos archivos MCC (formato MccTemplateFile).

Para ejecutar el programa se escribe la instrucción:

MccMatcher <MccTemplateFile1> <MccTemplateFile2> <MccMatchParametersFile> <OutputFile>

Donde:

- MccTemplateFile1: ruta del primer archivo de minucias de entrada guardado en formato de texto editable (.txt) (archivo tipo MccTemplateFile)
- MccTemplateFile2: ruta del segundo archivo de minucias de entrada guardado en formato de texto editable (.txt) (archivo tipo MccTemplateFile)
- MccMatchParametersFile: ruta del archivo con los parámetros de comparación (Match Parameters) guardado en formato XML. El archivo de parámetros utilizado es el MccSdk1.4OptimalMatchParameters.xml proporcionado junto con el código del programa.

- OutputFiles: el archivo de salida en formato de texto editable (.txt), donde se almacena una cadena de registro con el estado resultante de la comparación realizada entre los dos archivos de minucias. El resultado será OK si la comparación ha sido realizada o FAIL si la comparación no puede ser ejecutada por el algoritmo. También se indica la similitud entre ambos ficheros expresada con un número comprendido entre 0 y 1, donde 0 equivale a ninguna similitud y 1 equivale a máxima similitud.

A continuación se muestra un ejemplo de archivo resultante de comparación de minucias:

```
MCC2pi_M_minA0001.txt MCC2pi_M_postB0001.txt OK 0.213120731091964
MCC2pi_M_minA0002.txt MCC2pi_M_postB0002.txt OK 0.25516196119350842
MCC2pi_M_minA0003.txt MCC2pi_M_postB0003.txt OK 0.14099216890625485
MCC2pi_M_minA0004.txt MCC2pi_M_postB0004.txt OK 0.088512390583280171
MCC2pi_M_minA0005.txt MCC2pi_M_postB0005.txt OK 0.088611779628301079
MCC2pi_M_minA0006.txt MCC2pi_M_postB0006.txt OK 0.11801243514831646
MCC2pi_M_minA0007-1.txt MCC2pi_M_postB0007-1.txt OK 0.096411889568444112
MCC2pi_M_minA0007-2.txt MCC2pi_M_postB0007-2.txt OK 0.10617744697725395
MCC2pi_M_minA0008.txt MCC2pi_M_postB0008.txt OK 0.0945326383278223
MCC2pi_M_minA0009-1.txt MCC2pi_M_postB0009-1.txt OK 0.10284929385909042
MCC2pi_M_minA0009-2.txt MCC2pi_M_postB0009-2.txt OK 0.080959081782824044
MCC2pi_M_minA0010-1.txt MCC2pi_M_postB0010-1.txt OK 0.15775741957621356
MCC2pi_M_minA0010-2.txt MCC2pi_M_postB0010-2.txt OK 0.10870317816312096
MCC2pi_M_minA0011-1.txt MCC2pi_M_postB0011-1.txt OK 0.097361527226486111
MCC2pi_M_minA0011-2.txt MCC2pi_M_postB0011-2.txt OK 0.094961131032127527
MCC2pi_M_minA0012.txt MCC2pi_M_postB0012.txt OK 0.22553985412473682
MCC2pi_M_minA0013-1.txt MCC2pi_M_postB0013-1.txt OK 0.14180228261679195
MCC2pi_M_minA0013-2.txt MCC2pi_M_postB0013-2.txt OK 0.096768332865877507
MCC2pi_M_minA0014-1.txt MCC2pi_M_postB0014-1.txt OK 0.10787684863954598
MCC2pi_M_minA0014-2.txt MCC2pi_M_postB0014-2.txt OK 0.11421380757818467
MCC2pi_M_minA0015-1.txt MCC2pi_M_postB0015-1.txt OK 0.090302295900638277
MCC2pi_M_minA0015-2.txt MCC2pi_M_postB0015-2.txt OK 0.10808562781721628
```

Figura 65 – Ejemplo de archivo de salida de MccMatcher

4.2 Plan de pruebas y experimentos

Es preciso llevar a cabo una serie de experimentos para poder reflejar de alguna manera el avance que se ha producido. En resumen, es una manera de comprobar el fruto del trabajo realizado, las mejoras que nuestro trabajo ha propiciado en la comparación de huellas dactilares en el entorno forense.

Estas comparaciones se realizan con el software citado anteriormente: MCCSdk v1.4.

Para poder realizar esta tarea debemos comparar dos tipos de resultados: los obtenidos en las comparaciones antes del procesado de la base de datos, que ha sido el principal objeto de este proyecto, y los obtenidos después del procesado.

Por otro lado, como ya se ha explicado a lo largo de la memoria, el conjunto de archivos que componen esta base de datos son muy distintos: imágenes de huellas dubitadas, indubitadas, procesadas o sin alterar, archivos de minucias extraídos manualmente o con herramientas de extracción automática. Las distintas combinaciones que estos ofrecen son también una gran fuente de información para las conclusiones.

Las comparaciones entre huellas se realizan, obviamente, entre una huella de la que se desconoce el propietario, es decir dubitada, con una o varias huellas de las que se conoce el propietario, para comprobar si son de la misma persona.

En el caso de las huellas dubitadas (tipo A), la base de datos de la que disponemos, únicamente posee sets de minucias de tipo *min*, extraídas a mano por los expertos de la DGGC, y posteriormente digitalizados. La razón es que las imágenes de las huellas dubitadas tienen una calidad tan baja que es inviable utilizar en ellas el software de extracción de minucias automática.

Para las huellas indubitadas (tipo B) se pueden encontrar todos los tipos de archivos de minucias anteriormente mencionados: *min*, *veri*, *ideal* y *post*. Hay que tener en cuenta que *post* es la combinación y edición de los demás tras la localización de la ROI.

Por tanto, las comparaciones se realizan entre las huellas antes y después de la segmentación de la imagen de la huella indubitada. El archivo veri corresponde a la fase anterior a la segmentación, y el archivo post a la fase posterior, tras realizar la segmentación y extraer la ROI.

A continuación se explican los distintos experimentos y comparaciones realizadas.

4.2.1 Experimento 1: comparaciones genuinas

El primer tipo de comparación es de tipo genuino. Esto quiere decir que se compara una sola huella de un único individuo consigo misma. La diferencia es que se comparan

archivos de minucias de los distintos tipos mencionados anteriormente.

Por ejemplo, para el individuo 273 se compararán distintos archivos de minucias de su huella: el archivo de tipo min de la huella dubitada, con el veri, o con el post, de la indubitada.

El resultado que se espera obtener en estas comparaciones es positivo: es decir, encontrar la mayor coincidencia posible entre sus minucias como para poder afirmar que las dos huellas pertenecen a la misma persona.

Como se recordará de apartados anteriores de esta memoria, se utiliza la terminación "A" para denotar los archivos de minucias o imágenes de huellas dubitadas, y la "B" para denotar aquellos que pertenecen a las huellas indubitadas de la base de datos.

Por otro lado, los archivos de minucias tipo *min* contienen las minucias originales extraídas de forma manual por los expertos forenses de la DGGC. Los de tipo *veri* contienen todas las minucias extraídas con el software de extracción automática VeriFinger. Y los archivos tipo *post* son los que contienen la combinación de las minucias extraídas manualmente y las obtenidas con el software de extracción de minucias automático dentro de la región de interés de la huella.

En primer lugar tenemos las comparaciones:

1. Comparación MINA_XXX-POSTB_XXX
2. Comparación MINA_XXX-VERIB_XXX

Donde XXX representa el número de huella, que siempre será la misma entre las parejas de huellas que se estén comparando, ya que al tratarse de comparaciones genuinas, pertenecen al mismo usuario.

El resultado de esta comparación muestra la tasa de falso rechazo del sistema, según el umbral que decidamos fijar.

4.2.2 Experimento 2: comparaciones de impostores

El segundo tipo de experimento consiste en comparar parejas de huellas que no pertenecen a la misma persona, es decir, comparar la huella de la que se quiere averiguar el propietario con otra huella que sabemos que no es la correcta.

Este tipo de experimento permite comprobar qué rango de similitud se puede obtener entre huellas no pertenecientes al mismo individuo, lo cual resulta muy útil a la hora de establecer un umbral de aceptación o rechazo en un sistema de reconocimiento de huella dactilar.

Al igual que en el primer experimento existen varios tipo de comparaciones según las distintas combinaciones que permite realizar la base de datos con sus tipos de

archivos.

1. Comparación MINA_XXX-POSTB_YYY
2. Comparación MINA_XXX-VERIB_YYY

Donde XXX e YYY representan el número de identificación de la huella dentro de la base de datos. En este caso son siempre distintos entre las parejas de huellas que se estén comparando ya que las huellas no pertenecen al mismo individuo por tratarse de comparaciones de impostores.

El resultado de esta comparación muestra la tasa de falsa aceptación del sistema, según el umbral que decidamos fijar.

4.2.3 Conclusiones

En resumen, teniendo en cuenta que la base de datos está compuesta por 298 parejas de huellas (dubitadas e indubitada), los experimentos dan lugar a las siguientes comparaciones:

Comparaciones Genuinas			Comparaciones de impostores		
Huella Dubitada	Huella Indubitada	Número de comparaciones	Huella Dubitada	Huella Indubitada	Número de comparaciones
Min	Veri	298	Min	Veri	88.506
Min	Post	298	Min	Post	88.506

Figura 66 – Tabla resumen de las comparaciones realizadas en los experimentos

4.3 Resultados experimentales

Este apartado muestra los resultados obtenidos en los experimentos realizados a través de tablas y gráficas que facilitan su comprensión.

En la siguiente gráfica se muestran los resultados de las comparaciones genuinas para los 298 usuarios de la base de datos.

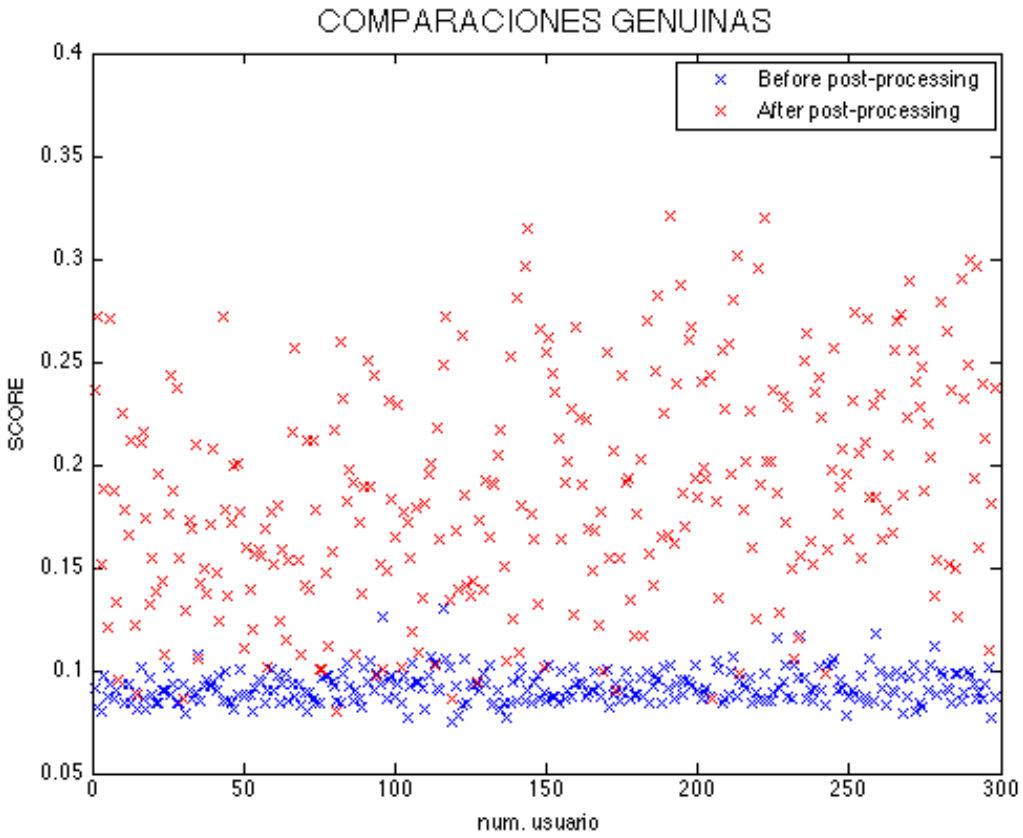

Figura 67 – Comparaciones genuinas antes y después del procesado

Como se puede apreciar en la imagen, mientras que los valores de los scores obtenidos antes del procesado si situaban en torno al 0.1, tras la extracción de la ROI y la eliminación de falsas minucias, el valor medio de los scores para las comparaciones genuinas aumenta notablemente.

Si observamos ahora un ejemplo de las comparaciones de impostores para un usuario de la base de datos, podemos ver que el efecto que tiene la eliminación de falsas minucias fuera de la ROI no es significativo. Una huella falsa, sigue siendo igual de falsa tenga o no minucias fuera de la ROI.

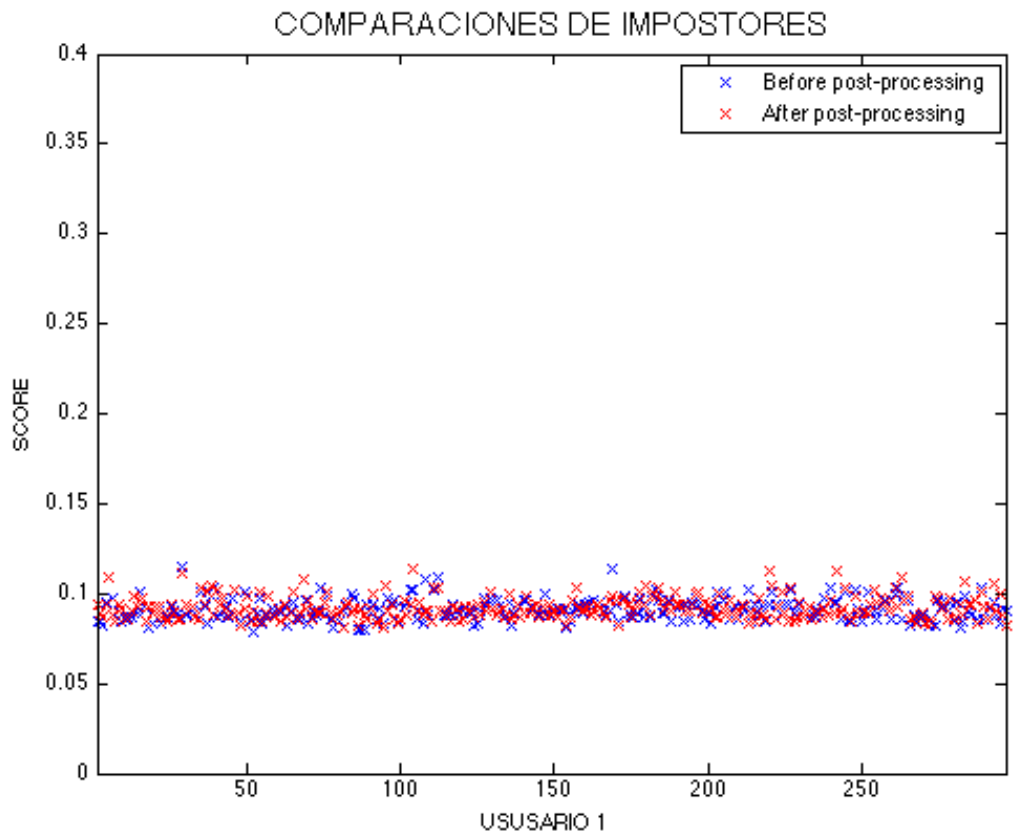

Figura 68 – Comparaciones de impostores para el usuario 1 antes y después del procesado

Por último, analicemos el efecto que tiene el procesado de las imágenes en algunos ejemplos de la base de datos.

Se representan en las siguientes gráficas en azul el resultado de los scores obtenidos de las comparaciones de impostores, y en rojo el valor del score de la comparación genuina. En la primera gráfica se muestran los resultados antes de la eliminación de minucias fuera de la ROI, y en la segunda después.

Claramente, se produce una mejora significativa en la identificación del usuario genuino. El valor del score para el usuario genuino tras la eliminación de falsas minucias sobresale por encima de los demás, lo que hace que sea fácilmente identificable.

En general, si observamos de nuevo las dos primeras gráficas, es mucho más fácil establecer un umbral de decisión tras el procesado de la base de datos, ya que los valores de los scores obtenidos para las comparaciones genuinas se encuentran muy por encima de los de los impostores.

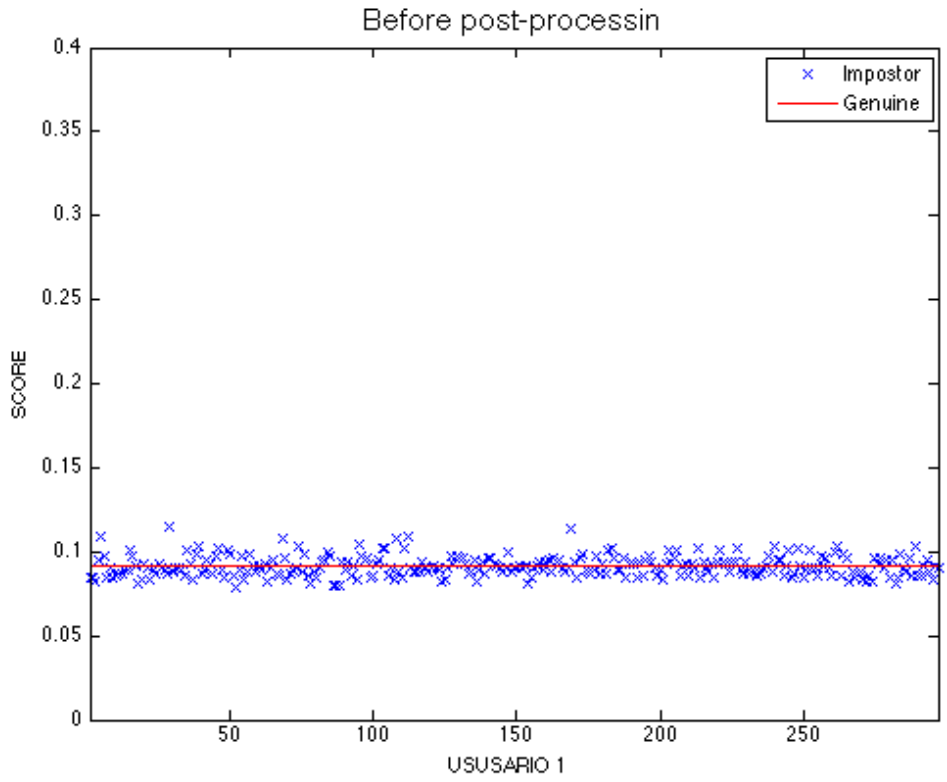

Figura 69 – Comparación de los scores de impostores y usuario genuino 1 antes del procesado

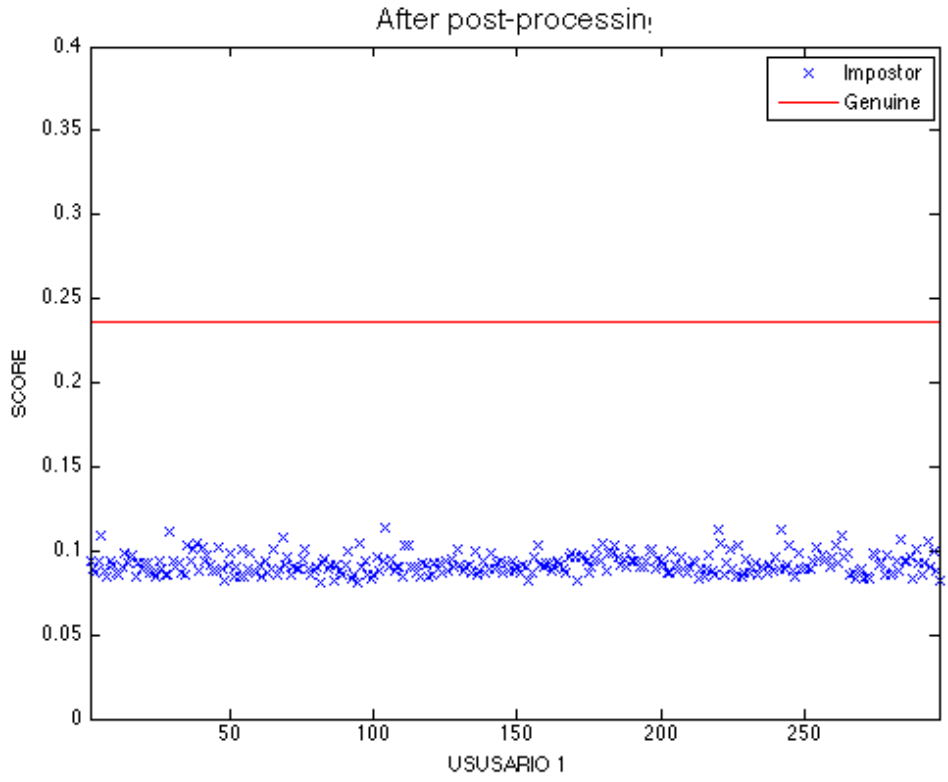

Figura 70 – Comparación de los scores de impostores y usuario genuino 1 tras el procesado

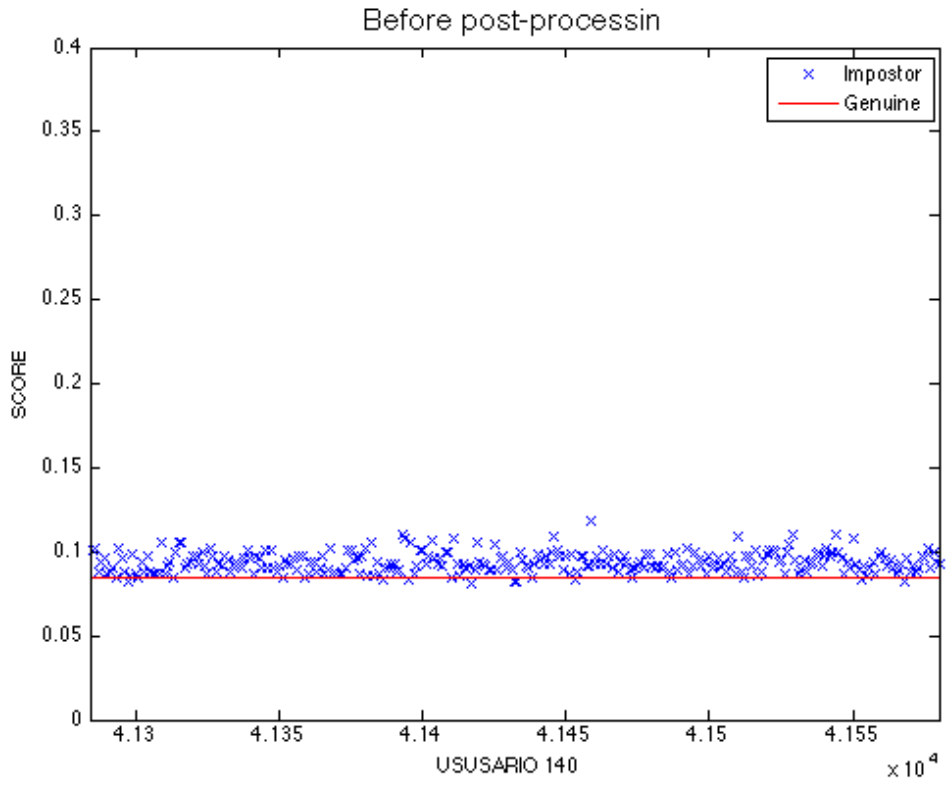

Figura 71 – Comparación de los scores de impostores y us. genuino 140 antes del procesado

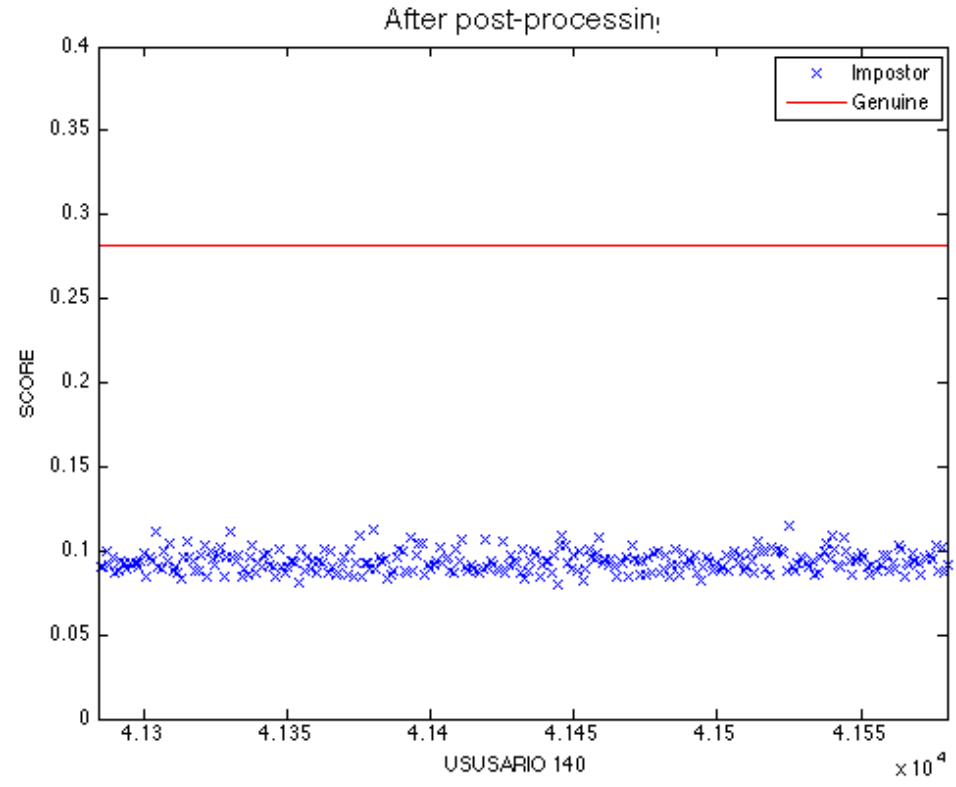

Figura 72 – Comparación de los scores de impostores y us. genuino 140 tras el procesado

4.3.1 Rendimiento del sistema

Las medidas de rendimiento del sistema pueden referirse al modo de funcionamiento de lista ordenada o identificación, o al modo de rendimiento de autenticación o verificación. Para el modo identificación, se suele utilizar el llamado Rank Identification Accuracy, representado por cuervas CMC. Para el modo autenticación, el valor de tasa de igual error (EER) suele ser utilizado para expresar la tasa de falsa aceptación cuando esta es igual a la tasa de falso rechazo, lo cual ocurre para un valor determinado de umbral de decisión.

4.3.1.1 EER

En los sistemas biométricos las tasas de Falsa Aceptación y Falso Rechazo definen el rendimiento del sistema.

El EER (Equal Error Rate) es una medida de definición de la calidad de un sistema de identificación ya que muestra el punto de intersección entre las curvas de falsa aceptación y falso rechazo. Es decir, el valor para el cual el sistema acepta y rechaza un individuo erróneamente de igual manera. Cuanto más bajo es el EER, se considera que el sistema posee un mayor poder de discriminación, ya que fallará menos veces en las decisiones de aceptación/rechazo que tome.

En este caso, los resultados obtenidos antes y después del procesado de las imágenes mediante filtros de Gabor muestran una mejora considerable en el sistema:

	Antes del procesado	Después del procesado
EER	49,66%	6,376%

Figura 73 – Tabla resumen del EER del sistema

4.3.1.2 Curva CMC

Cuando disponemos de un sistema de identificación, cuya salida es una lista de candidatos, se utilizan curvas CMC (Cumulative Match Characteristic) para poder analizar de manera visual los resultados obtenidos. Estas curvas no tienen en cuenta los Scores de salida del sistema, sino la posición del candidato genuino en la lista devuelta por el sistema. En ellas se representa para cada posición de la lista, el porcentaje de identificación del usuario genuino para esa posición y todas las anteriores en todas las búsquedas realizadas para cada tipo de experimento [18]. Estas curvas son siempre crecientes, ya que consisten en la acumulación de candidatos genuinos según se aumenta la posición en las listas de candidatos. En un sistema ideal, la curva sería una recta horizontal en el 100% de porcentaje de aparición, ya que todos los candidatos genuinos aparecerían la primera posición. En general, para un sistema biométrico, la curva debería alcanzar siempre la ordenada del 100%, ya que si no lo hace el sistema no está incluyendo el candidato genuino en la lista, por lo que no podrá ser identificado.

La curva CMC de nuestro sistema, antes y después del procesado basado en los filtros de Gabor es la siguiente:

	R-1	R-5	R-10	R-15	R-20
Antes	0.006711	0.01678	0.04362	0.0604	0.06711
Después	0.8691	0.9195	0.9295	0.943	0.9497

Figura 74 – Tabla con valores de las gráficas CMC del sistema

- Resultados obtenidos antes del procesado:

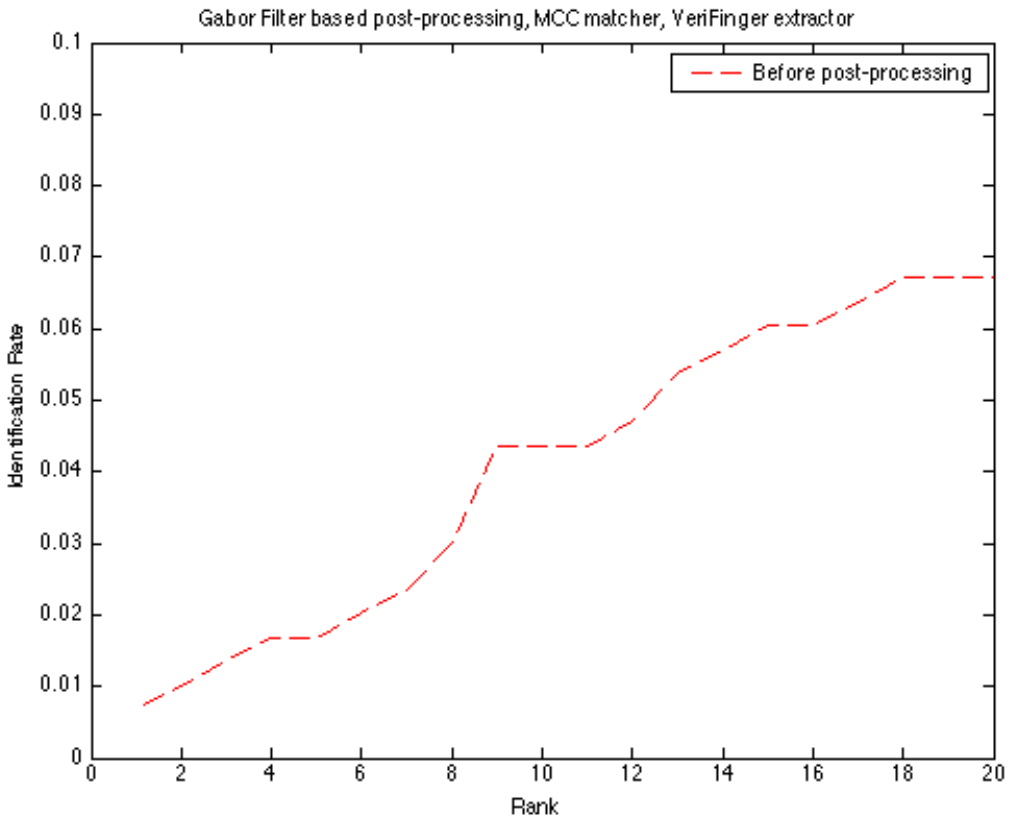

Figura 75 – Curva CMC del sistema antes del procesado

- Resultados tras el procesado:

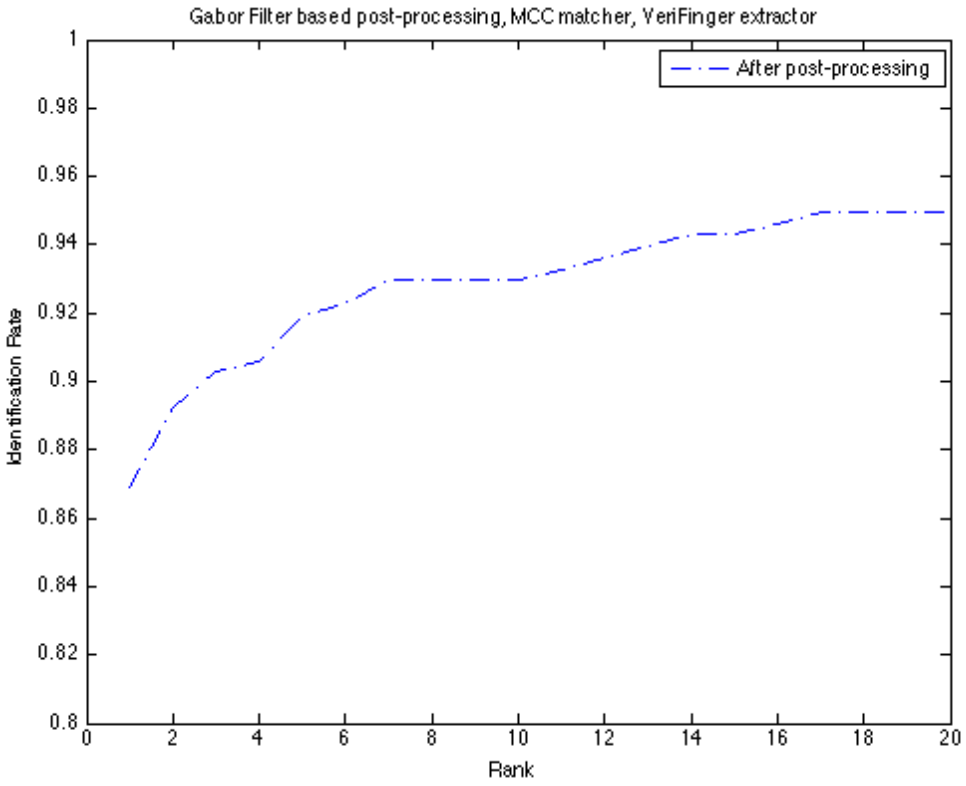

Figura 76 – Curva CMC el sistema tras el procesado

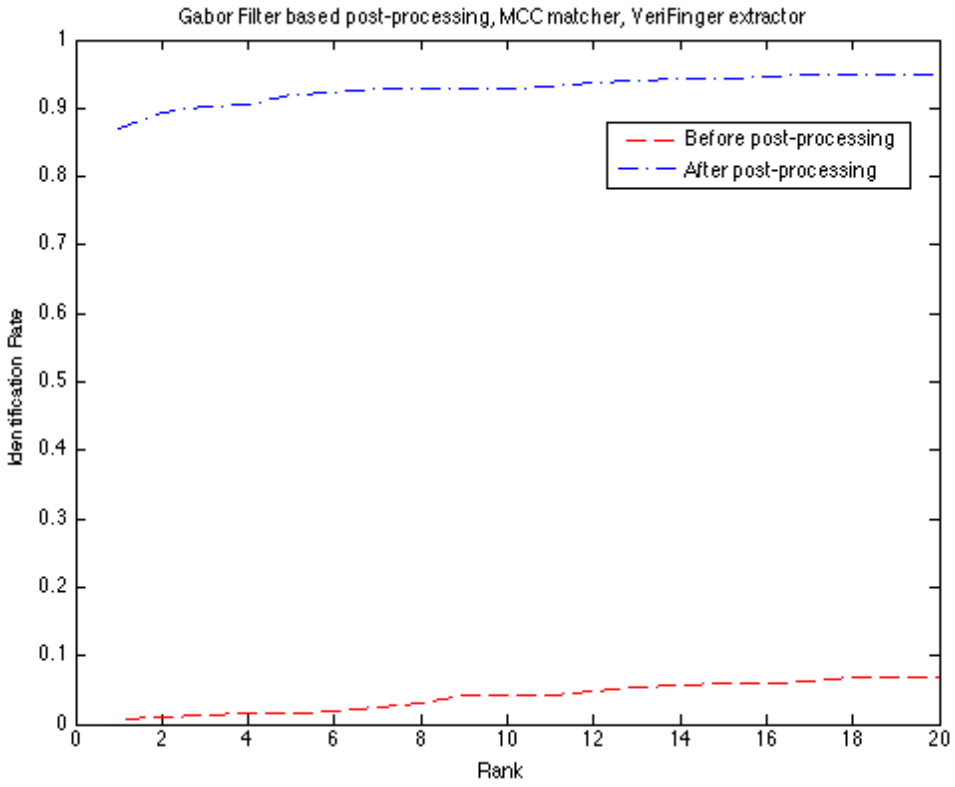

Figura 77 – Comparación de las curvas CMC del sistema antes y después del procesado

De esta gráfica se pueden extraer las siguientes conclusiones:

- Mientras que antes del procesado el porcentaje de encontrar el genuino entre los 20 primeros candidatos no superaba el 7%, tras el procesado es de un 94,97%.

- Aumenta considerablemente la probabilidad de acertar con el primer individuo identificado (pasa del 0,67% al 86,91%)

- Si bien es cierto que se ha utilizado software comercial para compara las huellas, es un software que claramente no está preparado para las huellas de la base de datos de la DGGC. A pesar de ser el mejor algoritmo para comparación de huellas que hay en el estado del arte, se hace totalmente necesario el paso de segmentación de la ROI de la huella dactilar, y selección de minucias dentro de dicha ROI.

5

Conclusiones y trabajo futuro

5.1 Conclusiones

Partiendo de los objetivos planteados al comienzo del proyecto, se puede afirmar que se han alcanzado los retos propuestos que a continuación se detallan:

- ✓ *Estudio del estado del arte en biometría forense, especialmente en sistemas automáticos de reconocimiento dactilar.*

  Se ha realizado un estudio del estado actual de los sistemas biométricos en general y más particularmente de los de reconocimiento de huella dactilar lo cual ha facilitado el desarrollo del PFC así como el entendimiento del área científica en el que se éste se localiza.

- ✓ *Evaluación del funcionamiento del sistema de identificación dactilar utilizando casuísticas y escenarios adaptados al trabajo diario del especialista forense, y con diversas herramientas de evaluación diferentes.*

  Se ha desarrollado dentro del proyecto una fase específica de trabajo práctico con el departamento de lofoscopia de la DGGC en la que se evaluaron los distintos métodos de trabajo.

- ✓ *Agilización en el proceso de adquisición de las evidencias mediante la creación de herramientas que faciliten la mecánica de la extracción de características en las huellas dactilares, así como el cálculo de LR.*

  Se detectaron las necesidades y se implementaron las mejoras necesarias en las herramientas de marcado de minucias y cálculo de relaciones de verosimilitud.

  Se ha digitalizado la base de datos disponible utilizando dichas herramientas, obteniendo como resultado una base de datos compuesta por 258 pares de huellas dactilares pertenecientes a 248 individuos distintos, y 20 pares de

huellas palmares de 14 individuos distintos. De cada una de las huellas se dispone además de un fichero que contiene las coordenadas, orientación y tipo de cada una de las minucias localizadas en la huella.

- ✓ *Mejora de algoritmos para la extracción de minucias mediante la aplicación de Filtros de Gabor a las imágenes de las huellas para la diferenciación de la región de interés.*

  La segunda fase de este proyecto se ha centrado en la mejora de herramientas de extracción automática de minucias y comparación de patrones de minucias. En concreto se ha desarrollado la fase de extracción de la región de interés en imágenes de huellas utilizando la base de datos adquirida en la fase anterior. Para ello se ha utilizado un método de aplicación de filtros de Gabor a las imágenes de las huellas para eliminación del fondo. Una vez extraída la región de interés de la huella se han eliminado las falsas minucias detectadas por el software de extracción de minucias automática (VeriFinger).

- ✓ *Realización de pruebas para evaluar los algoritmos de reconocimiento de huellas dactilares en condiciones forenses.*

  Tras la mejora llevada a cabo en el sistema, se ha aplicado a todos los archivos de minucias que componen la base de datos y se han realizado experimentos para comprobar su efectividad, dando como resultado una mejora considerable en el sistema que incrementa notablemente su funcionamiento.

5.2 Trabajo futuro

A partir del trabajo realizado se han detectado nuevas líneas de investigación para desarrollar en etapas posteriores:

- Creación de una base de datos mayor: se propone seguir aumentando el número de muestras de huellas en la base de datos creada con los nuevos cotejos que se vayan obteniendo.

- Aplicación del nuevo sistema mejorado a las huellas palmares de la base de datos.

- Estudio de la aplicación de este modelo de extracción y comparación de patrones de minucias a imágenes de huellas dactilares forenses extraídas en condiciones no controladas. Es decir, aplicación a las imágenes de la base de datos dubitadas, obtenidas directamente de la escena del crimen.

- Estudio de los resultados del funcionamiento del sistema biométrico en modo reconocimiento centrándose en la elección del umbral de decisión.

- Comprobación del funcionamiento del sistema sobre la base de datos dactilar una vez se haya aumentado su volumen considerablemente.

- Mejoras en las herramientas de marcado de minucias y obtención de LR tanto a nivel de interfaz como a nivel de sus algoritmos de funcionamiento.

- Búsqueda de nuevos métodos de extracción de la región de interés de la huella más efectivos en cuanto a carga computacional.

Referencias

[1] Galton, F. Finger Prints. Macmillan, London. , 1892.

[2] Champod, C. and Evett, I.W. "A probabilistic approach to fingerprint evidence". Journal of Forensic Identification, pages 101-122. Vol. 51(2), 2001.

[3] Maltoni, D., Maio, D., Jain, A.K. and Prabhakar, S. Handbook of fingerprint Recognition. Springer, 2009.

[4] Ratha, N., Karu, K. and Chen, S. "A real time matching system for large fingerprint database". IEEE Trans.on Pattern Analysis and Machine Intelligence, pages 799-813. Vol. 18, 1996.

[5] Dessimoz, D. and Champod, C. "Linkages Between Biometrics and Forensic Science". [ed.] A.K. Jain, P.J. Flynn and A. Ross. Handbook of biometrics. Springer, New York, 2007.

[6] Champod, C. "Forensic Applications, Overview". Encyclopedia of Biometrics. S.Z. Li, editor. Springer, 2009

[7] Jain, A.K., Flynn, P. and Ross, A.A. editors. Handbook of biometrics. Springer, 2007.

[8] E. Henry. "Classification an Uses of Finger Prints", Routledge, London, 1900.

[9] Champod, C., Lennard, C., Margot, P., and Stoilovic, M. Fingerprints and Other Ridge Skin Impressions. CRC Press, 2004.

[10] Ratha, N. and Bolle, R. Automatic Fingerprint Recognition Systems. Springer, 2003.

[11] Jain, A.K., Hong, L., Pankanti, S. and Bolle, R. "An identity authentication system using fingerprints". Proc. IEEE, pages 1365-1388. Vol. 85(9), 1997

[12] Jain, A.K., Prabhakar, S., Hong, L. and Pankanti, S. "Filterbank-based fingerprint matching". IEEE Trans. Image Processing, pages 846-859. Vol. 9(5), 2000

[13] http://www.policia.es/org_central/cientifica/servicios/id_identificacion.html

[14] Ramos, D. Forensic evaluation of the evidence using automatic speaker recognition systems. PhD. Thesis. UAM, Madrid, 2007.

[15] Aitken, C.G.G. and Taroni, F. Statistics and the Evaluation of Evidence for Forensic Science. John Wiley & Sons, Chichester , 2004.

[16] http://download.neurotechnology.com/VeriFinger_SDK_Brochure_2014-04-17.pdf

[17] http://www.neurotechnology.com/verifinger.html

[18] Asbaugh, D.R. Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology. The CRC Press, Boca Raton, FL, 1999.

[19] R. Cappelli, M. Ferrara and D. Maltoni, "Minutia Cylinder-Code: a new representation and matching technique for fingerprint recognition", IEEE Transactions on Pattern Analysis Machine Intelligence, vol.32, no.12, pp.2128-2141, December 2010.

[20] R. Cappelli, M. Ferrara, and D. Maltoni, "Fingerprint Indexing based on Minutia Cylinder Code," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 5, pp. 1051 - 1057, May 2011.

[21] M. Ferrara, D. Maltoni and R. Cappelli, "Noninvertible Minutia Cylinder-Code Representation", IEEE Transactions on Information Forensics and Security, vol.7, no.6, pp.1727-1737, December 2012.

[22] http://kime25.tripod.com/avance.htm

[23] http://sistemascontrolesbiometricos.blogspot.com.es/2011/12/sistemas-biometricos.html

[24] http://www.mianamnesis.com/2011/12/

[25] http://www.monografias.com/trabajos43/biometria/biometria2.shtml

[26] http://insigniass.blogspot.com.es/2011/09/dactiloscopia.html

[27] http://criminalistica.mx/areas-forenses/dactiloscopia/343-la-huella-perdida-identificaciersonal-utilizando-un-dedo

[28] http://www.papiloscopistas.org/forum/viewtopic.php?f=12&t=517

[29] http://oncedemarzo.net/ahmidan-el-origen/

Anexos

PRESUPUESTO

1) Ejecución Material

- Compra de ordenador personal (Software incluido)..... 2.000 €
- Alquiler de impresora láser durante 6 meses..... 50 €
- Material de oficina .....150 €
- Total de ejecución material ..... 2.200 €

2) Gastos generales

- 16 % sobre Ejecución Material..... 352 €

3) Beneficio Industrial

- 6 % sobre Ejecución Material ..... 132 €

4) Honorarios Proyecto

- 1280 horas a 15 € / hora ..... 19200 €

5) Material fungible

- Gastos de impresión ..... 200 €
- Encuadernación ..... 20 €

6) Subtotal del presupuesto

- Subtotal Presupuesto..... 22104 €

7) I.V.A. aplicable

- 21% Subtotal Presupuesto..... 10525,6 €

8) Total presupuesto

- Total Presupuesto..... 22115,5 €

Madrid, abril de 2015

El Ingeniero Jefe de Proyecto

Fdo.: Fátima García Donday
Ingeniero de Telecomunicación

PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un sistema de mejora de algoritmos de reconocimiento de huella dactilar en entornos forenses. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrata" y anteriormente llamado "Presupuesto de Ejecución Material" que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.