



UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR



-PROYECTO FIN DE CARRERA-

CÁLCULO DEL PESO DE LA EVIDENCIA FORENSE UTILIZANDO SISTEMAS BIOMÉTRICOS

**María Puertas Calvo
Febrero 2010**

Cálculo del Peso de la Evidencia Forense Utilizando Sistemas Biométricos

AUTOR: María Puertas Calvo
TUTOR: Daniel Ramos Castro



ATVS Grupo de Reconocimiento Biométrico
(<http://atvs.ii.uam.es>)
Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid

Resumen

Para la realización del proyecto se ha hecho uso de un sistema de reconocimiento de huella dactilar en el estado del arte, trabajando en modo de identificación. Dicho sistema ha sido utilizado para la realización de comparaciones con huellas procedentes de casos forenses reales. En concreto se han utilizado tres conjuntos de 50 huellas para lanzar comparaciones: uno de huellas latentes, uno de impresiones dactilares posadas y uno de impresiones dactilares rodadas, siendo estos dos últimos correspondientes a los mismos 50 usuarios. Las comparaciones realizadas con estos conjuntos de huellas se han hecho frente a una base de tarjetas decadactilares de 2,5 millones de reseñas reales, en la que estaba incluida una o más huellas genuinas para cada huella lanzada. El sistema devuelve por cada comparación una lista de 15 scores de similitud, que serán utilizados en el desarrollo del proyecto.

Se han diseñado una serie de experimentos que muestren la variación en los resultados producida por cambios en los factores de variabilidad presentes en las comparaciones de huellas. Estos parámetros de variación seleccionados son el número de puntos característicos, el área de la huella, la tolerancia a la distorsión y la calidad de la huella. A partir de los resultados obtenidos, se ha realizado la interpretación de los mismos, analizando el efecto de variación de cada uno de los factores para las distintas colecciones de huellas.

Se propone la utilización del sistema en modo de verificación, haciendo uso de la umbralización con los scores de las listas de candidatos. A partir de esto se estudia el rendimiento del sistema y el poder de discriminación de los conjuntos de scores, analizando las tasas de falso rechazo y falsa aceptación junto con el EER y las curvas DET.

A continuación se normalizan los conjuntos de scores obtenidos en los experimentos mediante el método *T-Norm*. Esta normalización reduce el desalineamiento que existe entre los scores de distintas comparaciones aumentando el poder de discriminación del sistema en modo verificación, lo cual se comprueba de nuevo midiendo las tasas de error y representando las curvas DET.

Finalmente, se propone 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. En este proyecto, se calculan valores de LR a partir de los conjuntos de scores obtenidos en los experimentos con el sistema de identificación dactilar. Para el cálculo de LR se utilizan cuatro métodos distintos encontrados en la literatura y se aplican a los conjuntos de scores antes y después de la normalización. Para terminar, se representan curvas ECE para realizar un análisis exhaustivo de los LR calculados. Estas curvas muestran la calibración y la discriminación de los LR, que determinan cuán apropiada es una técnica de evaluación de evidencias para el análisis forense.

Palabras Clave

Biometría, reconocimiento de huella dactilar, sistemas biométricos, ciencias forenses, relaciones de verosimilitud, identificación forense.

Abstract

In this project, an automatic fingerprint identification system was evaluated. The system was used to make comparisons with fingerprints from real forensic cases. More specifically, three different sets of fingerprints were selected. The first one is a set of 50 latent fingerprints found in real crime scenes. The second and third sets are 50 plain impressions and 50 inked impressions from the same index fingers of 50 individuals. These sets of fingerprints were compared against a database of about 2.5 millions of ten-print cards. In this database, the genuine cards corresponding with all the fingerprints (latent and impressions) used as query are included. When a comparison finishes, the system provides a list of 15 similarity scores. These lists are the basis of the experiments in this project.

A series of experiments was designed in order to evaluate the differences between the results produced by changes in the variability factors selected for the comparisons. These parameters are the number of minutiae marked on the fingerprint, the fingerprint's surface, the distortion tolerance and the quality of the fingerprint. A graphic interpretation of the results was made and the effects of the parameter variation were analyzed.

Then, the use of the system in a discrimination mode is proposed in order to see the effects of using a threshold within the scores obtained in the experiments. The error rates were calculated (FA, FR and EER) and the DET curves were represented to show the system's performance and the discriminating power of the different sets of scores.

Next, the sets of scores obtained in the experiments are normalized using T-norm. Normalization reduces the misalignments in the score ranges that appear among different trials, so the discriminating power of the system is improved. The new error rates are calculated and DET curves are plotted in order to see the new performance.

Finally, the LR (Likelihood Ratio) methodology is proposed in order to assess the weight of the evidence. The computation of LR values emulates DNA profiling, adapting forensic decisions to a probabilistic framework. Moreover, the LR methodology fulfills the new needs of forensic individualization, applying transparent and testable procedures. In this project, LR values are calculated from the sets of scores obtained in the experiments with the fingerprint recognition system. For the computation of LR values, four different techniques found in the literature are applied to the scores before and after normalization. Then, the results are analyzed using ECE curves. These curves show both discrimination and performance of the calculated LR, helping the fact finder in the task of finding the correct answer to the source attribution problem.

Key Words

Biometrics, fingerprint recognition, forensic science, biometric systems, likelihood ratios, forensic individualization.

Gracias

Gracias, Javier Ortega, por darme la oportunidad de formar parte del ATVS y depositar tu confianza en mí desde el principio.

Gracias, Daniel Ramos, por hacer que mi trabajo sea más fácil y agradable, por hacerme reír en los días malos (y en los buenos) y por ser como eres. Sin tu ayuda y tu apoyo esto no habría sido posible.

Gracias, miembros del ATVS, por hacer más amenas las horas de trabajo en el laboratorio y por los buenos momentos fuera de éste. En especial gracias a Pedro Tomé (mi primer tutor) y a Julián Fierrez (por todos los consejos y tu ayuda con el *paper*). Gracias Almu y Ali, siempre chicas ATVS.

Gracias al Capitán Nicomedes Expósito, al Teniente Francisco Molinero y a todo el equipo de SAID de la Guardia Civil, por hacer posible este proyecto y por enseñarme tantas cosas.

Gracias a todos mis profesores, del colegio y de la universidad, por aportar cada uno un granito de arena y contribuir positivamente a formar a la persona que soy ahora.

Gracias a mis amigos de teleco, por estos años de Bernabeu, parque de Pio XII, los viajes, la Deisi, el césped, las escaleras, los pasillos, los laboratorios, la biblioteca... Lo bueno lo habéis hecho mejor, y lo malo, menos malo. No creo que pueda agradecerlo uno por uno, porque seguro que me olvido de alguien que no merece ser olvidado, así que gracias a todos, sabéis quiénes sois.

Gracias a todos mis amigos de Roche (MIKA2P), de Boston (el mejor año de mi vida) y de Madrid. Todos y cada uno sois muy importantes para mí y estáis siempre conmigo en mis pensamientos.

Gracias, AMIGAS: Rocío, Yera, Ana y Sara, porque sin vosotras no sería quien soy ahora, porque siempre habéis estado ahí, desde hace 10 años, que se dice pronto. Gracias por crecer conmigo.

Gracias a mi hermana Car, a mis abuelos, a mis tíos y a mis primos, porque con una familia como la mía la vida es mejor.

Gracias a mis padres, porque la educación que me habéis dado, y que me seguís dando cada día, es lo más valioso que una persona puede recibir en la vida.

María Puertas Calvo
Febrero 2010

Glosario de Acrónimos

- **ADN:** Ácido Desoxirribonucleico
- **AFIS:** Automated Fingerprint Recognition System
- **CMC:** Cumulative Match Characteristic
- **DET:** Detection Error Trade-Off
- **ECE:** Empirical Cross Entropy
- **EER:** Equal Error Rate
- **FA:** False Accept
- **FR:** False Reject
- **KDF:** Kernel Density Function
- **LR:** Likelihood Ratio
- **PAV:** Pool Adjacent Violators

Índice de Contenidos

Resumen	I
Abstract	III
Gracias	V
Glosario de Acrónimos	VII
1. Introducción.....	1
1.1 Motivación.....	3
1.2 Objetivos	4
1.3 Contribuciones de este PFC.....	5
1.4 Organización de la memoria.....	7
2. Estado del arte	9
2.1 Biometría	9
2.1.1 Características de los rasgos biométricos	9
2.1.2 Clasificación de los rasgos biométricos	10
2.1.3 La biometría en ámbito forense.....	11
2.2 Identificación dactilar en ámbito forense.....	14
2.2.1 La huella dactilar	14
2.2.2 Impresiones dactilares.....	16
2.2.3 Huellas latentes	17
2.2.4 El proceso de identificación	18
2.3 Sistemas biométricos.....	20
2.3.1 Tipos de sistemas biométricos	20
2.3.2 Sistemas de reconocimiento de huella dactilar.....	22
2.3.3 Sistemas forenses de identificación dactilar.	24
2.3.4 Errores de los sistemas biométricos	25
2.3.5 Poder de discriminación de un sistema biométrico	26
2.3.6 Normalización de scores	29
3. Análisis de la evidencia forense basado en relaciones de verosimilitud	31
3.1 ¿Por qué un cambio de paradigma?	31
3.2 Teoría bayesiana: cálculo de LR	34
3.3 LR vs. Scores	38
3.4 Representación de LR mediante curvas ECE.....	39
3.5 Métodos de cálculo de LR.	41
3.5.1 Regresión logística.....	41
3.5.2 PAV (<i>Pool Adjacent Violators</i>).....	41
3.5.3 KDF (<i>Kernel Density Function</i>).....	42
3.5.4 <i>Gaussian</i>	42
4. Marco Experimental	43
4.1 Sistema utilizado.....	43
4.2 Bases de datos de huellas utilizadas.....	44
4.3 Protocolo experimental.....	45
4.4 Pruebas realizadas	47
4.4.1 Pruebas realizadas con huellas latentes	47
4.4.2 Pruebas realizadas con impresiones dactilares.....	48

5. Análisis de los factores de variabilidad	51
5.1 Pruebas de variación del número de puntos característicos.....	51
5.1.1 Experimentos con huellas latentes	52
5.1.2 Experimentos con impresiones posadas	54
5.1.3 Experimentos con impresiones rodadas	56
5.2 Pruebas de variación de la tolerancia a la distorsión	58
5.2.1 Experimentos con huellas latentes	58
5.2.2 Experimentos con impresiones posadas	59
5.2.3 Experimentos con impresiones rodadas	60
5.3 Pruebas de variación del área de la huella.....	61
5.3.1 Experimentos con huellas latentes	61
5.3.2 Experimentos con impresiones posadas	62
5.3.3 Experimentos con impresiones rodadas	62
5.4 Pruebas de variación de la calidad de las huellas	63
5.4.1 Experimentos con huellas latentes	63
5.4.2 Experimentos con impresiones posadas	65
5.4.3 Experimentos con impresiones rodadas	67
6. Rendimiento del sistema en modo verificación	69
6.1 Distribuciones de scores	70
6.1.1 Experimentos con huellas latentes	70
6.1.2 Experimentos con impresiones posadas	71
6.1.3 Experimentos con impresiones rodadas	72
6.2 Curvas DET.....	73
6.2.1 Experimentos con huellas latentes	73
6.2.2 Experimentos con impresiones posadas	75
6.2.3 Experimentos con impresiones rodadas	76
7. Cálculo de relaciones de verosimilitud (LR).....	77
7.1 LR en experimentos con huellas latentes.....	78
7.2 LR en experimentos con impresiones posadas.....	83
7.3 LR en experimentos con impresiones rodadas.....	87
8. Conclusiones y trabajo futuro	91
8.1 Conclusiones	91
8.2 Trabajo futuro	95
Bibliografía.....	97
A. Presupuesto.....	101
B. Pliego de Condiciones	105
C. Paper enviado a ICPR 2010	111

Índice de Figuras

FIGURA 1.1.1 HUELLAS DEL CASO MAYFIELD. A LA IZQUIERDA HUELLA LATENTE ENCONTRADA EN UNA BOLSA. A LA DERECHA, IMPRESIÓN DACTILAR DE BRANDON MAYFIELD. [6]	2
FIGURA 2.1.1 TRANSFERENCIA DE LA EVIDENCIA [9]	13
FIGURA 2.2.1 PATRONES DE HUELLA. A) ADELTO, B) SINISTRODELTO, C) DEXTRODELTO, D) BDELTO	14
FIGURA 2.2.2 BIFURCACIÓN DE CRESTA (IZQUIERDA) Y TERMINACIÓN DE CRESTA (DERECHA)	15
FIGURA 2.2.3 DETALLES DEL TERCER NIVEL	15
FIGURA 2.2.4 EJEMPLO DE TARJETA DECACTILAR	16
FIGURA 2.2.5 HUELLA LATENTE SOBRE UNA SUPERFICIE	17
FIGURA 2.2.6 FOTOGRAFÍA DE HUELLA LATENTE REVELADA	17
FIGURA 2.3.1 DIAGRAMA DEL MODO DE REGISTRO DE UN SISTEMA BIOMÉTRICO.	20
FIGURA 2.3.2 DIAGRAMA DEL MODO DE VERIFICACIÓN DE UN SISTEMA BIOMÉTRICO.	21
FIGURA 2.3.3 DIAGRAMA DEL MODO DE IDENTIFICACIÓN DE UN SISTEMA BIOMÉTRICO.	21
FIGURA 2.3.4 DIAGRAMA DE BLOQUES DEL EXTRACTOR DE CARACTERÍSTICAS DE UN SISTEMA DE RECONOCIMIENTO DE HUELLA BASADO EN MINUCIAS.	22
FIGURA 2.3.5 EJEMPLO DE FAUNAGRAMA	27
FIGURA 2.3.6 EJEMPLO DE CURVA TIPPETT	28
FIGURA 2.3.7 EJEMPLO DE CURVA DET	28
FIGURA 2.3.8 EJEMPLO DE CURVA CMC	29
FIGURA 2.3.9 FAUNAGRAMAS CON SCORES ANTES Y DESPUÉS DE LA NORMALIZACIÓN.	30
FIGURA 3.2.1 EJEMPLOS DE LA APORTACIÓN DEL LR A LAS DECISIONES [9]	36
FIGURA 3.4.1 EJEMPLO DE CURVA ECE	40
FIGURA 3.5.1 MÉTODO KDF DE GENERACIÓN DE LR [9].	42
FIGURA 3.5.2 MÉTODO GAUSSIAN DE GENERACIÓN DE LR [9].	42
FIGURA 5.1.1 CMC PARA LAS PRUEBAS DE VARIACIÓN DE PUNTOS CARACTERÍSTICOS EN LATENTES	52
FIGURA 5.1.2 CMC DESGLOSADA PARA LA PRUEBA PCAUTO EN LATENTES	53
FIGURA 5.1.3 CMC DESGLOSADA PARA LA PRUEBA PCM EN LATENTES	53
FIGURA 5.1.4 CMC PARA LAS PRUEBAS DE VARIACIÓN DE PUNTOS CARACTERÍSTICOS EN IMPRESIONES POSADAS	54
FIGURA 5.1.5 CMC DESGLOSADA PARA LA PRUEBA PCAUTO EN POSADAS	55
FIGURA 5.1.6 CMC PARA LAS PRUEBAS DE VARIACIÓN DE PUNTOS CARACTERÍSTICOS EN RODADAS	56
FIGURA 5.2.1 CMC PARA PRUEBAS DE VARIACIÓN DE DISTORSIÓN EN LATENTES	58
FIGURA 5.2.2 CMC PARA PRUEBAS DE VARIACIÓN DE LA DISTORSIÓN EN IMPRESIONES POSADAS	59
FIGURA 5.2.3 CMC PARA PRUEBAS DE VARIACIÓN DE LA DISTORSIÓN CON IMPRESIONES RODADAS	60
FIGURA 5.3.1 CMC PARA PRUEBAS DE VARIACIÓN DEL ÁREA CON HUELLAS LATENTES	61
FIGURA 5.3.2 CMC PARA PRUEBAS DE VARIACIÓN DEL ÁREA EN IMPRESIONES POSADAS	62
FIGURA 5.3.3 CMC PARA PRUEBAS DE VARIACIÓN DEL ÁREA EN IMPRESIONES RODADAS	62
FIGURA 5.4.1 CMC DE VARIACIÓN DE CALIDAD PARA PRUEBA PCAUTO EN LATENTES	63
	XI

FIGURA 5.4.2 CMC DE VARIACIÓN DE CALIDAD PARA PRUEBA PCM EN LATENTES	64
FIGURA 5.4.3 CMC DE VARIACIÓN DE CALIDAD PARA PRUEBA PC12 EN LATENTES	64
FIGURA 5.4.4 CMC DE VARIACIÓN DE CALIDAD PARA PRUEBA PC8 EN LATENTES	65
FIGURA 5.4.5 CMC DE PRUEBAS DE VARIACIÓN DE LA CALIDAD PARA PCAUTO EN POSADAS	65
FIGURA 5.4.6 CMC DE PRUEBAS DE VARIACIÓN DE CALIDAD PARA PC12 EN POSADAS	66
FIGURA 5.4.7 CMC DE PRUEBAS DE VARIACIÓN DE CALIDAD PARA PC8 EN POSADAS	66
FIGURA 5.4.8 CMC DE PRUEBAS DE VARIACIÓN DE CALIDAD PARA PC12 EN RODADAS	67
FIGURA 5.4.9 CMC DE PRUEBAS DE VARIACIÓN DE CALIDAD PARA PC8 EN RODADAS	67
FIGURA 6.1.1 FAUNAGRAMAS DE LOS EXPERIMENTOS PCM CON HUELLAS LATENTES, SIN Y CON NORMALIZACIÓN.	70
FIGURA 6.1.2 CURVAS FA/FR SIN Y CON NORMALIZACIÓN PARA EXPERIMENTOS PCM CON LATENTES.	70
FIGURA 6.1.3 FAUNAGRAMAS DE LOS EXPERIMENTOS PCAUTO CON IMPRESIONES POSADAS, SIN Y CON NORMALIZACIÓN.	71
FIGURA 6.1.4 CURVAS FA/FR SIN Y CON NORMALIZACIÓN PARA EXPERIMENTOS PCAUTO CON IMPRESIONES POSADAS.	71
FIGURA 6.1.5 FAUNAGRAMAS DE LOS EXPERIMENTOS PCAUTO CON IMPRESIONES RODADAS, SIN Y CON NORMALIZACIÓN.	72
FIGURA 6.1.6 CURVAS FA/FR SIN Y CON NORMALIZACIÓN PARA EXPERIMENTOS PCAUTO CON IMPRESIONES RODADAS.	72
FIGURA 6.2.1 CURVAS DET PARA EXPERIMENTOS CON HUELLAS LATENTES.	74
FIGURA 6.2.2 CURVAS DET PARA EXPERIMENTOS CON IMPRESIONES POSADAS.	75
FIGURA 6.2.3 DET PARA EXPERIMENTOS CON IMPRESIONES RODADAS.	76
FIGURA 7.1.1 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PCAUTO PARA EXPERIMENTOS CON LATENTES	78
FIGURA 7.1.2 LR PARA LOS SCORES NORMALIZADOS EN PCAUTO PARA EXPERIMENTOS CON LATENTES	78
FIGURA 7.1.3 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PCM PARA EXPERIMENTOS CON LATENTES	79
FIGURA 7.1.4 LR PARA LOS SCORES NORMALIZADOS EN PCM PARA EXPERIMENTOS CON LATENTES	79
FIGURA 7.1.5 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PC12 PARA EXPERIMENTOS CON LATENTES	80
FIGURA 7.1.6 LR PARA LOS SCORES NORMALIZADOS EN PC12 PARA EXPERIMENTOS CON LATENTES	80
FIGURA 7.1.7 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PC8 PARA EXPERIMENTOS CON LATENTES	81
FIGURA 7.1.8 LR PARA LOS SCORES NORMALIZADOS EN PC8 PARA EXPERIMENTOS CON LATENTES	81
FIGURA 7.2.1 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PCAUTO PARA EXPERIMENTOS CON IMPRESIONES POSADAS	83
FIGURA 7.2.2 LR PARA LOS SCORES NORMALIZADOS EN PCAUTO PARA EXPERIMENTOS CON IMPRESIONES POSADAS	83
FIGURA 7.2.3 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PC12 PARA EXPERIMENTOS CON IMPRESIONES POSADAS	84
FIGURA 7.2.4 LR PARA LOS SCORES NORMALIZADOS EN PC12 PARA EXPERIMENTOS CON IMPRESIONES POSADAS	84
FIGURA 7.2.5 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PC8 PARA EXPERIMENTOS CON IMPRESIONES POSADAS	85

FIGURA 7.2.6 LR PARA LOS SCORES NORMALIZADOS EN PC8 PARA EXPERIMENTOS CON IMPRESIONES POSADAS	85
FIGURA 7.3.1 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PCAUTO PARA EXPERIMENTOS CON IMPRESIONES RODADAS	87
FIGURA 7.3.2 LR PARA LOS SCORES NORMALIZADOS EN PCAUTO PARA EXPERIMENTOS CON IMPRESIONES RODADAS	87
FIGURA 7.3.3 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PC12 PARA EXPERIMENTOS CON IMPRESIONES RODADAS	88
FIGURA 7.3.4 LR PARA LOS SCORES NORMALIZADOS EN PC12 PARA EXPERIMENTOS CON IMPRESIONES RODADAS	88
FIGURA 7.3.5 LR PARA LOS SCORES DADOS POR EL SISTEMA EN PC8 PARA EXPERIMENTOS CON IMPRESIONES RODADAS	89
FIGURA 7.3.6 LR PARA LOS SCORES NORMALIZADOS EN PC8 PARA EXPERIMENTOS CON IMPRESIONES RODADAS	89

1

Introducción

Desde principios del siglo XX el reconocimiento dactilar ha sido utilizado en ámbito forense con el objetivo de identificar a criminales que dejan sus huellas dactilares en las escenas del crimen. Fue un poco antes, en 1892, cuando Galton [1] presentó los axiomas básicos del reconocimiento dactilar acerca de unicidad, permanencia y clasificación de los patrones dactilares. La teoría de unicidad de la huella dactilar, que afirma que no existen dos huellas dactilares iguales y está basada en datos empíricos [2], está ampliamente aceptada en todo el Mundo [3].

Con el paso de los años son muchos los avances que se han realizado en el campo de la huella dactilar en particular, y la biometría forense en general. El crecimiento de las bases de datos forenses de huellas dactilares ha hecho que la indexación y la comparación manual de huellas sean cada vez más complicadas. Los avances de la tecnología han permitido la creación de sistemas automáticos de identificación dactilar, conocidos como AFIS [4]. 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].

La dactiloscopia forense se ha considerado una técnica infalible, casi perfecta y libre de errores a lo largo de décadas. Sin embargo, en los últimos años se ha comprobado que, como en todo, se pueden cometer errores también en este ámbito. Uno de los más destacados fue cometido en mayo de 2004, cuando los expertos del FBI acusaron al abogado de Oregón, Brandon Mayfield, alegando haber encontrado su huella dactilar en una bolsa relacionada con los atentados terroristas de Madrid. Tras unas semanas, las autoridades españolas verificaron que la misma huella pertenecía a Ouhnane Daoud, de origen argelino. Tres agentes del FBI habían examinado la huella cometiendo el mismo error: el análisis dactilar que había llevado al arresto de Mayfield era erróneo [6].

En el informe publicado por el departamento de justicia de EEUU acerca del caso Mayfield [6], se atribuía la falsa identificación a un error humano, alegando que un factor importante había sido que la búsqueda fue realizada frente a una base de datos de unos 470 millones de huellas, por lo que la probabilidad de que la huella latente se pareciera mucho a una de las huellas de la base de datos es muy alta. Es necesaria, por tanto, la realización de evaluaciones empíricas y rigurosas de los sistemas de reconocimiento dactilar, ya que pueden llevar fácilmente a cometer errores cuando se trata con bases de datos tan grandes.

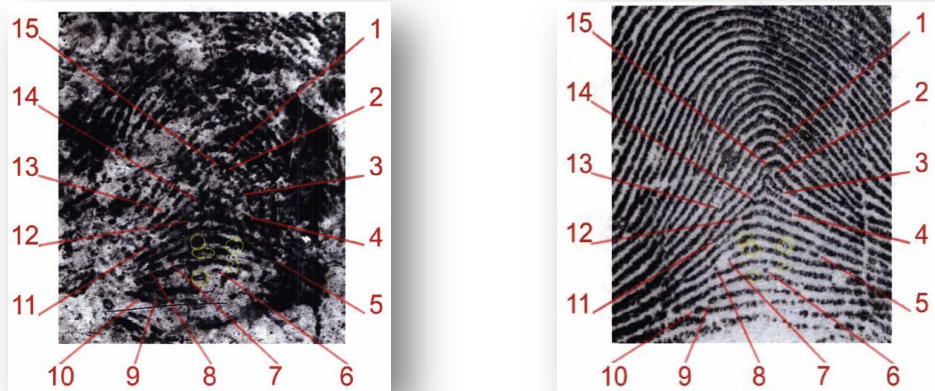


Figura 1.1.1 Huellas del caso Mayfield. A la izquierda huella latente encontrada en una bolsa. A la derecha, impresión dactilar de Brandon Mayfield. [6]

Por el procedimiento vigente de identificación dactilar forense, el resultado de la comparación por parte de un experto entre la huella anónima (latente o dubitada) dejada en la escena del crimen, y una impresión en tinta de la huella del sospechoso (indubitada) da un resultado determinista: identificación o exclusión. Sin embargo, este resultado está basado en la experiencia del experto y tiene un carácter subjetivo.

A raíz de la implantación de las reglas Daubert en Estados Unidos [7], para que una evidencia científica sea admitida en un juicio tiene que cumplir las siguientes premisas:

- 1) Estar basada en una prueba empírica en condiciones reales.
- 2) Que se emplee una técnica con precisión conocida.
- 3) Que dicha técnica esté revisada y publicada en foros científicos.
- 4) Existencia de estándares que definan el uso de dicha técnica.
- 5) Aceptación general de dicha técnica por parte de la comunidad científica.

La implantación de estas reglas junto con la aparición de errores de identificación en biometría forense han hecho que muchos científicos forenses de todo el mundo se cuestionen si las técnicas de evaluación de la evidencia forense utilizadas hasta este momento cumplen los requisitos de objetividad y transparencia exigidos por las reglas Daubert.

A raíz de todo esto, Saks y Koehler [8] proponen un *cambio de paradigma* en la identificación forense. Afirman que las ciencias de identificación forense necesitan dar un giro y tomar como modelo la identificación de ADN.

El modelo de identificación de ADN fue utilizado por primera vez en un juicio en la década de 1980. Tras un periodo de constante evolución y adaptación, en la actualidad los protocolos utilizados en ADN se tienden a considerar como la metodología modelo para el resto de disciplinas [9]. Los procedimientos utilizados en la identificación de ADN son sólidos y científicos. Además, los resultados se presentan en un marco probabilístico, evitando así los dictados de identificación o exclusión para determinar la fuente de la evidencia [10].

La identificación de ADN utiliza frecuentemente LR (*Likelihood Ratios*) para presentar resultados. Una de las grandes ventajas de la metodología LR es su universalidad. Es muy sencillo para cualquier disciplina forense calcular LR modelando estadísticamente la variabilidad entre las muestras analizadas, siempre y cuando existan datos representativos de la población de potenciales autores que se maneja. Es por esto que la metodología LR ha sido propuesta como estándar común para el análisis de la evidencia forense en todas las disciplinas [2] [10] [11], lo cual permite a la vez dos cosas: 1) que todas las disciplinas adapten sus procedimientos a un marco científico y probabilístico, y 2) que todas las disciplinas tengan procedimientos unificados y regulados, facilitando la tarea de los tribunales en la decisión final.

En este capítulo se detalla la motivación que lleva al desarrollo de este proyecto, seguida de los objetivos que se persiguen en la realización del mismo. A continuación se exponen las contribuciones originales desarrolladas en este PFC. Finalmente se expone de forma detallada el contenido de las distintas partes de este documento.

1.1 Motivación

Los sistemas forenses actuales de reconocimiento biométrico son una herramienta muy potente para ayudar a encontrar la fuente de una marca dejada en la escena de un crimen. Sin embargo, su funcionamiento se limita a comparar la huella lanzada con la base de datos, calculando una medida cuantitativa del parecido y devolviendo listas con los candidatos que más puntuación han obtenido. Además, la utilización de estos sistemas con bases de datos de millones de huellas facilita que se produzcan errores, tanto de falso positivo (porque una huella candidata sea muy parecida a la lanzada), como de falso negativo (porque la tipicidad, la mala calidad u otros factores de la huella hagan que el candidato genuino no aparezca en la lista de candidatos).

Desde nuestro punto de vista, los sistemas biométricos forenses pueden ser la base de la adaptación de las técnicas de reconocimiento forense a un marco científico y probabilístico, tal y como se ha descrito anteriormente. A partir de los scores devueltos por el sistema, pueden realizarse múltiples estudios probabilísticos acerca del funcionamiento del propio sistema, ayudando así a los usuarios del mismo a mejorar los resultados de las búsquedas.

Además, los resultados obtenidos del sistema en forma de listas de scores, pueden ser transformados, adaptándolos a un marco probabilístico tal y como se hace en identificación de ADN. De esta forma se darán herramientas a los especialistas que toman la decisión final que les ayuden a evaluar el peso de la evidencia forense.

1.2 Objetivos

Para la realización de este proyecto se ha realizado un estudio en el marco de un proyecto con el Departamento de Identificación de la Guardia Civil. En este estudio se ha utilizado un sistema de identificación dactilar en el estado del arte y bases de huellas procedentes de casos forenses reales, tanto anónimas recogidas de escenas de crimen como impresiones decadactilares de tinta.

Los objetivos de este estudio son:

- 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.
- Medida del impacto que tienen los factores de variabilidad de las comparaciones en el rendimiento del sistema de identificación dactilar para ofrecer al especialista indicaciones de bajo qué circunstancias la fiabilidad del sistema se degrada particularmente.
- Desarrollo de herramientas que traten los scores que devuelve el sistema en modo discriminación, aplicando un umbral de decisión. Para ello se realizará un paso previo de normalización, que reduzca el desalineamiento existente entre distintos conjuntos de scores procedentes de distintas comparaciones. Análisis del rendimiento del sistema cuando funciona en este modo.
- Desarrollo de herramientas que, mediante análisis estadístico forense basado en relaciones de verosimilitud (LR), obtenidas a partir de los scores que da un sistema de identificación dactilar, ayuden al especialista a evaluar el peso de la evidencia en relación con la identidad de los dactilogramas. Los valores de LR normalizan los scores que da un sistema biométrico en un marco probabilístico siguiendo métodos transparentes y científicos.

1.3 Contribuciones de este PFC

- Se realiza un estudio del estado del arte en biometría, sistemas biométricos y medida de rendimiento, reconocimiento de huella dactilar, identificación en ámbito forense y teoría bayesiana de cálculo de relaciones de verosimilitud.
- Se proponen una serie de protocolos experimentales que midan el rendimiento del sistema para distintos factores de variabilidad (número de puntos característicos, tolerancia a la distorsión, área de la huella y calidad). Se aplican los protocolos diseñados a tres conjuntos de huellas:
 - Huellas latentes: este es el grupo de máximo interés, debido a que es el que peor rendimiento presenta cuando trata con sistemas automáticos de identificación. Su imparcialidad, distorsión y mala calidad hacen más complicada la identificación de huellas latentes. Se realiza una comparación entre la extracción automática de minucias y la extracción manual por parte de un experto. Se compara también el rendimiento frente al de impresiones dactilares.
 - Impresiones dactilares rodadas: son tomadas en una situación controlada maximizando la calidad de la huella y su área. Se comparan los resultados obtenidos en los experimentos con los otros conjuntos de huellas.
 - Impresiones posadas: también tomadas bajo situación controlada, presentan menos área y menos distorsión que las rodadas. Se hace un análisis comparativo de los resultados frente a otros conjuntos de huellas.
- Se propone un nuevo modo de funcionamiento del sistema para, por ejemplo, su utilización para discriminar cuándo un par de huellas debe ser enviado a análisis dactiloscópico. Puede ser que haya coincidencias que presenten scores muy bajos y aparezcan al principio de la lista si no existen scores mejores. Ese efecto puede dar lugar a que se analicen dactiloscópicamente huellas que aparentemente son muy similares, pero que no lo sean. La solución a esto es utilizar la magnitud del score como indicador del parecido de las huellas, en lugar de la posición en la lista. Ese modo de funcionamiento propuesto se conoce como modo de discriminación o de verificación. Además, exige un paso previo de normalización, de cara a homogeneizar los rangos de variación de los scores y así poder utilizar el mismo umbral en todas las comparaciones. Esta normalización se ha aplicado sobre el conjunto experimental analizado, dando lugar a resultados excelentes.

- Se proponen varios algoritmos de cálculo de relaciones de verosimilitud (LR) a partir de los scores del sistema. El cálculo de LR se puede utilizar para aportar información útil al experto en la evaluación de la evidencia forense. Pero además de ello, se puede utilizar el LR calculado como un score normalizado que permita el uso del sistema biométrico en modo de verificación (ver punto anterior). De esta forma, se podría decidir mandar dos huellas a análisis dactiloscópico si y sólo si su LR es mayor que un determinado umbral. Si por ejemplo, ese umbral es 1000, sólo se mandarían a análisis dactiloscópico pares de huella cuya comparación supusiera un apoyo 1000 a 1 a la hipótesis de que ambas huellas provienen de la misma fuente. La facilidad de uso de un LR como discriminador a la hora de establecer umbrales sólo es posible gracias a su interpretación probabilística.

Una parte de este proyecto ha sido plasmada en un artículo científico de título “*Towards a Better Understanding of the Performance of Latent Fingerprint Recognition in Realistic Forensic Conditions*”. Dicho artículo ha sido enviado a la *20th International Conference on Pattern Recognition, ICPR 2010* y se encuentra pendiente de aceptación. El contenido del artículo se encuentra en el Anexo C de este documento.

1.4 Organización de la memoria

Esta memoria consta de los siguientes capítulos:

Capítulo 1. Introducción

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

Capítulo 2. Estado del arte

Este capítulo empieza con una introducción a la biometría donde se revisa la literatura al respecto, haciendo hincapié en la identificación dactilar en 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.

Capítulo 3. Análisis de la evidencia forense basado en relaciones de verosimilitud

Este capítulo comienza con una descripción de la estadística bayesiana y su aplicación a la evaluación de la evidencia forense. A continuación se dan los conceptos teóricos sobre teoría bayesiana y relaciones de verosimilitud, así como la forma de representarlos y calcularlos automáticamente por distintos métodos.

Capítulo 4. Marco experimental

Este capítulo describe los procedimientos y protocolos utilizados para el desarrollo de los experimentos en este proyecto, así como los sistemas y bases de datos utilizados.

Capítulo 5. Análisis de los factores de variabilidad

Capítulo en el que es analizado el funcionamiento del sistema de identificación dactilar para distintos escenarios entre los que se varían diferentes parámetros en las comparaciones. Los resultados se presentan en forma de curvas CMC.

Capítulo 6. Rendimiento del sistema en modo verificación

En este capítulo se introducen las ventajas de analizar el sistema de identificación en modo verificación para analizar así su poder de discriminación. Además, se realiza un paso previo de normalización de los scores, con el objetivo de aumentar el poder de discriminación de los conjuntos de scores. Se presentan los resultados en forma de faunagramas, curvas Tippett y curvas DET.

Capítulo 7. Cálculo de relaciones de verosimilitud

Se presentan los resultados de calcular relaciones de verosimilitud (*Likelihood Ratios*, LR) con los distintos métodos de cálculo a partir de los scores obtenidos de un sistema automático de identificación dactilar, con huellas anónimas e impresiones dactilares.

Capítulo 8. Conclusiones y trabajo futuro.

2

Estado del arte

2.1 Biometría

Al igual que las personas somos capaces de identificar a nuestros conocidos por su cara, fisionomía, forma de caminar, de hablar etc., la biometría busca métodos que sean capaces de identificar a las personas de forma automática [12].

La biometría es la ciencia que se ocupa del reconocimiento automático de personas a partir de sus características fisiológicas o de comportamiento [12]. Estas características se conocen como rasgos biométricos.

2.1.1 Características de los rasgos biométricos

Los rasgos biométricos son aquellas características intrínsecas a cada persona que la diferencian del resto y que pueden ser medidos y utilizados por un sistema de reconocimiento automático para distinguir al individuo de los demás.

Para que un rasgo personal sea válido para que un sistema biométrico sea capaz de distinguir a las personas a partir de él, debe cumplir las siguientes premisas:

- **Universalidad:** toda persona debe poseer dicho rasgo biométrico.
- **Unicidad:** personas distintas deben poseer rasgos diferenciados.
- **Permanencia:** el rasgo debe ser invariante con el tiempo a corto plazo.
- **Perennidad:** el rasgo debe ser perpetuo, es decir, invariante con el tiempo a largo plazo.
- **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 a 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.

2.1.2 Clasificación de los rasgos biométricos

En general, los rasgos biométricos se dividen en dos grandes grupos: rasgos fisiológicos y rasgos de comportamiento. Se conocen como rasgos fisiológicos aquellas características que son intrínsecas a la naturaleza física de la persona: huella dactilar, cara, iris...

Los rasgos de comportamiento son aquellos que denotan el modo en que una persona realiza una determinada acción: escritura, voz, forma de caminar...

A continuación se presenta una breve introducción a los rasgos biométricos más comunes:



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 la vida. Esto hace de las huellas dactilares un rasgo biométrico muy atractivo para sistemas de reconocimiento. Su uso está muy extendido en aplicaciones comerciales, pero también en ámbito forense, en el que se trata de identificar criminales que dejaron sus huellas en la escena del crimen. La unicidad de las huellas dactilares está totalmente asumida pese a ser un hecho concebido a partir de datos empíricos.



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 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.



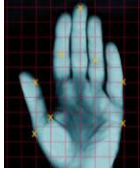
Cara: la cara es uno de los rasgos biométricos más aceptados, ya que es el más empleado por los humanos para reconocernos entre nosotros de manera visual. Una de las grandes ventajas es que el método de adquisición no es intrusivo, basta con tomar una imagen del rostro del individuo. Sin embargo, no cumple con los requisitos de permanencia debido a que sufre grandes cambios a lo largo de la vida del individuo. Las diferentes posibles condiciones de captura de la imagen por causa de la iluminación, posición, artefactos... hacen que el reconocimiento automático de caras sea un tema al que le queda mucho que perfeccionar.



Voz: este rasgo es una combinación de características físicas (fisonomía del tracto vocal) y de comportamiento (ritmo, acento). Estas últimas no son invariantes a lo largo del tiempo y hacen difícil reconocer dos muestras de voz de un individuo muy espaciadas en el tiempo. Su obtención no requiere intrusión y es un rasgo muy aceptado socialmente para reconocimiento de personas. No obstante, el riesgo de aparición de impostores es grande al ser un rasgo fácil de robar y de imitar.

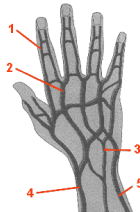


Firma: la firma es un rasgo de comportamiento que es característico de cada individuo. A lo largo de la historia, la firma ha sido el medio de identificación más común. Sin embargo, este rasgo presenta gran variabilidad a corto y largo plazo, y un alto riesgo de falsificación por parte de otros individuos, lo cual lo convierte en un rasgo difícil para un reconocimiento automático fiable.



Geometría de la mano

La longitud de los dedos y la separación entre ellos así como las distancias entre puntos clave de la mano hacen que este rasgo biométrico alcance cierto grado de unicidad. Para la adquisición es necesaria la cooperación del usuario, por lo que es considerado un método intrusivo. Es útil cuando se dispone de poco espacio de almacenamiento de datos, ya que las imágenes necesitan poco espacio en memoria para ser almacenadas.



Venas de la mano: la geometría de las venas en una mano tiene un alto grado de unicidad y es a su vez un rasgo de alta seguridad debido a la gran dificultad de falsificación. Su adquisición puede hacerse por imágenes con iluminación infrarroja.

Además de los rasgos descritos anteriormente, existen otros rasgos menos estudiados cuyo uso se está extendiendo en el mundo de la biometría. Por ejemplo, la forma de la oreja, la forma de caminar, la escritura, la dinámica de tecleo o el olor.

2.1.3 La biometría en ámbito forense

Actualmente, la biometría se usa en todos los campos donde se requiere la identificación de personas. Es común encontrar un sensor de huella para el control de acceso a un edificio, o para acceder a una cuenta en un ordenador portátil. En algunos aeropuertos el personal entra en zonas de acceso restringido gracias a un reconocedor de su iris. Pero aparte de las aplicaciones comerciales y de seguridad, en este proyecto nos interesaremos por las aplicaciones forenses de la biometría.

Empezaremos introduciendo el concepto 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] [13].

La biometría es una parte muy importante dentro de las ciencias forenses ya que ayuda a establecer la identidad de una persona o a asociar a un individuo con una fuente desconocida. Los científicos forenses han demostrado que los rasgos físicos y de comportamiento pueden informar acerca de la identidad de personas implicadas en crímenes [13].

Como ya se ha definido, la biometría consiste en la utilización de sistemas automáticos para identificar personas a partir de rasgos característicos [12]. Sin embargo, la biometría forense hoy en día no persigue la identificación automática de personas a partir de las marcas dejadas en la escena del crimen, sino un filtrado automático de fuentes potenciales de la marca, a partir del cual se realizan comparaciones más exhaustivas por parte de expertos humanos [5].

Otra diferencia significativa 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 hacen que la variabilidad entre las marcas dejadas por un mismo individuo sea mucho menor que en los casos de biometría forense, en los que las marcas son dejadas accidentalmente en escenas del crimen. En biometría forense la calidad de las marcas biométricas es mucho más pobre, por lo que el poder de discriminación disminuye considerablemente.

En general, la biometría forense se usa en las siguientes situaciones [14]:

- 1) Cuando una persona desconocida, viva o muerta, necesita ser identificada.
- 2) Cuando se encuentran marcas de rasgos biométricos en un lugar de interés y se requiere averiguar la identidad de su fuente.
- 3) Cuando se quiere relacionar dos o más marcas biométricas averiguando si pertenecen a la misma fuente.

En las primeras dos situaciones es necesaria la comparación de material biométrico proveniente de una fuente desconocida y de una fuente conocida. La tercera situación pretende relacionar dos marcas desconocidas aunque la fuente de éstas permanezca sin identificar.

Son numerosos los rasgos biométricos que se utilizan en la identificación de las posibles marcas o trazas encontradas en escenarios de interés:

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

A su vez, la biometría forense es una rama más de todas las técnicas de identificación forense, como la balística, que trata de identificar el arma con el que una bala ha sido disparada, o el reconocimiento de fibras, cristales, pintura y otros materiales. En todas estas disciplinas se estudia la evidencia forense.

La evidencia forense es la relación entre una marca cuya fuente es desconocida (por ejemplo, una huella anónima encontrada en un coche robado) y otro material originado por una fuente conocida (impresión de la huella de un sospechoso), ambos relacionados de alguna manera con un crimen u ofensa [10]. En cualquier caso, dispondremos de dos tipos de materiales.

- El primer tipo es conocido como el **material recuperado, muestra o marca**, y es transferido a la escena del crimen desde una persona implicada (voz grabada, huellas dactilares...) o viceversa (trozos de cristal encontrados en la ropa de un sospechoso).
- 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 (restos de una ventana rota) o directamente de un sospechoso (impresiones dactilares).

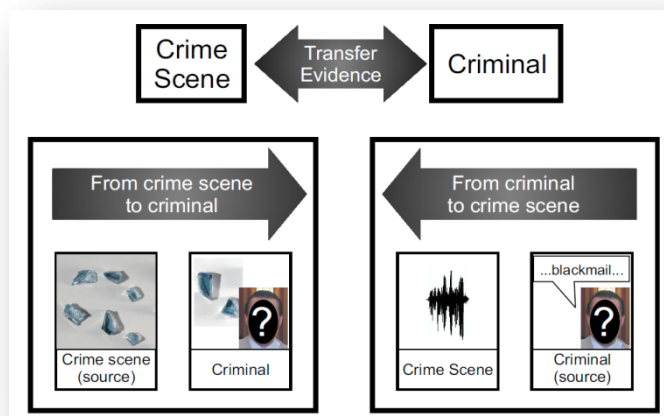


Figura 2.1.1 Transferencia de la evidencia [9]

El papel del científico forense es examinar el material recuperado y el material de control para evaluar la contribución de estos hallazgos a la decisión entre dos hipótesis contrarias. Cuando las hipótesis tratan sobre si la fuente de ambas muestras analizadas es la misma, estamos frente a un problema de atribución de fuentes [15], en el que se trata de responder a la pregunta: *¿Pertenece el material recuperado y el material de control a la misma fuente?*

2.2 Identificación dactilar en ámbito forense

Han pasado ya más de cien años desde que Alphonse Bertillon concibió la idea de utilizar rasgos físicos individuales con el objetivo de resolver crímenes [3] [13]. En 1893, el ministerio de interior del Reino Unido aceptó que no hay dos individuos que tengan huellas dactilares iguales. Pronto se empezaron a tomar las huellas de los criminales, para solucionar el problema de delincuentes reincidentes que utilizaban nombres falsos cada vez que eran detenidos, y empezaron a archivarlos para facilitar su posterior identificación. Además, con la comparación de estos registros con las huellas anónimas encontradas en escenas de crimen, 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 principios del siglo XX hasta ahora la identificación dactilar forense ha sufrido cambios muy importantes. Por un lado, la creciente movilidad de los criminales ha creado la necesidad de que las policías de diferentes regiones y países tengan que compartir las bases de huellas. Por otro lado, el problema del tamaño de las bases de datos hace difícil la clasificación e indexado de las reseñas y las huellas anónimas, y aun más las comparaciones. Sin embargo, gracias a los avances de la tecnología, hoy se dispone de sistemas informáticos que son capaces de almacenar imágenes de millones de huellas, que permiten que se compartan y se envíen en cuestión de segundos a cualquier parte del mundo, y que son capaces de hacer comparaciones de una huella frente a una base de datos de varios millones en cuestión de pocos minutos.

2.2.1 La huella dactilar

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

Nivel 1: el primer nivel es el más global y determina la forma general del dactilograma, conocida como fórmula o patrón general. Los puntos singulares a tener en cuenta en este nivel son el núcleo y los deltas. El núcleo es centro aproximado del dactilograma y el delta es el punto donde confluyen crestas orientadas en distintas direcciones. Según la aparición y distribución de núcleo y deltas, se tienen distintos tipos de dactilograma que se clasifican como:

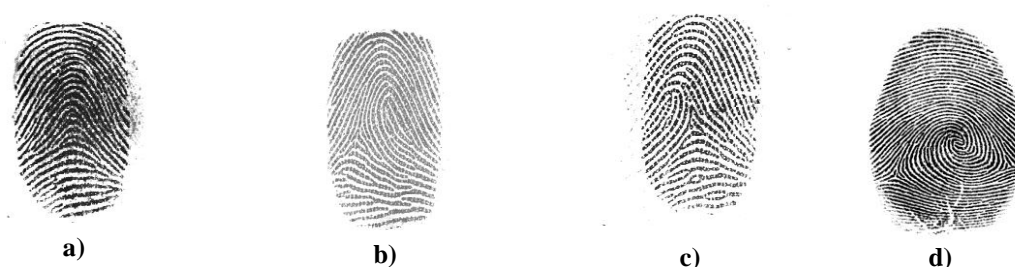


Figura 2.2.1 Patrones de huella. a) adelto, b) sinistrodelto, c) dextrodelto, d) bidelto

- **Adelto** o en arco: comprenden los dactilogramas que carecen de delta o de sistema nuclear.
- **Sinistrodelto:** reciben este nombre los dactilogramas que tienen sólo un delta a la izquierda del núcleo y tienen la cabeza del núcleo orientada hacia el mismo lado.
- **Dextrodelto:** reciben este nombre los dactilogramas que tienen sólo un delta a la derecha del núcleo y tienen la cabeza del núcleo orientada hacia el mismo lado.
- **Bidelto:** estos dactilogramas tienen dos deltas, generalmente a ambos lados del núcleo, que puede ser también doble en este tipo de patrones.

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

Nivel 2: es el nivel de características más utilizado en identificación comercial y forense. En este nivel, se definen y comparan puntos característicos de las crestas papilares. Los tipos de puntos característicos más importantes, conocidos como minucias, son dos: final de cresta (punto donde finaliza una cresta) y bifurcación de cresta (punto donde una cresta se divide en dos o más crestas). La localización de los puntos característicos y la distribución y orientación de éstos son la clave de la unicidad de las huellas dactilares. Las cicatrices se incluyen también en este nivel de características, pero se tiene en cuenta el carácter no permanente de éstas.



Figura 2.2.2 Bifurcación de cresta (izquierda) y terminación de cresta (derecha)

Nivel 3: es el nivel más detallista ya que se fija en las características internas de cada cresta. Los puntos característicos de este nivel incluyen la forma del contorno de las crestas, la anchura de éstas y los poros de la piel. Es el nivel menos estudiado ya que requiere que la calidad de la imagen sea excepcional para detectar estos detalles. Su uso a nivel automático es aún muy limitado pero puede utilizarse en comparaciones de huellas por parte de un experto.



Figura 2.2.3 Detalles del tercer nivel

2.2.2 Impresiones dactilares

Las impresiones dactilares son generalmente el material de control utilizado en ámbito forense. Cuando una persona es detenida por las fuerzas de seguridad, se le crea una reseña decadactilar. Estas reseñas han sido hasta la fecha tomadas sobre una cartulina haciendo uso de tinta, aunque cada vez más se usa un scanner conocido como *live scan*.

El formulario muestra los siguientes campos:

- APPLICANT:** Leave Blank
- TYPE OF PRINT ALL INFORMATION IN BLACK:** FIBER TONE
- FBI:** Leave Blank
- LAST NAME:** Leave Blank
- ALIAS:** AKA, O, A, I
- RESIDENCE OF PERSON FINGERPRINTED:** Leave Blank
- STATE OF BIRTH:** EDCB
- CITIZENSHIP:** PRP, US
- DATE OF BIRTH:** Leave Blank
- SEX:** M, F
- HAIR:** B, BR, BK, BL, BRN, GR, RD, WH
- EYES:** BLU, BRN, GRN, YEL, OTH
- FACE OF BIRTH:** F, M, F
- POB:** Leave Blank
- SIGNATURE OF OFFICIAL TAKING FINGERPRINTS:** Leave Blank
- CLASS:** Leave Blank
- EMPLOYER AND ADDRESS (if applicable):** Leave Blank
- REASON FINGERPRINTED:** Leave Blank
- SPONSOR IDENTIFYING:** 000-10-1111
- REASON IDENTIFYING:** Leave Blank

Las huellas dactilares se muestran en una cuadrícula de 3x5:

- Fila 1: 1. R. THUMB, 2. R. INDEX, 3. R. MIDDLE, 4. R. RING, 5. R. LITTLE
- Fila 2: 6. L. THUMB, 7. L. INDEX, 8. L. MIDDLE, 9. L. RING, 10. L. LITTLE
- Fila 3: Huellas de los pulgares de la mano izquierda (L. THUMB) y derecha (R. THUMB).

Figura 2.2.4 Ejemplo de tarjeta decadactilar

Las reseñas decadactilares constan de:

- Impresiones rodadas: impresiones de todos los dedos tomadas una por una rodando el dedo sobre la superficie del papel o del scanner.
- Impresiones posadas: huellas de los dedos índice, corazón, anular y auricular (meñique en dactiloscopia) posadas a la vez sobre el papel o el scanner, y de los pulgares posados individualmente.
- Marca de escritor: impresión posada del canto exterior de la mano
- Reseña palmar: impresión posada de la palma de la mano.

En el caso de las reseñas de tinta, éstas son escaneadas y digitalizadas, y tras añadir la información demográfica perteneciente al individuo, pasan a formar parte de la base de reseñas. En el caso de las reseñas tomadas por *live scan*, la digitalización de las reseñas se hace simultáneamente con la adquisición.

La captura de reseñas dactilares tiene que ser un proceso controlado por un especialista debido a que los individuos reseñados suelen negarse a colaborar fácilmente. Sin embargo, por norma general, las impresiones dactilares son imágenes que contienen una gran cantidad de información dactiloscópica gracias a la alta calidad y resolución de las imágenes. Además, al tomarse las huellas de todos los dedos por duplicado, la combinación de toda la información disponible en una tarjeta aumenta mucho el poder de las mismas para discriminar entre individuos.

2.2.3 Huellas latentes

Se usa el término latente para las huellas anónimas encontradas en escenas del crimen u otros escenarios de interés forense. Sin embargo, el término correcto debería ser huellas anónimas o marcas dactilares (en inglés, *fingermarks*). Dentro de este tipo de huellas podemos encontrar dos subconjuntos: las huellas visibles y las huellas latentes (ahora sí es el término correcto). Las huellas visibles o patentes son aquellas que han quedado marcadas sobre algún tipo de material (tinta, pintura, sangre...) que hace que la huella pueda ser percibida por el ojo humano sin ningún tipo de revelado. Las huellas latentes (palabra que viene del latín y que significa *escondido*) son aquellas que se marcan sobre las superficies gracias a las secreciones cutáneas, pero que son invisibles a los ojos humanos. Para ver estas huellas es necesaria una luz especial, y para su obtención, un revelado con productos químicos.



Figura 2.2.5 Huella latente sobre una superficie

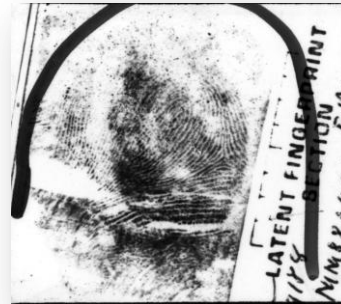


Figura 2.2.6 Fotografía de huella latente revelada

Las huellas anónimas constituyen el material recuperado dentro de la evaluación de la evidencia forense, ya que son aquellas cuya fuente es desconocida. Por lo general estas huellas son de baja calidad, tienen artefactos y no están completas. 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 en tinta.

2.2.4 El proceso de identificación

En el campo de la dactiloscopia, la palabra identificación es sinónimo de individualización y representa la declaración de certeza de que una marca particular fue hecha por las crestas papilares de la piel de un determinado individuo [16].

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 una discusión para crear un estándar numérico para la identificación forense de huellas dactilares, sugiriendo las siguientes reglas:

1. Si hay más de 12 minucias coincidentes y la huella anónima es nítida, entonces hay identificación (en ausencia de diferencias significativas).
2. Si existen entre 8 y 12 puntos coincidentes, la certeza 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, la orientación de éstas y el valor angular de las bifurcaciones.
3. Si existen menos de 8 minucias coincidentes, no se puede identificar la huella.

Estas reglas fueron ampliamente aceptadas por la comunidad dactiloscópica forense, aunque lamentablemente la tercera regla fue bastante ignorada [16].

Hoy en día el proceso de individualización de las huellas dactilares ha evolucionado bastante desde entonces y se ha formalizado en un proceso de 4 pasos conocido como ACEV (Análisis-Comparación-Evaluación-Verificación).

El procedimiento consiste en:

1. Análisis de la marca y de la impresión dactilar.
2. Comparación de ambas huellas
3. Evaluación y decisión sobre la identificación o la no identificación, basándose en las similitudes y diferencias entre las huellas.
4. Verificación de la decisión por parte de otro experto diferente.

Esta práctica varía entre continentes y países, y a veces incluso dentro de un mismo país. En particular, el paso de evaluación puede seguir 2 corrientes: la corriente del umbral cualitativo y la corriente del umbral cuantitativo.

El umbral cuantitativo se utiliza en la mayoría de los países europeos y sudamericanos. Esta corriente fija un número mínimo de minucias coincidentes entre la huella anónima y la impresión dactilar para la identificación de dicha huella, tal y como dicen las reglas de Locard. El número mínimo de minucias coincidentes es diferente para cada país, variando entre las 7 de Rusia y las 16 de Italia, siendo 12 el umbral en la mayoría de países, incluido España.

El umbral cualitativo es la corriente seguida en EEUU y se basa en un informe de la Asociación Internacional de Identificación (IAI) que afirma que “no existe una base científica para requerir un número mínimo predeterminado de características papilares que deban existir en dos impresiones para que se establezca una identificación positiva” [17]. A partir de este informe, se aceptó que el concepto de identificación no podía reducirse a contar minucias en las huellas. El argumento para ello fue 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 [16].

Una vez finalizado el proceso de individualización, se obtiene como resultado una decisión formal, sea cual sea la corriente utilizada en el paso de evaluación. Esta decisión se reporta en un informe pericial en el que se establece una de las tres opiniones cualitativas posibles: identificación, exclusión o inconcluyente. Para la toma de esta decisión determinista, los expertos hacen uso implícito de las probabilidades de rareza de las características de la huella que sostienen su opinión. Sin embargo, estas probabilidades se sustentan en la experiencia del experto y no en ninguna base científica.

Tal y como se ha demostrado en los últimos años, la dactiloscopia no es infalible, sino que en la toma de la decisión acerca de la identificación de una huella, pueden cometerse errores. Existen únicamente dos tipos de errores en este ámbito, ya que existen sólo dos tipos de decisión que el especialista puede tomar acerca de la identidad de una huella dactilar. El primero es una falsa individualización, o un falso positivo, que se comete cuando el experto afirma que una huella anónima y una impresión provienen del mismo individuo, cuando en realidad provienen de individuos distintos. Por otro lado está la falsa exclusión o falso negativo, que ocurre cuando un especialista decide que una marca y una impresión no corresponden al mismo individuo, cuando en realidad proceden del mismo individuo [6]. Para los expertos en dactiloscopia, el error de falsa identificación, o falso positivo, es mucho más grave que el de falso negativo ya que puede traer consecuencias desastrosas, como la condena de alguien inocente. Por el contrario, cometer un error de falso negativo se considera mucho menos importante y suele atribuirse a falta de información relevante en alguna de las huellas comparadas [16].

2.3 Sistemas biométricos

Los sistemas biométricos son sistemas automáticos de reconocimiento de patrones que intentan determinar la identidad de los usuarios a partir de características individuales fisiológicas o de comportamiento [12]. En general, un sistema biométrico comprende las funciones de:

- 1) Adquisición de datos: hace uso de un transductor para digitalizar el rasgo biométrico. De este proceso depende la cantidad y calidad de la información de la que dispone el sistema para aplicar en las siguientes fases.
- 2) Preprocesamiento: fase de acondicionamiento de la señal digital obtenida, para eliminar ruido u otros factores no deseados.
- 3) Extracción de características: se computan las características que diferencian al individuo del resto. Depende de la naturaleza del reconocedor.
- 4) Cálculo de la similitud: Se compara el o los modelos almacenados en la base de datos con la plantilla de características extraídas del usuario y se calcula un score o score, que supone una medida cuantitativa del parecido entre ambas muestras.
- 5) Toma de decisión: en los sistemas de verificación, si el score dado por el sistema para la comparación supera cierto umbral, se decide que el usuario es genuino. En caso contrario se decide que es un impostor.

2.3.1 Tipos de sistemas biométricos

Los sistemas biométricos constan de dos modos distintos de operación: un sistema biométrico puede ser o de verificación o de identificación. A su vez, existe otro modo de operación que es complementario y común a los dos anteriores, llamado modo de registro. A continuación se explica brevemente cada uno de los modos de operación:

Modo de registro: es el modo en el que los usuarios son dados de alta en el sistema. Para ello es necesario que el usuario introduzca su identidad y su rasgo biométrico. A continuación se realiza la extracción de características y se almacenan en a base de datos del sistema. En ocasiones se pide al usuario la aportación de su rasgo biométrico repetidas veces, para robustecer el sistema.

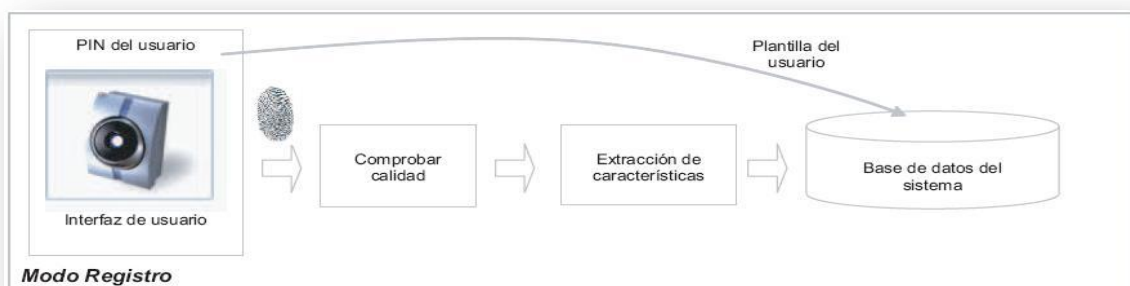


Figura 2.3.1 Diagrama del modo de registro de un sistema biométrico.

Modo de verificación: un sistema biométrico de verificación se encarga de confirmar a partir de un rasgo biométrico si un usuario es quien dice ser. Durante la fase de reconocimiento, el usuario hace una reclamación de identidad, por medio de una tarjeta o código. El sistema adquiere su rasgo biométrico y tras la extracción de características, lo compara con la plantilla almacenada en la base de datos para el usuario reclamado. El resultado es la decisión acerca de si el usuario es o no quien dice ser. Por lo 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”.

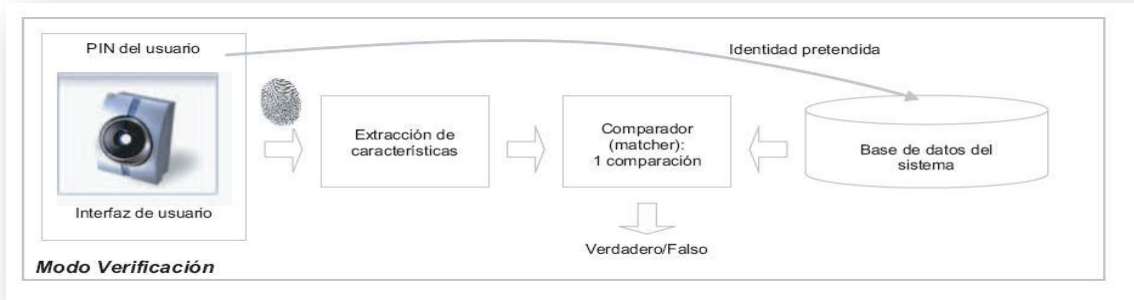


Figura 2.3.2 Diagrama del modo de verificación de un sistema biométrico.

Modo de identificación: en el modo identificación, el usuario no reclama ninguna identidad, sino que se comprueba si el usuario está o no entre los almacenados en la base de datos. Como el rasgo introducido se compara con todas la plantillas de los usuarios contenidos en la base de datos, se dice que es una comparación “uno a muchos”. La salida de un sistema de identificación puede ser o bien determinista (el usuario si/no se encuentra en la base de datos) o puede ser una lista de candidatos ordenados de mayor a menor score (puntuación de similitud). Al realizarse N comparaciones, siendo N el número de usuarios registrados, el coste computacional de este modo de operación es mucho mayor que en el modo de verificación. Sin embargo, este modo es necesario en sistemas en los que el usuario cuya identidad se busca no va a aportar información sobre su identidad, porque su objetivo es no ser identificado. Este es el caso más habitual de los sistemas forenses de reconocimiento biométrico.

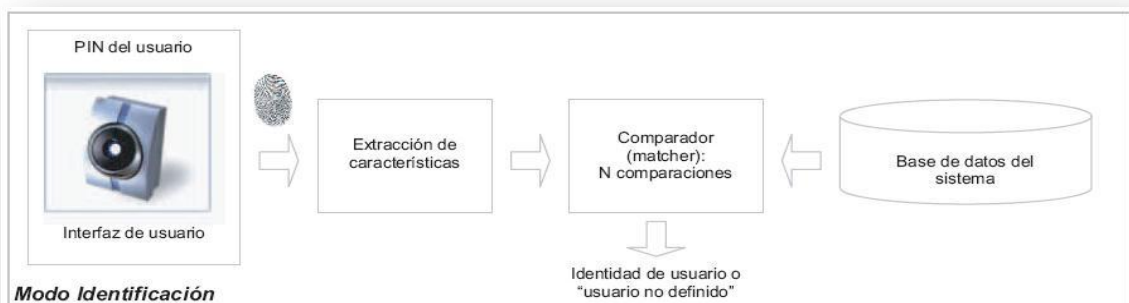


Figura 2.3.3 Diagrama del modo de identificación de un sistema biométrico.

2.3.2 Sistemas de reconocimiento de huella dactilar

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 [18]. Estas características no tienen por qué ser necesariamente minucias, ya que hay numerosos tipos de características que pueden ser utilizados en el proceso de identificación. En general, un sistema de reconocimiento de huella tiene dos partes diferenciadas: el extractor de características y el comparador.

Extracción de características:

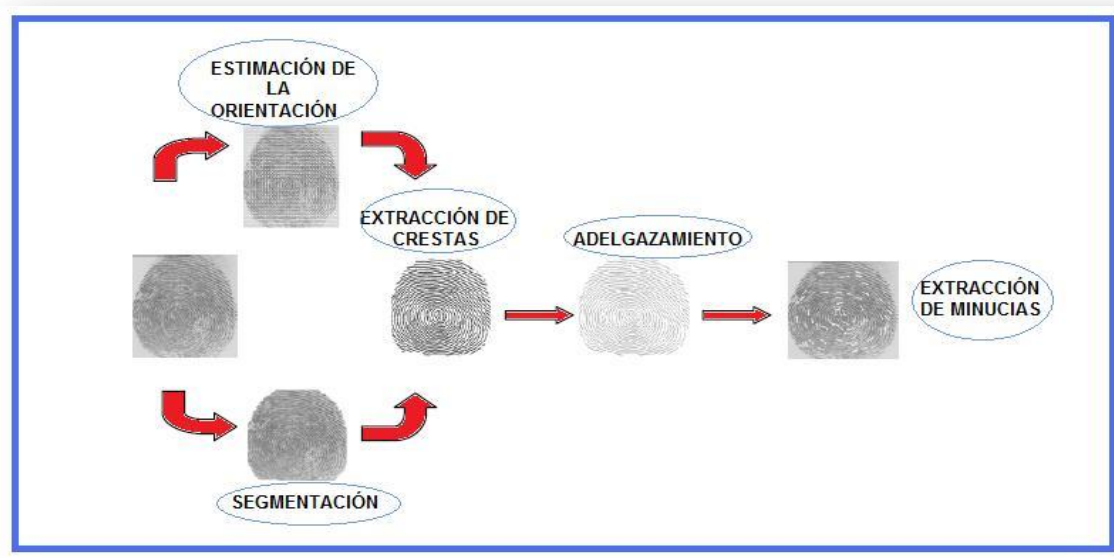


Figura 2.3.4 Diagrama de bloques del extractor de características de un sistema de reconocimiento de huella basado en minucias.

En la figura 2.3.4 se muestran los pasos del proceso de extracción de minucias de un sistema de reconocimiento de huella dactilar. Estos pasos se detallan a continuación [3]:

- 1) **Segmentación:** este proceso consiste en diferenciar la parte que se corresponde con la huella dactilar del fondo de la imagen. Esta segmentación se basa en la gran diferencia de nivel de gris que existe entre el fondo de la imagen y la huella dactilar, por lo que pueden diferenciarse ambas partes fácilmente.
- 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 pixel. 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 primer lugar se utilizan filtros para mejorar la claridad de las crestas y valles, de forma que facilite la extracción de características. Los filtros que suelen utilizarse con este propósito son filtros de Gabor, ya que se adaptan a la orientación frecuencial de las crestas. Una vez hecha la mejora, el sistema realiza la binarización de crestas, transformando la imagen que estaba en escala de grises en otra con sólo dos posibles valores: negro para las crestas y blanco para el resto.

- 4) **Adelgazamiento:** este procedimiento consiste en reducir la anchura de las crestas a un solo pixel en la imagen, ayudando a eliminar ruido y artefactos, y a realizar la extracción de minucias.
- 5) **Extracción de minucias:** se seleccionan todos los puntos en los que existe una terminación o bifurcación de cresta. Si un pixel negro sólo tiene un pixel negro a su alrededor, será una terminación de cresta. Si tiene tres pixeles negros a su alrededor, será una bifurcación. Posteriormente se realiza una etapa de post-procesado para eliminar minucias espurias, y agrupar minucias demasiado cercanas en una sola.

Comparador:

La comparación o *matching* 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. Uno gran motivo de la dificultad de este proceso radica en la gran variabilidad, no en el propio rasgo cuando pertenece a un mismo individuo, sino en la forma en que este rasgo ha sido calculado. Es decir, aunque el patrón de la huella de una persona permanece invariante, la forma en que esta huella es capturada puede sufrir cambios debido a presión, orientación, desplazamiento, curvatura de la superficie, estado de la piel y otros factores [12].

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 son muy dependientes de la calidad de las huellas. Por ello, la identificación automática con huellas latentes sigue siendo un tema que requiere mucha investigación.

En general existen tres grandes grupos de comparadores en sistemas de huella dactilar:

- **Comparador basado en minucias:** una vez realizada la extracción de minucias, se almacena una plantilla conocida como patrón de minucias. Al comparar dos huellas, son estos patrones los que se comparan en el sistema, dando lugar a una medida cuantitativa de similitud, conocida como *score*.
Antes de la fase de comparación, suele pasarse por un proceso de alineación de los dos patrones que sirva para tomar un punto de referencia que sea común a ambas huellas y a partir del cual empezar la comparación.
Existen numerosos métodos de comparación de patrones de minucias [4] [19]. Uno de los más populares consiste en pasar las minucias a coordenadas polares, tomando una minucia como punto de referencia. A continuación se ordenan las minucias en orden creciente de ángulo y distancia, obteniendo así cadenas que se van comparando entre los dos patrones y observando cuáles minucias son coincidentes y cuáles no. Con esto se calcula una medida global de similitud que da el *score*.
- **Comparador basado en textura:** este método utiliza el patrón del campo de orientación y de la frecuencia espacial de la imagen de la huella [18] [20]. Se emplean filtros de Gabor con diferente orientación. La ventaja de este método reside en que es más robusto en presencia de ruido o con imágenes de baja calidad, en las que la extracción de minucias se hace más complicada. El vector de características es constante y la carga computacional es menor. Sin embargo, para imágenes de buena calidad, este proceso tiene mayor tasa de error.

- **Comparador basado en correlación:** se superponen las imágenes y se calcula la correlación entre píxeles correspondientes. Cuando la correlación supera un umbral, se considera que ambas huellas comparadas provienen de la misma fuente. También existen sistemas que calculan la correlación como la multiplicación de ambas imágenes en el dominio frecuencial, aunque su coste computacional es mayor debido a la necesidad de conversión al dominio espectral. Otra opción es dividir la imagen en partes y calcular la correlación en cada sector, en lugar de en toda la imagen. En general las técnicas de comparación basadas en correlación se enfrentan a mayores problemas cuando las huellas no están alienadas o cuando existe deformación no lineal [18].

Hoy en día, los sistemas combinan las distintas técnicas explicadas de comparación, para hacer más robusta la comparación.

2.3.3 Sistemas forenses de identificación dactilar.

Como se ha explicado anteriormente, los sistemas biométricos utilizados en ámbito forense son en su gran mayoría sistemas de identificación. El motivo es que en las aplicaciones forenses, al contrario que en las comerciales, el individuo que aporta el rasgo biométrico no quiere ser identificado, por lo que su identidad es en principio desconocida.

Sin embargo, lo único que diferencia a un sistema de identificación de uno de verificación es el número de comparaciones de realiza el comparador cada vez que se inicia una búsqueda. En el caso de un sistema de verificación, sólo es necesaria una comparación, en cambio, en un sistema de identificación, se realizarán tantas comparaciones como plantillas de usuario existan en la base de datos. Otra diferencia significativa es que en los sistemas de identificación forense, cuando se lanza, por ejemplo, una huella anónima para buscar contra la base de huellas (plantillas decadactilares de impresiones), el resultado de la búsqueda es una lista de candidatos ordenados de mayor a menor score, mientras que en los sistemas comerciales, la salida suele ser una determinación sobre si el usuario se encuentra o no en la base de datos.

Por lo tanto, un sistema forense de reconocimiento dactilar será un sistema biométrico de identificación que recibe como entrada una huella anónima, y tras la extracción de características, la compara con las plantillas almacenadas en la base de datos, devolviendo como resultado una lista con los candidatos que más puntuación han obtenido en el comparador y que serán analizados posteriormente por un experto humano. Estos sistemas se conocen como AFIS (*Automated Fingerprint Identification System*).

La gran mayoría de los sistemas AFIS basan sus búsquedas en comparación de patrones de minucias, que contienen para cada punto característico, sus coordenadas, orientación y en algunos casos, la cuenta de crestas entre minucias. La gran ventaja de los sistemas AFIS es que son capaces de lanzar comparaciones tanto con una única huella como con una tarjeta decadactilar entera, cubriendo así todos los casos [5].

2.3.4 Errores de los sistemas biométricos

Aunque existen errores en las fases de adquisición y extracción de características, en este apartado nos centraremos en los errores del comparador.

Los errores en las comparaciones en un sistema biométrico se dan cuando el sistema no distingue adecuadamente si una comparación es *target* o *non-target*. Los motivos de la existencia de errores en las comparaciones son claros. Cuando se diseña un sistema biométrico, los parámetros se ajustan empíricamente, probando el sistema con bases de datos de rasgos biométricos reales y reajustando el sistema para que los resultados sean lo más correctos posibles. Sin embargo, ningún sistema puede probarse con todas las muestras posibles ni estar adaptado a todas las condiciones de adquisición, por lo que la existencia de fallos es difícilmente evitable. Para disminuir al máximo estos errores, los sistemas deben entrenarse con bases de datos grandes que modelen adecuadamente las variaciones que sufrirán las muestras biométricas para el funcionamiento operativo del sistema.

Los errores que los sistemas biométricos pueden cometer son dependientes del modo de operación de éstos.

Sistemas de verificación

Cuando se dispone de un sistema biométrico trabajando en modo verificación, existen dos tipos de errores que el sistema puede cometer a la hora de evaluar si dos muestras de material biométrico pertenecen o no a la misma fuente [12].

- **Error de Falsa Aceptación:** el sistema acepta la muestra de entrada como perteneciente al usuario al que dice pertenecer, cuando en realidad las muestras comparadas no pertenecen a la misma fuente. Esto quiere decir que el score a la salida del comparador supera el umbral de decisión pese a surgir de dos muestras pertenecientes a distintos individuos.
- **Error de Falso Rechazo:** el sistema no reconoce la muestra de entrada como perteneciente al usuario al que dice pertenecer. En otras palabras, la puntuación entre ambas muestras a la salida del comparador no supera el umbral de decisión siendo ambas muestras de la misma fuente.

Para medir el rendimiento del sistema, se calcula la frecuencia con la que un sistema comete estos errores para cada valor del umbral de decisión y un conjunto experimental de scores. Esto se conoce como Tasa de Falsa Aceptación (FAR) y Tasa de Falso Rechazo (FRR).

Además, existe un valor del umbral de decisión para el cual $FAR=FRR$. En este punto se define el EER (*Equal Error Rate*), punto donde ambas tasas de error son iguales y que suele definir el rendimiento del sistema con un único valor. En la sección siguiente se analizarán con más detalle los métodos de análisis del rendimiento de un sistema a partir de sus tasas de error.

Sistemas de Identificación

En sistemas de identificación es más difícil definir los tipos de errores porque depende de lo que devuelva cada sistema.

En general, cuando la salida del sistema es la aceptación o rechazo del rasgo introducido como perteneciente o no a un usuario existente en la base de datos, los errores que pueden cometerse son el de falsa identificación positiva, que ocurre cuando se acepta a un individuo no registrado, o falsa identificación negativa, que ocurre cuando un individuo registrado no es reconocido [21].

Sin embargo, en el caso de los sistemas de identificación forense, cuya salida es una lista de candidatos, existen las siguientes consideraciones:

- En comparaciones lanzadas con material de control, como por ejemplo, una tarjeta de impresiones decadactilar, se considera un error si el sistema no devuelve el candidato genuino en primera posición de la lista.
- En comparaciones lanzadas con material recuperado (por ejemplo, huellas latentes), se considera un error si el candidato genuino (considerando que el candidato genuino se encuentre en la base de datos) no aparece en la lista de candidatos.

2.3.5 Poder de discriminación de un sistema biométrico

Denominamos poder de discriminación del sistema a su capacidad para diferenciar las comparaciones genuinas u objetivo de las impostoras o no objetivo [9] [10].

Se entiende por comparación genuina o *target* la que se realiza con dos muestras de material biométrico que provienen de la misma fuente, y por comparación impostora o *non-target* la que genera scores con muestras provenientes de diferente fuente. Este poder de discriminación determina la fiabilidad del sistema cuando trabaja en modo verificación, es decir, el modo en el que se establece un umbral de decisión a partir del cual se considera si las muestras pertenecen o no a la misma fuente.

Para mostrar el poder de discriminación del sistema en los experimentos realizados, utilizaremos tres tipos de representación.

Faunagramas

El primer tipo de figura, conocido como faunagrama, sirve para analizar el comportamiento de un sistema cuya salida son las diferentes puntuaciones de las comparaciones entre el rasgo introducido y los almacenados en la base de datos. Este tipo de figura muestra el comportamiento de todos los scores de cada experimento, haciendo sencilla la visualización del alineamiento de los scores entre las distintas comparaciones. Cada fila del gráfico se corresponde con una lista de candidatos para una búsqueda distinta, y en cada una de las filas aparecen los scores correspondientes a la muestra genuina en forma de círculo azul, y los scores de muestras no genuinas en forma de cruz roja. El EER se muestra para cada comparación con un asterisco negro.

La figura 2.3.5 siguiente muestra un ejemplo de faunagrama.

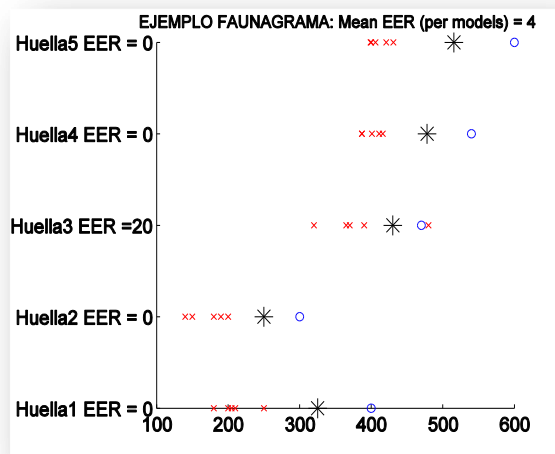


Figura 2.3.5 Ejemplo de Faunagrama

Curvas FA/FR, o curvas Tippett

En el segundo tipo de figura representaremos las curvas FA/FR, conocidas como curvas *Tippett* y utilizadas para mostrar la función de distribución acumulada de los scores. Estas curvas se basan en la aplicación de un umbral de decisión al conjunto de scores, por el cual se aceptarían como pertenecientes a la misma fuente las muestras que generaran scores por encima de dicho umbral, e impostoras las que puntúen por debajo del umbral. Los errores de FA y FR que se comenten al utilizar ese umbral para tomar decisiones se representan en función de dicho umbral, y son los siguientes [3]:

- La curva FA (Falsa Aceptación) representa la probabilidad de que el score dado por una huella no objetivo (comparación que genera un llamado score impostor) quede por encima del umbral de decisión. Esta curva es siempre decreciente, ya que al aumentar el umbral sólo podemos rechazar más scores de impostor, y no aceptarlos falsamente.
- La curva FR (Falso Rechazo) representa la probabilidad de que el score dado por una huella objetivo (comparación que genera un llamado score genuino) quede por debajo del umbral de decisión. Esta curva es siempre creciente, ya que al aumentar el umbral sólo podemos rechazar falsamente más scores genuinos, y no aceptarlos.

El punto en el que ambas probabilidades (FA y FR) se cruzan se denomina punto de *Equal Error Rate* (EER, tasa de error igual), y resume en un único número el poder de discriminación de un sistema biométrico para un conjunto determinado de prueba.

La figura 2.3.6 muestra un ejemplo de curva *Tippett*.

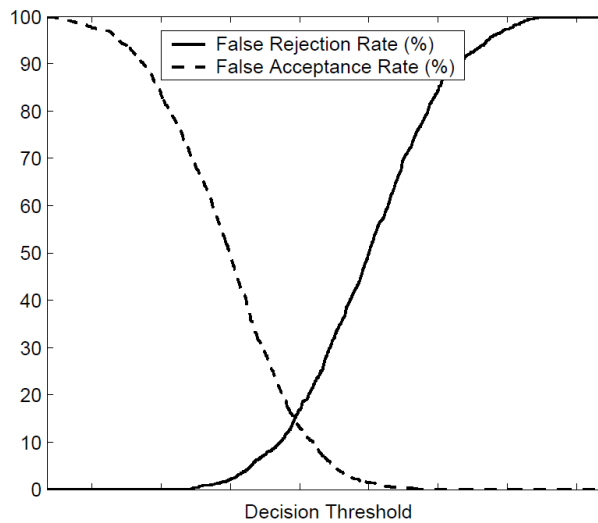


Figura 2.3.6 Ejemplo de curva *Tippett*

Curvas DET

Finalmente, se representan curvas DET (*Detection Error Tradeoff*). Estas curvas se calculan representando las probabilidades de FA y FR respectivamente en los ejes x e y de la gráfica, y se calculan fácilmente a partir de las curvas FA/FR. Estas curvas DET son muy útiles para ver de forma fácil y visual el poder discriminativo del conjunto de scores. Cuanto más cercanas están estas curvas al origen de coordenadas, mejor es el poder de discriminación del sistema para ese conjunto [22]. Además, el punto de EER se corresponde con el cruce de la curva DET con la bisectriz de los ejes, en la cual la probabilidad de FA y de FR se igualan. La figura siguiente muestra un ejemplo de curva DET. La curva más cercana al origen tiene mejor poder de discriminación.

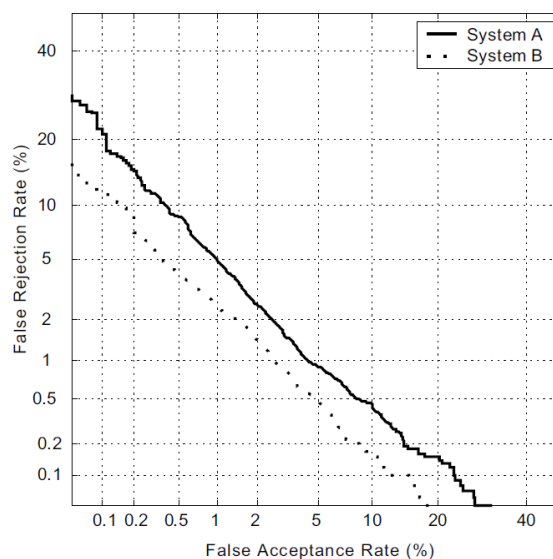


Figura 2.3.7 Ejemplo de curva DET

Curvas 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 [23]. La figura 2.3.8 muestra un ejemplo de curva CMC.

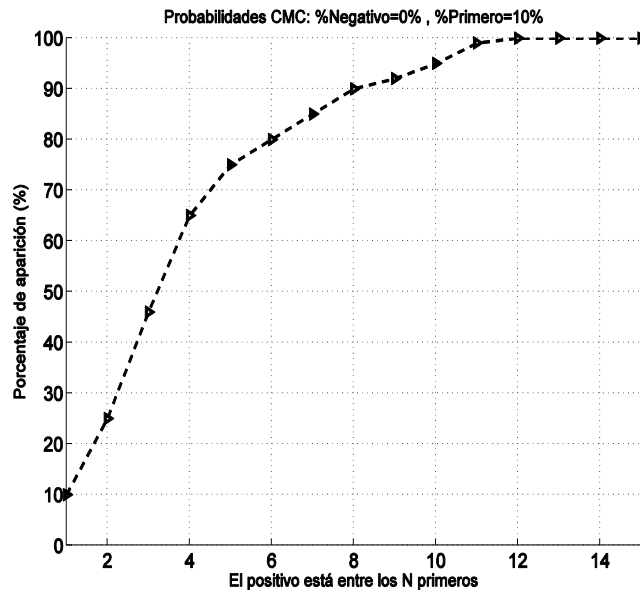


Figura 2.3.8 Ejemplo de curva CMC

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 en 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.

2.3.6 Normalización de scores

La normalización de scores es una técnica ampliamente aceptada para reducir el desalineamiento de los rangos de scores de salida de un sistema biométrico para diferentes comparaciones (entradas) [24]. Aunque existen muchas técnicas de normalización de scores, la más extendida es la llamada “*impostor-centric*” [25], en la que los parámetros para la normalización se estiman a partir de los scores obtenidos de las comparaciones de impostor o *non-target*, es decir, scores de comparaciones en la que ambas muestras no pertenecen a la misma fuente. El motivo de la popularidad de esta técnica es que para una determinada base de datos, existirán muchas más comparaciones *non-target* que comparaciones *target*, por lo que la normalización será mucho más robusta.

De entre todas las técnicas de normalización “*impostor-centric*”, en este proyecto vamos a describir el método *T-Norm*, o *Test Normalization*. Este método trata de alinear los scores de distintas comparaciones ajustándolos a una gaussiana de media cero y desviación típica uno. Para ello, se calcula la media y la desviación típica de un conjunto de entrenamiento de scores *non-target* calculados con la muestra biométrica de test y una cohorte de impostores, denotadas por μ_{Tnorm} y σ_{Tnorm} . A partir de estos datos se puede alinear cualquier conjunto de scores generados con dicha muestra de test aplicando la siguiente expresión:

$$S_{Tnorm} = \frac{S - \mu_{Tnorm}}{\sigma_{Tnorm}} \quad (2.1)$$

Donde S_{Tnorm} será el score normalizado y S es el score sin normalizar.

Cuando no se dispone de datos suficientes para tener un conjunto de entrenamiento y un conjunto de test diferenciados, podemos usar el conjunto de scores que queremos normalizar para calcular la distribución. Para ello, cada vez que se normaliza un score, debe calcularse μ_{Tnorm} y σ_{Tnorm} a partir de todos los demás conjuntos de scores *non-target* obtenidos de las comparaciones de las otras muestras. Esta técnica es útil cuando no se dispone de un conjunto de scores de entrenamiento, sin embargo la gran correlación entre los conjuntos de test y de entrenamiento hacen que la normalización sea más ideal que real. En la figuras 2.3.7 y 2.3.8, puede apreciarse el efecto de la normalización de los scores, alineando todos los scores *non-target* en un mismo rango.

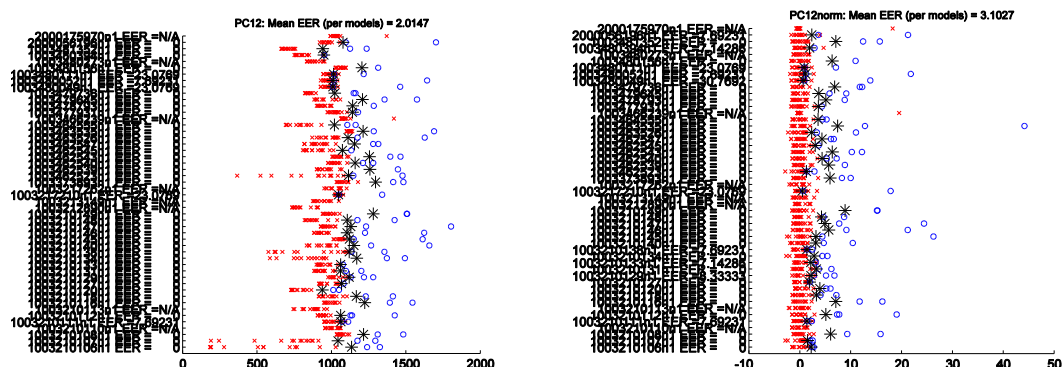


Figura 2.3.9 Faunagramas con scores antes y después de la normalización.

3

Análisis de la evidencia forense basado en relaciones de verosimilitud

3.1 ¿Por qué un cambio de paradigma?

Eventos recientes, como el caso Mayfield y otras condenas erróneas, han dado lugar a que algunos de los métodos utilizados en identificación forense se cuestionen por parte de juristas y científicos. Según Saks y Koehler [8], la identificación forense debe sufrir lo que ellos llaman un “cambio de paradigma” (*Paradigm Shift*), es decir, una transición de los procedimientos utilizados en identificación para adecuarlos a un método más científico que cumpla con unos requisitos de transparencia y testabilidad.

Los motivos que Saks y Koehler argumentan para justificar este cambio de paradigma se identifican a continuación:

Cambios en las leyes:

Las reglas Daubert [7] han establecido el primer paso para este cambio en EEUU. Estas reglas establecen que para que la evidencia sea admitida en un juicio, las técnicas utilizadas tienen que cumplir los siguientes requisitos:

- 1) Que se base en una prueba empírica en condiciones reales.
- 2) Que se emplee una técnica con precisión (o tasa de error) conocida.
- 3) Que dicha técnica esté revisada y publicada en foros científicos.
- 4) Que existan estándares que definan el uso de dicha técnica.
- 5) Que exista aceptación general de dicha técnica por parte de la comunidad científica.

Aunque estas reglas sólo se aplican en Estados Unidos, su aparición ha repercutido inmensamente en la comunidad a nivel mundial, ya que establecen un gran cambio conceptual en muchos ámbitos que además coincide con las opiniones de muchos expertos [8] [14]. Estas reglas facilitan la admisibilidad de pruebas de carácter científico, y a su vez dificultan la de metodologías no basadas en métodos científicos tales como testimonios basados en la experiencia de expertos humanos.

Condenas erróneas:

En los últimos años se han descubierto casos de condenas basadas en identificaciones erróneas, que han llevado a personas inocentes a prisión. Esto ha provocado que se cuestionen las técnicas de identificación forense, consideradas infalibles [26]. Uno de los casos más conocidos debido a la difusión en los medios de comunicación es el caso Mayfield, que llevó a una detención injusta debido a una identificación errónea de una huella dactilar relacionada con los atentados de Madrid en 2004 [6]. En [8], varios casos de identificación errónea han sido analizados para llegar a las causas de los mismos, y se ha llegado a conclusiones de que en un porcentaje significativo de los casos (63%) las causas parciales o totales han sido errores en la identificación forense.

Cálculo de tasas de error:

Aunque se incluye en las reglas Daubert, el cálculo de la precisión de una técnica en forma de tasas de error con conjuntos experimentales y bases de datos no es común en ciencia forense. La falta de investigación en este tema dificulta la tarea de establecer la precisión de los métodos utilizados en identificación forense.

ADN como modelo científico:

La metodología de identificación por ADN es una técnica relativamente novedosa comparada con otras disciplinas, como el reconocimiento dactilar, ya que sus orígenes datan de un siglo después. Desde sus orígenes en la década de 1980, la identificación genética es un método científico que evita en sus conclusiones la existencia de opiniones de expertos basadas en la experiencia y no en objetividad. Se trata de una metodología basada en procedimientos claros y transparentes que elimina los métodos no científicos. Además, los informes de identificación de ADN no son deterministas, es decir, eliminan las afirmaciones rotundas de identificación o exclusión que existen en otras disciplinas. En su lugar, se presentan informes probabilísticos objetivos basados en la información disponible y apoyados en un marco experimental y repetible.

Las opiniones probabilísticas de ADN se expresan en relaciones de verosimilitud dentro de un contexto Bayesiano, tal y como muchos expertos recomiendan [9] [10] [27].

A raíz de la identificación de estos obstáculos, surgen una serie de necesidades o cambios a realizar para que la identificación forense sea una técnica plenamente basada en métodos científicos:

1) **Transparencia de los procedimientos:**

La transparencia es esencial para que en los juicios se puedan evaluar los métodos e identificar y eliminar las posibles prácticas no científicas. La claridad en la presentación de los resultados forenses es esencial a la hora de evaluar el peso de la evidencia forense y la precisión de la disciplina forense en cada caso.

2) Testabilidad de las técnicas utilizadas:

La medida de la precisión de una disciplina científica forense debería estar basada en resultados experimentales representando condiciones reales en la medida de lo posible. La existencia y disponibilidad de los datos es fundamental a la hora de poder realizar experimentos repetibles. Son necesarias técnicas de evaluación comunes y compartición de recursos para llegar a estándares que faciliten la comparación y la mejora del rendimiento de las diferentes técnicas de identificación.

3) Precisión :

La precisión es el grado de conformidad de una cantidad medida o calculada con respecto de su valor verdadero. Es importante la selección de técnicas comunes de medida de precisión para evitar confusiones y malentendidos a la hora de presentar a un tribunal los resultados sobre la precisión de las distintas técnicas forenses.

4) Procedimientos comunes:

Es importante que los forenses adopten metodologías comunes a la hora de presentar los resultados a un tribunal. Este requerimiento es necesario para evitar confusiones debidas a la incongruencia de los resultados entre disciplinas o incluso dentro de una misma disciplina. Esta convergencia debe motivarse en todos los pasos del proceso de identificación forense.

Para la satisfacción de estas necesidades, pueden adaptarse los procedimientos de identificación forense tomando como modelo la metodología seguida en identificación por ADN. En esta disciplina, se sigue una corriente probabilística: el cálculo de LR. La metodología LR cumple los requisitos descritos en las reglas Daubert para la admisibilidad de la evidencia forense, aportando un apoyo probabilístico acerca del peso de la evidencia forense y evitando así las opiniones deterministas basadas en la experiencia.

3.2 Teoría bayesiana: cálculo de LR

En este apartado se introduce la teoría bayesiana [28] de relaciones de verosimilitud (en inglés *Likelihood Ratios* y a partir de aquí LR) y su aplicación a la biometría forense. En concreto, se computarán LR a partir de los scores de salida de un sistema biométrico de identificación. El objetivo del cálculo de LR es apoyar estadísticamente una de las hipótesis del problema de atribución de fuentes, y hacerlo de manera transparente y a partir de toda la información disponible.

La metodología LR se resume a continuación [10]:

Suponemos las siguientes hipótesis:

- **H_P** : Hipótesis de la fiscalía. Implica que el acusado sea el autor de las muestras incriminatorias. Según esta hipótesis, ambas muestras comparadas pertenecen a la misma fuente.
- **H_D** : Hipótesis de la defensa. Conlleva que el autor de las muestras incriminatorias es otro individuo distinto al acusado, es decir, que las dos muestras comparadas no pertenecen a la misma fuente.

Y las siguientes fuentes de información:

- **E** : Resultado de la comparación entre ambas muestras conocido como Evidencia. En un sistema biométrico, la evidencia sería el score.
- **I** : Información adicional del caso que no se incluye en la evidencia, pero que puede afectar al cálculo de las probabilidades. Ésta puede contener otra información proveniente de una evidencia distinta a la evaluada (E), o características conocidas del autor de la marca que no se muestran en la evidencia (como el sexo, la raza, la profesión, etc.), conocidas, por ejemplo, gracias a la declaración de un testigo. En general, I se utiliza para definir un conjunto de posibles fuentes de la marca, conocido como población potencial [29].

La probabilidad de que el sospechoso sea el autor de la muestra, para la que se cumple la relación:

$$P(H_P|E, I) = 1 - P(H_D | E, I) \quad (3.1)$$

Debido a que son sucesos complementarios, independientes y excluyentes. El teorema de Bayes relaciona las probabilidades de antes y después del análisis de la evidencia:

$$P(H_P|E, I) = \frac{P(E | H_P, I) P(H_P|I)}{P(E | I)} \quad (3.2)$$

Siendo igual la expresión para la hipótesis de la defensa:

$$P(H_D|E, I) = \frac{P(E | H_D, I) P(H_D|I)}{P(E | I)} \quad (3.3)$$

A partir de las ecuaciones anteriores llegaríamos a la expresión que relaciona ambas hipótesis antes y después de la evaluación de la evidencia mediante el LR [4]:

$$\frac{P(H_P | E, I)}{P(H_D | E, I)} = LR \cdot \frac{P(H_P | I)}{P(H_D | I)} \quad (3.4)$$

Siendo la relación LR la que aporta el peso de la evidencia para esta probabilidad:

$$LR = \frac{P(E | H_P, I)}{P(E | H_D, I)} \quad (3.5)$$

La ecuación (3.4) será la que relaciona las probabilidades de las hipótesis de la acusación y la defensa antes y después del análisis de la evidencia forense. En esta ecuación se demuestra que la relación final entre ambas probabilidades a posteriori, teniendo en cuenta tanto la evidencia forense (E) y el resto de la información del caso (I), se puede desglosar como el producto de dos factores:

- 1) El primero tiene en cuenta las probabilidades a priori, sin tener en cuenta la evidencia forense: $P(H_P | I) = 1 - P(H_D | I)$
- 2) El valor de LR, que es calculado por el científico forense a partir de la evidencia y de la información adicional [10].

El valor de LR es el cociente entre dos probabilidades, y se puede demostrar que es igual al cociente de dos densidades de probabilidad en el caso de variables continuas. En el numerador tendremos la función densidad de probabilidad (*fdp*) conocida como la distribución intra-fuente, que muestra la variabilidad de evidencias correspondientes a muestras pertenecientes a la misma fuente, es decir, la distribución correspondiente a comparaciones genuinas. En el denominador encontramos la distribución entre-fuentes, que es calculada a partir de comparaciones no genuinas y muestra la tipicidad o rareza de la muestra incriminatoria con respecto a una población relevante.

La tarea de los científicos forenses será calcular estas probabilidades siguiendo una metodología transparente y seleccionar para su cálculo una población relevante adecuada para el caso a partir de la información disponible (I).

Por lo tanto, antes del cálculo del LR, la única información es la dada por I y por lo tanto la relación entre ambas hipótesis será

$$\frac{P(H_p | I)}{P(H_D | I)}$$

de ello que se deduce que si $P(H_p | I) > P(H_D | I)$, entonces la relación anterior será mayor que uno y la hipótesis de la acusación será la predominante frente a la defensa.

En caso contrario, cuando $P(H_p | I) < P(H_D | I)$, será la hipótesis de la defensa la predominante.

Después del cálculo del LR a partir del análisis de la evidencia forense, las probabilidades a priori se modifican al multiplicarse por el valor del LR.

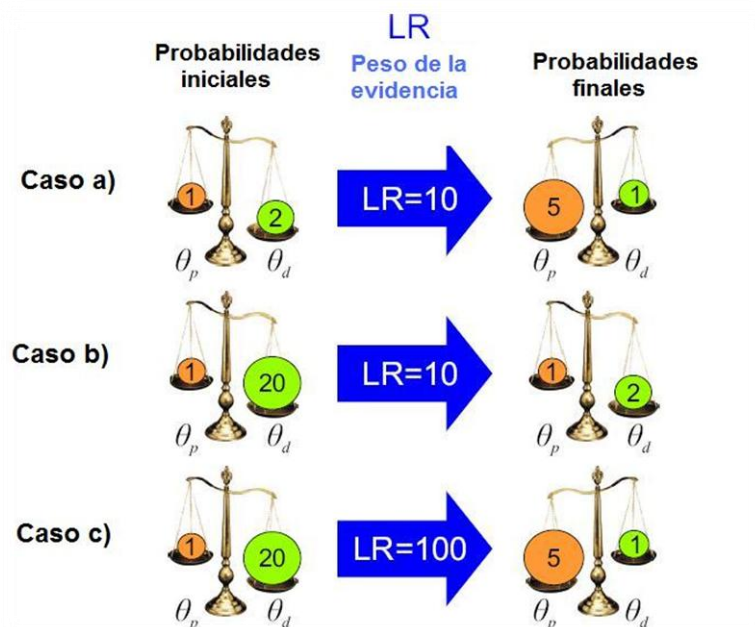


Figura 3.2.1 Ejemplos de la aportación del LR a las decisiones [9]

Como el LR es un factor que se multiplica por la relación inicial entre las hipótesis, es fácil deducir que:

- Si el $LR > 1$, entonces la evidencia apoyará la hipótesis de la acusación, es decir, las probabilidades de variabilidad intra-fuente, y de tipicidad de la muestra incriminatoria apoyan el que ambas muestras comparadas pertenecen a la misma fuente.
- Si el $LR < 1$, entonces la evidencia apoyará la hipótesis de la defensa, es decir, que ambas muestras comparadas no pertenecen a la misma fuente.

Finalmente, la decisión de un tribunal debe basarse en las probabilidades a posteriori, calculadas a partir de la combinación de las probabilidades a priori y del peso de la evidencia forense, representado en forma de LR. Si los errores cometidos al tomar decisiones tuvieran la misma importancia (lo cual no es el caso general), las decisiones se toman de la siguiente forma:

- Si $P(H_p | E, I) > P(H_D | E, I)$ se decidirá a favor de la hipótesis de la acusación.
- Si $P(H_p | E, I) < P(H_D | E, I)$ se decidirá a favor de la hipótesis de la defensa.

Este marco de cálculo de LR presenta numerosas ventajas en el campo de las ciencias forenses:

- Permite a los forenses evaluar y determinar el LR, un valor lleno de significado, ya que por sí mismo aporta el peso de la evidencia forense al caso.
- Define claramente el papel del científico forense, siendo éste el de evaluar cuál es el peso de la evidencia forense y dejando el papel de decidir sobre la culpabilidad o inocencia del acusado al tribunal encargado del caso.
- Las probabilidades pueden ser interpretadas como grados de certeza acerca del problema de atribución de fuentes, dejando a un lado las decisiones categóricas y subjetivas.

3.3 LR vs. Scores

Como se detalla en el capítulo 2, los sistemas biométricos de identificación (dactilar o de cualquier otro rasgo) en ámbito forense, devuelven una lista de candidatos. Estos candidatos aparecerán ordenados de mayor a menor score, siendo cada score una medida del parecido, obtenida de la comparación entre la muestra de entrada y cada candidato. Por lo tanto, al igual que los LR, los scores obtenidos de un sistema biométrico son también una forma de expresar cómo de parecidas son ambas muestras biométricas para así decidir si pertenecen o no a la misma fuente. Sin embargo es necesario establecer las diferencias entre estos dos tipos de valores.

Un score obtenido de la comparación de una muestra recuperada y una de control será un número real, que tendrá mayor valor cuanto más parecidas sean ambas muestras. Es decir, cuanto mayor sea el score, más se apoya la hipótesis de la acusación frente a la de la defensa. Sin embargo, un score por sí sólo no establece un grado de apoyo a ninguna de las dos hipótesis. Para ello es necesario aportar más información, como un umbral de decisión, el rango de variación de los scores del sistema utilizado o las distribuciones de scores de comparaciones *target* y *non-target*.

Sin embargo, un LR tiene un significado intrínseco que no necesita ir acompañado de otra información para expresar el grado de apoyo que da a cada una de las hipótesis [9]. Con un LR, se sabe que cuando es mayor que 1, la evidencia apoya la hipótesis de la acusación, es decir, que ambas muestras pertenecen a la misma fuente. Además, cuanto mayor sea este valor, mayor será el grado de apoyo a esta hipótesis. Por el contrario, si el valor de un LR es menor que 1, la hipótesis apoyada será la de la defensa, siendo mayor el apoyo cuanto menor sea su valor. Además, si expresamos este valor como una fracción, por ejemplo $LR = \frac{1}{2}$, la evidencia apoya 2 frente a 1 la hipótesis de la defensa. Si por el contrario el valor fuera $LR = 3$, puede decirse que el peso de la evidencia es de 3 para la hipótesis de la acusación frente a 1 para la defensa.

3.4 Representación de LR mediante curvas ECE

En este proyecto, se propone la utilización de curvas ECE (*Empirical Cross Entropy*) para representar la calibración y el poder de discriminación de los LR calculados. La entropía es un concepto creado en teoría de la información, y representa el grado de incertidumbre acerca de una variable determinada con respecto a la información conocida [30]. En nuestro marco forense, utilizaremos la entropía para representar la incertidumbre que existe en cada caso acerca del valor verdadero de las hipótesis [9].

Para cada probabilidad inicial (antes del análisis de la evidencia) en el rango [0,1], se computan las probabilidades finales utilizando los valores de LR calculados a partir de los scores del sistema biométrico. Como la representación mediante curvas ECE es una función de $P(H_p)$, representaremos en primer lugar la ecuación del teorema de Bayes como una función de $P(H_p)$.

Como las hipótesis de la acusación y de la defensa son eventos complementarios, tenemos que:

$$P(H_p | I) = 1 - P(H_D | I) \quad (3.6)$$

Podemos definir las probabilidades a priori:

$$O(H_p) = \frac{P(H_p|I)}{P(H_D|I)} = \frac{P(H_p|I)}{1-P(H_p|I)} \quad (3.7)$$

y las probabilidades a posteriori:

$$O(H_p | E) = \frac{P(H_p|E,I)}{P(H_D |E,I)} = \frac{P(H_p |E,I)}{1-P(H_p|E,I)} \quad (3.8)$$

Por lo tanto, podemos sustituir las expresiones anteriores en la ecuación (3.4) obteniendo:

$$O(H_p |E) = LR \cdot O(H_p) \quad (3.9)$$

La ECE viene expresada como [9]:

$$ECE = \frac{P(H_p)}{N_p} \sum_{i \in \text{targets}} \log_2 \left(1 + \frac{1}{LR_i \cdot \frac{P(H_p)}{P(H_D)}} \right) + \frac{P(H_D)}{N_D} \sum_{j \in \text{non-targets}} \log_2 \left(1 + \frac{1}{LR_j \cdot \frac{P(H_p)}{P(H_D)}} \right) \quad (3.10)$$

Donde N_p y N_D serán el número de scores *target* y *non-target* respectivamente.

El valor de ECE considera no sólo el poder de discriminación del conjunto de LR evaluados, sino también su calibración.

El poder de discriminación de un conjunto, ya descrito en el capítulo 2, se refiere a la separación entre los *scores target* y los *non-target*. Cuanto mayor es el valor ECE, más información se necesita para saber cuál de las dos hipótesis de la verdadera por lo que peor es el poder de discriminación de los LR calculados.

La calibración se refiere a la interpretación estadística de un conjunto de LR. En general, la calibración mide a posteriori si una predicción es acertada o no después de que el hecho predicho ocurra [31]. Por lo tanto, cuando tras el cálculo de LR se conoce cuál es la hipótesis verdadera, puede comprobarse cómo de aptos son los LR para ser interpretados probabilísticamente [9]. La calibración tiene el objetivo de presentar la información contenida en los *scores* dentro del contexto bayesiano y así ayudar a tomar la decisión correcta [31] [32].

En la figura 3.4.1 aparece un ejemplo de una curva ECE para un sistema de muestra:

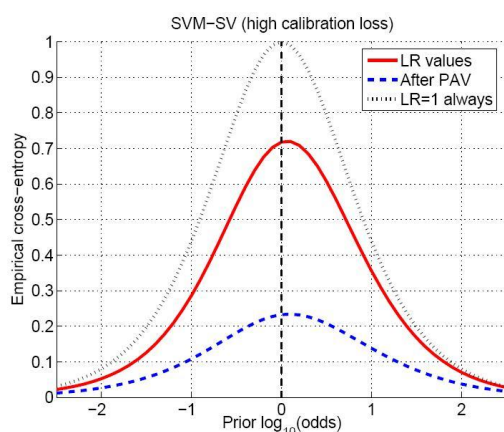


Figura 3.4.1 Ejemplo de curva ECE

En la figura 3.4.1 podemos distinguir tres curvas:

- La curva sólida es la entropía cruzada, es decir, la pérdida media de información de los valores de LR calculados. Cuanto más alta es esta curva, más información se necesita para saber cuál de las dos hipótesis enfrentadas es la verdadera, y por lo tanto, peor es el poder de discriminación del sistema. Además de esta curva, se representan otros dos sistemas para la comparación
- La curva de puntos azules representa el sistema calibrado, que optimiza la ECE conservando la discriminación y obtenida por el algoritmo PAV (*Pool Adjacent Violators*) [31]. Cuanto más cercanas estén la curva azul y la roja, mejor calibrado estará nuestro sistema. Además, la curva azul es una medida del poder de discriminación del sistema, y por tanto curvas DET iguales dan lugar a curvas azules iguales.
- La curva negra puntuada representa el rendimiento de un sistema con LR=1 siempre, conocido como sistema neutral. Esta curva nos sirve como referencia para la curva sólida, que siempre debe estar por debajo para que el sistema tenga algún poder de discriminación. Si la curva sólida estuviera por encima, el sistema estaría perdiendo más información con el cálculo de LR que si la decisión se basara solamente en la información inicial del caso.

Hay un software para dibujar curvas ECE en <http://arantxa.ii.uam.es/~dramos>.

3.5 Métodos de cálculo de LR.

En esta sección se describen los distintos métodos utilizados en este proyecto para transformar los scores que devuelve un sistema biométrico en valores de LR.

3.5.1 Regresión logística

La regresión logística es un método muy utilizado en calibración y otros ámbitos como fusión de sistemas biométricos [33]. El objetivo de la regresión logística es el de obtener una transformación lineal (desplazamiento y escalado) de un conjunto de entrada para optimizar una función objetivo [9].

La transformación realizada por el modelo de regresión logística puede definirse como:

$$f_{lr} = \log(O(H_P|E)) = a_0 + a_1 \cdot E_1 + a_2 \cdot E_2 + \dots + a_k \cdot E_k \quad (3.11)$$

Despejando de la ecuación del teorema de Bayes (3.9) obtenemos que:

$$\begin{aligned} \log(LR) &= a_0 + a_1 \cdot E_1 + a_2 \cdot E_2 + \dots + a_k \cdot E_k - \log(O(H_P)) \\ &= a_0' + a_1 \cdot E_1 + a_2 \cdot E_2 + \dots + a_k \cdot E_k \end{aligned} \quad (3.12)$$

Si deshacemos la transformación logarítmica, llegamos al modelo de regresión logística denotado por la siguiente expresión:

$$P(O(H_P|E)) = \frac{1}{1 + e^{-\log(LR) - \log(O(H_P))}} \quad (3.13)$$

Los valores de los pesos $\{a_0, a_1, a_2 \dots a_k\}$ pueden obtenerse a partir de un conjunto de scores de entrenamiento, haciendo que $P(O(H_P|E))$ sea lo más cercano a 1 para los scores *target* y lo más cercano posible a 0 para scores *non-target* [9].

3.5.2 PAV (Pool Adjacent Violators)

Otra aproximación de cálculo de LRs fue propuesta en [31] denominada algoritmo PAV (*Pool Adjacent Violators*). Este algoritmo transforma un conjunto de scores en un conjunto de LR calibrados. Con este algoritmo sólo es posible calcular una transformación óptima cuando la verdadera hipótesis para cada score es conocida. Sin embargo, puede aplicarse una transformación óptima si se entrena con un conjunto de scores cuya hipótesis verdadera es conocida y después se aplica la transformación entrenada a conjunto de test de hipótesis verdadera desconocida. Esta transformación es escalonada y puede sobreentrenarse a los datos de entrenamiento, por lo que no es fiable con un conjunto reducido de datos.

3.5.3 KDF (Kernel Density Function)

Proporciona una distribución más flexible, poco adecuada cuando se dispone de pocos datos ya que sobreentrena el conjunto de entrenamiento. La transformación que aplica puede observarse en la figura 3.5.1. A la izquierda aparecen las fdp de los scores *target* y *non-target*, y a la derecha la transformación que se aplica para el cálculo de LR. Para más información acerca de este método se refiere al lector a [34] [35] [36].

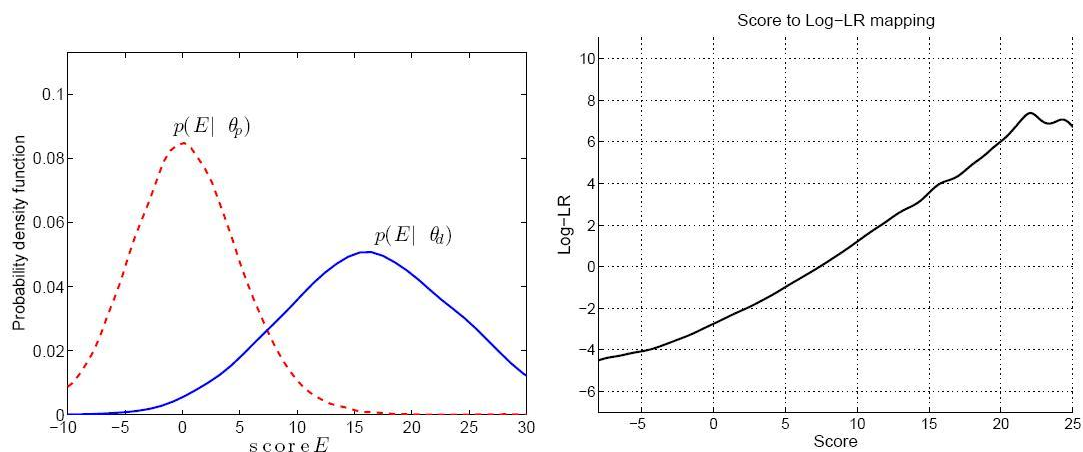


Figura 3.5.1 Método KDF de generación de LR [9].

3.5.4 Gaussian

Transformación en LR que supone una distribución gaussiana del conjunto de scores. Se basa en la teoría de que toda distribución tiende a ser gaussiana cuando es muy grande, por lo tanto no funciona bien cuando se dispone de pocos datos. Más información puede ser encontrada en [37] [38].

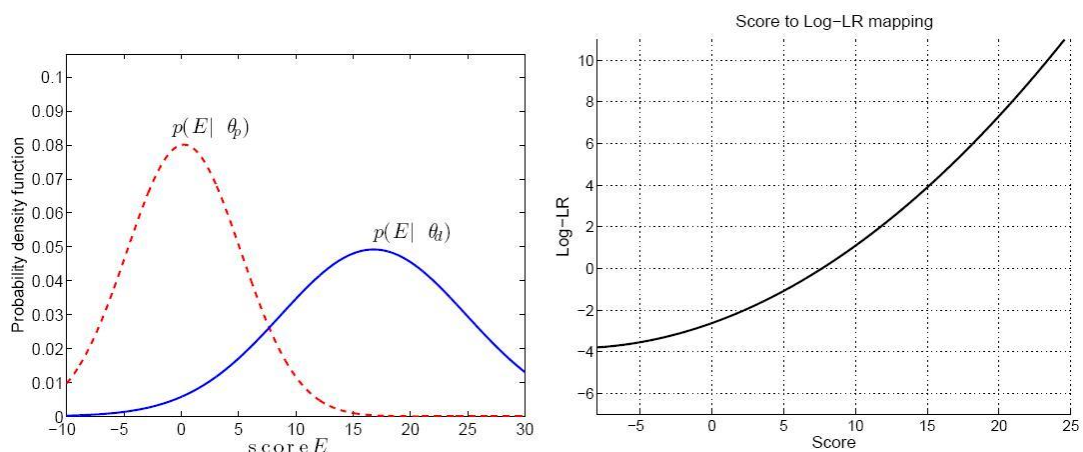


Figura 3.5.2 Método gaussian de generación de LR [9].

4

Marco Experimental

4.1 Sistema utilizado

Para la realización de los experimentos desarrollados en este proyecto, se ha utilizado un sistema biométrico de reconocimiento de huella dactilar en el estado del arte. El sistema elegido trabaja en modo identificación.

El funcionamiento del comparador se basa en características de primer y segundo nivel. Siendo capaz de comparar huellas utilizando la localización de núcleo, deltas y minucias. Además puede también ayudarse con comparaciones basadas en la textura de las imágenes. El resultado de cada comparación es un score de similitud entre ambas huellas.

La entrada al sistema es la imagen de una huella dactilar que quiere compararse con las de una base de datos a la que el sistema tiene acceso. Las minucias de la huella de entrada, al igual que el núcleo y los deltas, pueden marcarse de forma automática por el sistema, o de forma manual por el usuario. Para cada entrada que recibe el sistema para ser comparada, el sistema permite la selección de varias opciones de búsqueda, como el margen de rotación, la tolerancia a la distorsión, el área de la huella, etc.

La salida del sistema es una lista de 15 candidatos ordenados de mayor a menor score. Además de esto, el sistema computa la calidad de la huella de entrada.

El funcionamiento del sistema es el siguiente: El usuario introduce la imagen digital de una huella previamente archivada. Una vez hecho esto, el sistema computa de forma automática la extracción de minucias. La configuración automática de minucias puede quitarse si se quiere y marcarlas de forma manual. A continuación se marcan, también opcionalmente, el núcleo y los deltas de la huella y la zona de área que quiere compararse. Finalizada la edición, se seleccionan los parámetros de las comparaciones (tolerancia a la distorsión, margen de rotación, utilización de textura...) y la huella ya está lista para el comparador. Una vez realizada la búsqueda contra la base de datos, el sistema devuelve una lista de 15 candidatos junto con los scores correspondientes.

4.2 Bases de datos de huellas utilizadas

Para la realización de los experimentos, se dispone de 3 conjuntos de datos.

- 1) Una base de tarjetas decadactilares provenientes de reseñas reales. La base de datos consta de aproximadamente 2,5 millones de tarjetas decadactilares, con 10 dedos por tarjeta tomados de dos formas distintas: huellas posadas y huellas rodadas. Cada individuo reseñado en la base de datos puede tener una o más tarjetas decadactilares en la base de datos. Las reseñas dactilares de este conjunto pueden haber sido adquiridas en papel y tinta o por medio de un scanner de alta resolución. En esta base de datos, existen imágenes de 500 dpi y de 1000 dpi.
- 2) Una colección de 100 impresiones dactilares, que consiste en 50 posadas y 50 rodadas de los dedos índices de la mano derecha de 50 individuos diferentes. Es decir, un conjunto de test de 50 usuarios y dos impresiones por cada usuario (rodada y posada). Además, para cada uno de los individuos de esta colección existen una o más reseñas decadactilares en la base de tarjetas cuyas impresiones son otras diferentes de las de esta colección. Las reseñas dactilares de este conjunto han sido adquiridas en papel y tinta, y posteriormente escaneadas a 1000 dpi.
- 3) Una colección de 50 huellas latentes obtenidas en casos forenses reales y seleccionadas de forma que una o más tarjetas decadactilares de los individuos a los que corresponden cada una de las huellas se encuentren en la base de tarjetas decadactilares utilizada.

Para la realización de los experimentos, el primer conjunto será utilizado como la base de datos de la que dispone el sistema de identificación para realizar las comparaciones. Es decir, para cada entrada a sistema, este realizará la comparación con todas las huellas de la base de datos.

El segundo y tercer conjunto de huellas se utilizarán como entradas al sistema, de forma que serán las huellas comparadas con las de las tarjetas decadactilares de la base de datos. Hay que destacar que para todas las huellas de entrada existirá al menos un candidato genuino en la base de datos.

Es necesario mencionar que la cantidad de datos utilizados para la realización de estos experimentos es escasa para aportar resultados suficientemente fiables estadísticamente. El motivo de esta limitación en los datos radica en la necesidad de marcar los puntos característicos (en latentes) y los parámetros de las búsquedas de forma manual, lo cual consume una enorme cantidad de tiempo y recursos.

Sin embargo, con los datos disponibles, la evaluación del rendimiento del sistema dará una idea general su funcionamiento si se extrapolan los resultados.

4.3 Protocolo experimental

Para definir el protocolo a seguir en los experimentos, se realizó un análisis de los factores que se pueden variar en cada comparación lanzada utilizando el sistema de identificación, identificando los siguientes:

- **Número de puntos característicos.** En el protocolo seguido por los especialistas forenses, los puntos característicos son identificados por el especialista en búsquedas utilizando huellas latentes y automáticamente por el sistema en búsquedas utilizando imágenes procedentes de tarjetas decadactilares (impresiones).
- **Área de la imagen de la huella.** Se puede variar recortando la imagen o seleccionando sólo el área deseada mediante un lazo de corte provisto por la herramienta de edición de la aplicación del sistema.
- **Tolerancia del sistema a la distorsión.** Variable discreta que toma un valor entre un máximo y un mínimo.
- **Margen de rotación.** Define la rotación máxima permitida en las búsquedas.
- **Textura.** Parámetro que al ser seleccionado para su utilización en las búsquedas utiliza información relevante adicional a la geografía de los puntos característicos para realizar las comparaciones.
- **Núcleo y deltas.** Se marcan de forma manual en las comparaciones y pueden utilizarse o no en las comparaciones.
- **Fórmula.** Definición del tipo de dactilograma. En la actualidad no se utiliza en las búsquedas por los especialistas por diversas razones, y por ello no se utilizará en las pruebas definidas en este protocolo.
- **Búsqueda por número de dedo.** Restringe la búsqueda al dedo al que se indique que pertenece la huella que con la que se va a realizar la comparación.

Considerando todo ello, durante las primeras semanas del proyecto, se ha intentado valorar cuáles de estos factores son más importantes para el trabajo diario del especialista. Los criterios de valoración han sido los siguientes:

- Frecuencia de uso en la operativa para el especialista, o dicho de otra forma, qué factor de variabilidad se usa más en comparaciones en un caso.
- Grado de sorpresa del especialista sobre el impacto de los experimentos preliminares realizados.
- Grado de desconocimiento por parte del especialista acerca de los efectos del factor analizado en el rendimiento del sistema.
- Interés científico de acuerdo con el estado del arte de la tecnología de reconocimiento biométrico dactilar.

De acuerdo a los criterios expuestos, se han seleccionado para este protocolo los siguientes factores de variación:

1. **Número de puntos característicos.** De acuerdo con la operativa del especialista, en latentes estos puntos se seleccionan a mano por el especialista. En estas pruebas incluimos además un experimento con búsquedas lanzadas con extracción automática de puntos. En el caso de búsquedas con impresiones dactilares, los puntos marcados serán los seleccionados de forma automática por el sistema.
2. **Área de la imagen de la huella.** Se utilizará el lazo de corte del editor de huellas del sistema para variar el área de la huella y observar la variabilidad de los resultados.
3. **Tolerancia del sistema a la distorsión.** La distorsión se identifica como un factor de capital importancia en identificación dactilar por lo que se realizarán pruebas variando la tolerancia a la distorsión del sistema.
4. **Calidad de la huella.** Se almacenará la calidad mostrada por el sistema para cada huella lanzada con el objetivo de hacer análisis del funcionamiento del sistema en función de la calidad de la huella

4.4 Pruebas realizadas

4.4.1 Pruebas realizadas con huellas latentes

De acuerdo a los factores de variabilidad seleccionados anteriormente, el protocolo seguido se compone de 8 pruebas por cada huella para comparaciones entre huellas anónimas previamente identificadas por expertos en dactiloscopia y la base de reseñas decadactilares. Estas pruebas se diferencian en la modificación de una o varias de los factores de variabilidad seleccionados, y se resumen a continuación (resaltamos en cursiva el factor de variabilidad modificado en cada prueba con respecto al resto):

- **Pruebas de variación de puntos característicos**
 - Prueba “PCAUTO”: *selección automática de puntos característicos*, área total de la huella, tolerancia a la distorsión media.
 - Prueba “PCM”: *máximo de puntos característicos seleccionados manualmente*, área total de la huella, tolerancia a la distorsión media (1/2 del valor máximo).
 - Prueba “PC12”: *12 puntos característicos de los seleccionados manualmente*, área total de la huella, tolerancia a la distorsión media (1/2 del valor máximo).
 - Prueba “PC8”: *8 puntos característicos de los seleccionados manualmente*, área total de la huella, tolerancia a la distorsión media (1/2 del valor máximo).

- **Pruebas de variación de área de la huella**
 - Prueba “AR2”: *12 puntos característicos, mitad aproximada del área total de la huella*, tolerancia a la distorsión media (1/2 del valor máximo).
 - Prueba “AR3”: *12 puntos característicos, un tercio aproximado del área total de la huella*, tolerancia a la distorsión media (1/2 del valor máximo).

- **Pruebas de variación de la tolerancia a la distorsión**
 - Prueba “DL”: *12 puntos característicos, área total de la huella, tolerancia a la distorsión baja (1/4 del valor máximo)*.
 - Prueba “DH”: *12 puntos característicos, área total de la huella, tolerancia a la distorsión máxima (3/4 del valor máximo)*.

El resto de factores de variabilidad identificados se dejaron con valores por defecto, definidos a continuación:

- Margen de rotación: $\pm 90^\circ$
- Textura: marcar la casilla “Utilizar textura”.

- Núcleo y deltas: Como estos han de ser marcados de forma manual, no han sido utilizados en la prueba PCAUTO. Para el resto de pruebas se marcaron los núcleos y los deltas siempre que estaban disponibles, y sin contabilizarse para el número de puntos característicos marcados en la huella.
- Fórmula: de acuerdo al protocolo operativo del especialista, no se utilizó la fórmula en las búsquedas.
- Búsqueda por número de dedo: en muchas huellas latentes no es posible determinar el dedo del que proviene la misma. Por lo tanto, no se utilizó el número de dedo en pruebas de huellas latentes contra decadaclares.

Cada una de las 50 huellas anónimas seleccionadas fue buscada 8 veces siguiendo el protocolo de 8 experimentos descrito anteriormente, obteniendo para cada experimento una lista de 15 candidatos con los scores correspondientes. Además de los scores obtenidos en las comparaciones, se almacenaron también la calidad mostrada por el sistema de cada huella lanzada y el número de puntos característicos con los que se lanzó la huella en las pruebas PCAUTO y PCM. Una vez realizados los 8 experimentos con huellas latentes, tendremos 400 listas de 15 scores cada una para analizar.

4.4.2 Pruebas realizadas con impresiones dactilares

De acuerdo a los factores de variabilidad seleccionados anteriormente, el protocolo seguido se compone de 7 pruebas por cada huella para comparaciones entre huellas anónimas identificadas y la base de reseñas decadaclares. Estas pruebas se diferencian en la modificación de uno o varios de los factores de variabilidad seleccionados, y se resumen a continuación (resaltamos en cursiva el factor de variabilidad modificado en cada prueba con respecto al resto):

- **Pruebas de variación de puntos característicos**
 - Prueba “PCAUTO”: *selección automática de puntos característicos*, área total de la huella, tolerancia a la distorsión media.
 - Prueba “PC12”: *12 puntos característicos de los seleccionados automáticamente*, área total de la huella, tolerancia a la distorsión media (1/2 del valor máximo).
 - Prueba “PC8”: *8 puntos característicos de los seleccionados automáticamente*, área total de la huella, tolerancia a la distorsión media (1/2 del valor máximo).
- **Pruebas de variación de área de la huella**
 - Prueba “AR2”: *12 puntos característicos, mitad aproximada del área total de la huella*, tolerancia a la distorsión media (1/2 del valor máximo).
 - Prueba “AR3”: *12 puntos característicos, un tercio aproximado del área total de la huella*, tolerancia a la distorsión media (1/2 del valor máximo).

- **Pruebas de variación de la tolerancia a la distorsión**
 - Prueba “DL”: 12 puntos característicos, área total de la huella, *tolerancia a la distorsión baja (1/4 del valor máximo)*.
 - Prueba “DH”: 12 puntos característicos, área total de la huella, *tolerancia a la distorsión máxima (3/4 del valor máximo)*.

El resto de factores de variabilidad identificados se dejaron con valores por defecto, definidos a continuación:

- Margen de rotación: $\pm 90^\circ$
- Textura: marcar la casilla “Utilizar textura”.
- Núcleo y deltas: Como estos han de ser marcados de forma manual, no han sido utilizados en la prueba PCAUTO. Para el resto de pruebas se marcaron los núcleos y los deltas siempre que estaban disponibles, y sin contabilizarse para el número de puntos característicos marcados en la huella.
- Fórmula: como se ha dicho anteriormente, y de acuerdo al protocolo operativo del especialista, no se utilizó la fórmula en las búsquedas.
- Búsqueda por número de dedo: no se utilizó el número de dedo en pruebas de impresiones contra reseñas decadactilares.

Cada una de las 50 huellas seleccionadas fue buscada 7 veces siguiendo el protocolo de 7 experimentos descrito anteriormente.

A diferencia de los experimentos con huellas anónimas, en los experimentos con impresiones dactilares la selección de puntos es siempre automática, utilizando para PCAUTO la selección inicial de puntos del sistema, y para el resto de pruebas seleccionando el número de puntos necesario (12 u 8 puntos) de los seleccionados automáticamente por el sistema.

El resultado de cada comparación es una lista de 15 candidatos ordenados de mayor a menor score. Además de los scores obtenidos en las comparaciones, se almacenaron también la calidad mostrada por el sistema de cada huella lanzada y el número de puntos característicos con los que se lanzó la huella en la prueba PCAUTO. Una vez finalizados los experimentos se dispone de 700 listas de scores (350 para rodadas y 350 para posadas) de 15 scores cada una para analizar.

5

Análisis de los factores de variabilidad

En este capítulo se hace un análisis comparativo y cualitativo del funcionamiento del sistema para todas las pruebas realizadas, tanto para huellas latentes como para las impresiones rodadas y posadas.

5.1 Pruebas de variación del número de puntos característicos

En esta sección se analizan los resultados obtenidos de los experimentos de variación del número de puntos característicos de las huellas lanzadas en forma de curvas CMC. Se han realizado distintas comparaciones con cada huella, variando el número de puntos característicos en cada caso. El objetivo es ver la influencia del número de puntos característicos de la huella lanzada en la posición del usuario genuino en la lista de candidatos devueltos por el sistema. Tenemos así cuatro tipos de búsqueda para cada huella:

- PCAUTO: Puntos característicos seleccionados por el sistema.
- PCM: Puntos característicos seleccionados manualmente (sólo en latentes).
- PC12: 12 puntos característicos (automáticos para impresiones y manuales para latentes).
- PC8: 8 puntos característicos (automáticos para impresiones y manuales para latentes).

5.1.1 Experimentos con huellas latentes

En la figura 5.1.1 se representan las curvas CMC obtenidas con los scores de las pruebas de variación de puntos característicos para las comparaciones con huellas latentes.

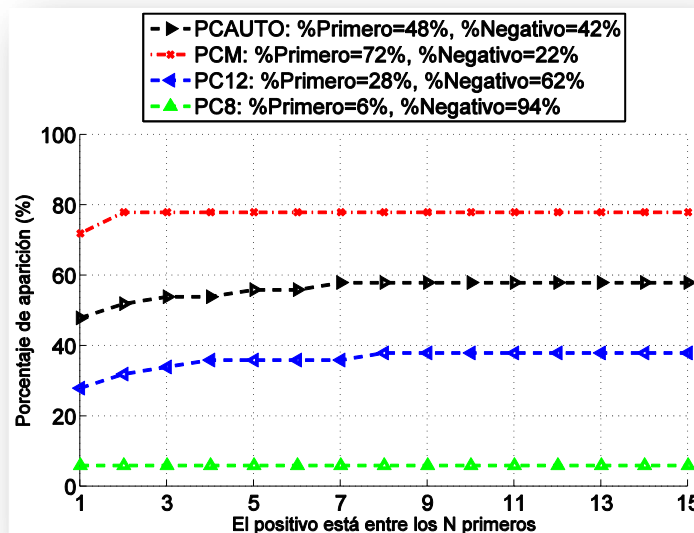


Figura 5.1.1 CMC para las pruebas de variación de puntos característicos en latentes

Como puede apreciarse en la figura, en ninguno de los 4 experimentos se consigue una curva que llegue al 100%. Es decir, no hay ningún experimento en el que todas las huellas genuinas aparezcan en su lista como uno de los 15 candidatos.

El mejor de los casos ocurre cuando se marcan manualmente los puntos característicos, siendo el número de éstos el máximo que se encuentre en la huella. En un 72% de los casos la huella genuina aparece en el primer candidato. En un 22% de los casos, la huella genuina no aparece en la lista de candidatos.

El siguiente mejor resultado ocurre en comparaciones realizadas con la extracción automática de puntos característicos. En este caso, un 48% de los experimentos devuelven la huella genuina como primer candidato. En un 42% de los casos la huella genuina no aparece en la lista de candidatos.

En los experimentos realizados con 12 puntos característicos de los seleccionados manualmente, la huella genuina aparece como primer candidato en un 28% de casos, siendo el porcentaje de no aparición en toda la lista del 62%. Resulta muy llamativo que la reducción en el rendimiento sea mucho menor si el sistema selecciona automáticamente los puntos característicos que si se toman 12 puntos manualmente, que es un número considerable.

Finalmente, cuando el número de puntos característicos utilizados en la comparación es 8, encontramos que sólo un 6% de las huellas genuinas aparecen en la lista de candidatos, encontrándose todas ellas en la primera posición. El restante 94% de las comparaciones no dieron la huella genuina como uno de los 15 candidatos.

Además del resultado anterior, podemos desglosar los experimentos PCAUTO (extracción automática de puntos) y PCM (máximo número de puntos marcados manualmente) en función del número de puntos en cada huella. En las siguientes figuras se observa la frecuencia de aparición en cada posición en la lista para estas pruebas.

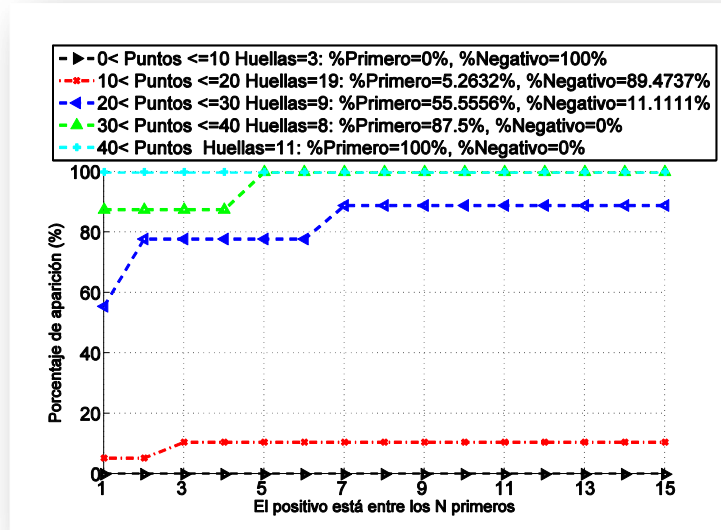


Figura 5.1.2 CMC desglosada para la prueba PCAUTO en latentes

En el caso de PCAUTO (figura 5.1.2), podemos comprobar como el número de puntos característicos es crucial a la hora de que el usuario genuino aparezca en una posición más alta de la lista de candidatos. Es apreciable que para huellas con menos de 20 puntos característicos seleccionados por el sistema, los resultados no son satisfactorios. Sin embargo, para huellas con más de 40 puntos, existe un 0% de negativos, es decir, el usuario genuino aparece siempre como candidato.

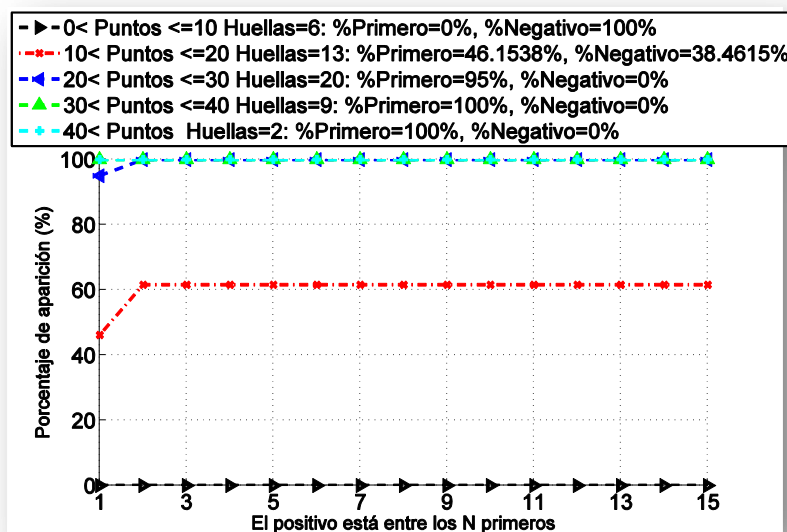


Figura 5.1.3 CMC desglosada para la prueba PCM en latentes

En el caso de PCM (figura 5.1.3), considerada la prueba más importante de todos los experimentos, ya que es la casuística principal del experto (marcar manualmente el máximo número de puntos característicos en huellas latentes), los resultados son notablemente mejores que para PCAUTO. Para huellas latentes con más de 30 puntos característicos, el 100% de los experimentos realizados devuelve al candidato genuino dentro de la lista de candidatos.

Cuando el número de puntos característicos disminuye, baja notablemente el rendimiento del sistema. Se aprecia que para huellas de 10 a 20 puntos característicos, se dio únicamente un 48% de positivos en primera posición y un 38,46% de las huellas que nunca aparecieron en la lista. Para huellas con menos de 10 puntos característicos, el usuario genuino nunca apareció como candidato. Aunque el tamaño de la muestra utilizada para estos experimentos es pequeño, podemos observar que el sistema no funciona de la manera deseada para huellas latentes con menos de 20 puntos característicos.

5.1.2 Experimentos con impresiones posadas

En este apartado se muestran los resultados de la influencia del número de puntos característicos en las pruebas realizadas con impresiones dactilares posadas. En este caso se elimina la prueba PCM, ya que la casuística de los expertos en este tipo de huellas es que el sistema marque los puntos característicos de forma automática. En la figura 5.1.4 se muestran los resultados obtenidos:

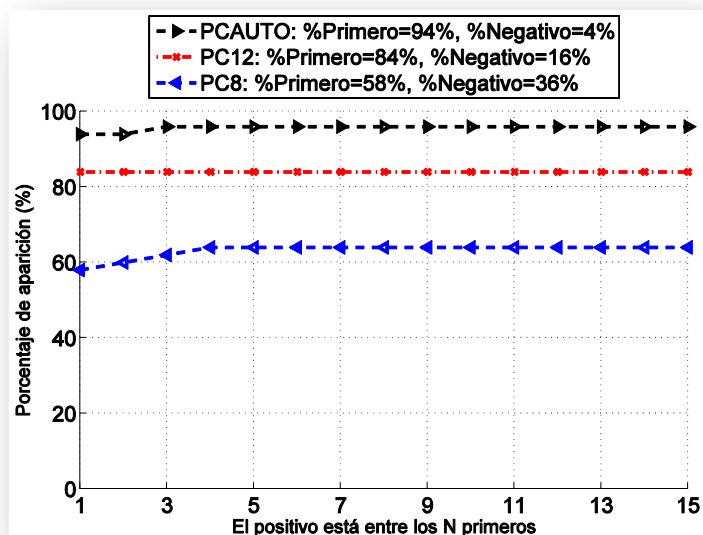


Figura 5.1.4 CMC para las pruebas de variación de puntos característicos en impresiones posadas

En la figura 5.1.4 puede observarse la estrecha relación entre el número de puntos característicos y el porcentaje de positivos que ofrece el sistema. Debido a la buena calidad de las huellas de estos experimentos, en comparación con las huellas anónimas, la selección automática de puntos característicos es mucho más exacta y por lo tanto el porcentaje de positivos es mayor. Así, puede observarse que sólo en un 4% de los casos, la huella genuina no aparece en la lista de candidatos para PCAUTO., frente a un 42% de negativos en huellas

latentes para la misma prueba y un 22% para PCM. Además, un 94% de los experimentos de PCAUTO para impresiones dactilares muestra el candidato genuino en primera posición, mientras que en los experimentos de huellas latentes esta cifra supone sólo el 48% en PCAUTO y el 72% en PCM.

Para PC12, prueba en la que se buscan las huellas con 12 puntos característicos de los seleccionados previamente por el sistema de manera automática, el número de positivos es el mismo en toda la curva. Esto quiere decir que todos los candidatos genuinos que aparecieron (84% de los casos), lo hicieron siempre en primera posición, siendo negativo el resultado del restante 16% de las búsquedas.

Para las pruebas con 8 puntos característicos (también elegidos de entre los extraídos automáticamente), el porcentaje de positivos en primera posición es de 58% mientras que un 36% de usuarios genuinos no aparece nunca en su lista de candidatos.

Tanto los resultados de las pruebas PC12 como los de PC8 para impresiones dactilares posadas son notablemente mejores que los de los mismos experimentos realizados con huellas latentes, como era de esperar.

En la figura 5.1.5 se desglosa la curva CMC para la prueba PCAUTO, en función de intervalos de número de puntos característicos.

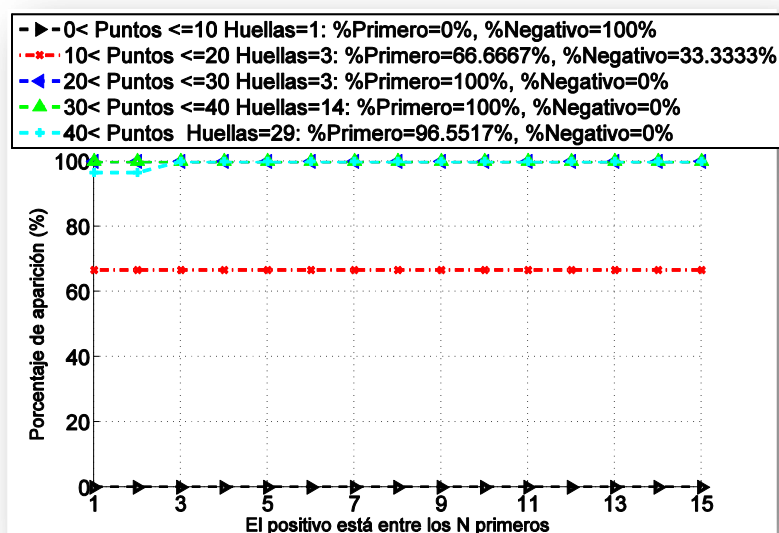


Figura 5.1.5 CMC desglosada para la prueba PCAUTO en posadas

Como se aprecia en la figura 5.1.5, el número de puntos característicos que el sistema extrae de las huellas es en cierto modo proporcional al porcentaje de positivos en primera posición de la lista. Para los experimentos con huellas con menos de 20 puntos, el porcentaje de negativos es del 50%. Para todas las huellas con más de 20 puntos, todas las huellas genuinas aparecen en la lista de candidatos, es decir, el porcentaje de negativos es 0%.

5.1.3 Experimentos con impresiones rodadas

En las impresiones rodadas, el área de la huella es mayor, por lo que el sistema dispone de más cantidad de información para utilizar en las comparaciones entre plantillas.

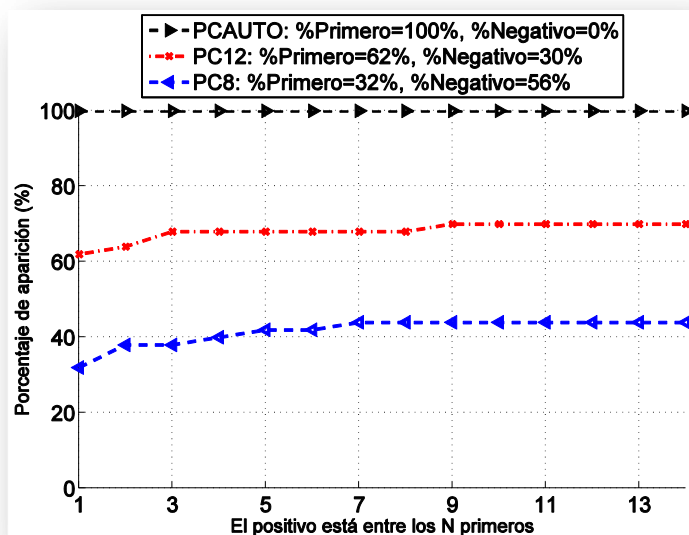


Figura 5.1.6 CMC para las pruebas de variación de puntos característicos en rodadas

Puede comprobarse observando la figura 5.1.6 que en impresiones rodadas, el sistema encuentra la huella genuina en primera posición en un 100% de los casos para la prueba en la que los puntos característicos son seleccionados por el sistema. Sin embargo, tal y como se aprecia en la figura, cuando el número de puntos característicos se reduce a 12 y 8 puntos, el porcentaje de positivos en primera posición disminuye considerablemente a un 62% y a un 32% respectivamente. A su vez, en estas dos últimas pruebas, el porcentaje de veces en las que la huella objetivo no aparece en la lista de candidatos aumenta, siendo del 30% para PC12 y del 56% para PC8.

Puede comprobarse como los resultados para PCAUTO en rodadas son mejores que en posadas, y sin embargo, para el resto de pruebas, el porcentaje de positivos es mayor en las pruebas de huellas posadas. Esto se debe a que las impresiones rodadas sufren mayor distorsión que las posadas, dando lugar a peores resultados cuando el número de puntos es igual.

En la tabla 5.1.1 se realiza una comparativa del porcentaje de aparición de la huella genuina en la lista de candidatos para todos los experimentos de variación del número de puntos característicos.

Porcentaje de aparición de la huella genuina como primer candidato				
Tipo de huella	PCM	PCAUTO	PC12	PC8
Latentes	72%	48%	28%	6%
Posadas		94%	84%	58%
Rodadas		100%	62%	32%

Tabla 5.1.1 Resultados para experimentos de variación de puntos característicos

La tabla muestra el porcentaje de positivos en primera posición para todas las pruebas de variación del número de puntos característicos de la huella lanzada. En ella se puede comparar de un vistazo el rendimiento del sistema para cada una de las pruebas.

De ella podemos deducir:

- Que el funcionamiento de la selección automática de puntos característicos es muy bueno en impresiones dactilares, sobre todo en rodadas. Sin embargo, en el caso de las huellas latentes sólo el 48% de huellas aparecieron en primera posición cuando se utilizó la selección automática, frente a un 72% en la selección manual.
- Que la reducción del número de puntos característicos a 12 y a 8, empeora notablemente el rendimiento del sistema en todas las pruebas. Además es curiosa la observación de que para estas dos pruebas, el rendimiento en impresiones posadas es mejor que en impresiones rodadas, debido la existencia de mayor distorsión en estas últimas.

5.2 Pruebas de variación de la tolerancia a la distorsión

La distorsión es un parámetro de gran importancia en el ámbito de la identificación dactilar. La presión del dedo, el movimiento, la curvatura de la superficie y su textura; estos y más factores producen que distintas marcas procedentes del mismo dedo no sean iguales.

Además, la distorsión se hace mucho más notable en huellas anónimas, ya que las superficies pueden ser de muchas naturalezas y difieren de la superficie sobre la que se reseñan las impresiones decadactilares policiales.

La tolerancia a la distorsión en el sistema aparece como una variable discreta con 5 valores posibles entre un mínimo y un máximo, que puede ser seleccionada por el usuario al lanzar una búsqueda. Se han realizado 3 experimentos para mostrar la variabilidad en los resultados producida por la variación de la tolerancia a la distorsión seleccionada en las comparaciones. Estos 3 experimentos consisten en 3 búsquedas de cada una de las 50 huellas utilizando 12 puntos iguales para las tres búsquedas, que son:

- PC12: tolerancia a la distorsión media (por defecto en el sistema).
- DH: tolerancia a la distorsión alta.
- DL: tolerancia a la distorsión baja.

5.2.1 Experimentos con huellas latentes

En la figura 5.2.1 se presentan las curvas CMC para los tres experimentos de variación de la tolerancia a la distorsión del sistema para huellas latentes.

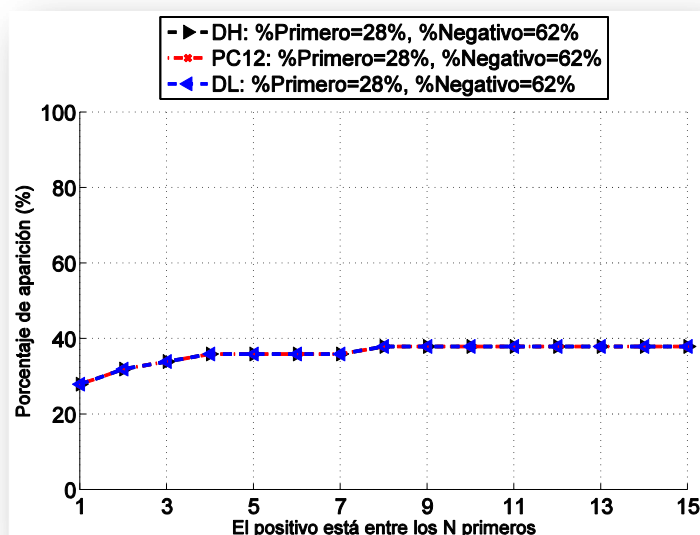


Figura 5.2.1 CMC para pruebas de variación de distorsión en latentes

Como puede comprobarse observando la figura, las tres curvas correspondientes a las pruebas PC12, DH y DL son iguales. La selección de distintas tolerancias a la distorsión no ofrece variabilidad en los resultados obtenidos. Las listas de candidatos obtenidas para los tres experimentos son las mismas.

5.2.2 Experimentos con impresiones posadas

En la figura 5.2.2 se presentan las curvas CMC para los tres experimentos de variación de la tolerancia a la distorsión del sistema para impresiones posadas.

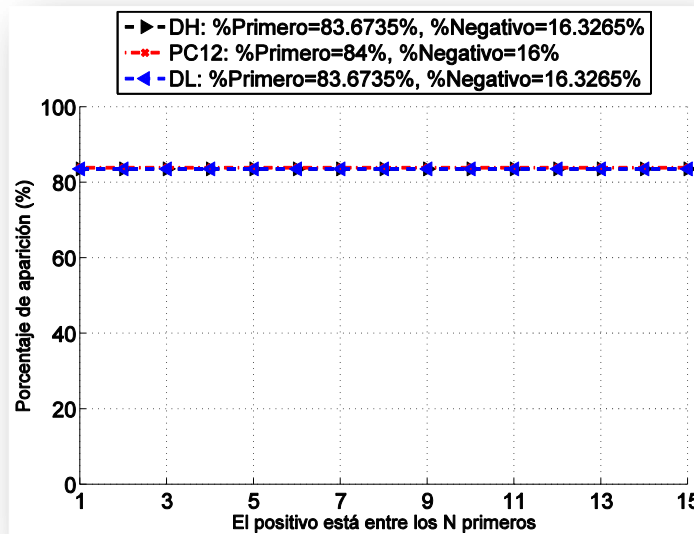


Figura 5.2.2 CMC para pruebas de variación de la distorsión en impresiones posadas

Para los experimentos con impresiones posadas, la variación de la tolerancia a la distorsión en las comparaciones, tampoco produce ningún cambio en las listas de candidatos devueltos por el sistema. Tal y como se aprecia en la figura, las curvas CMC para las 3 pruebas son iguales.

5.2.3 Experimentos con impresiones rodadas

En la figura 5.2.3 se presentan las curvas CMC para los tres experimentos de variación de la tolerancia a la distorsión del sistema para impresiones rodadas.

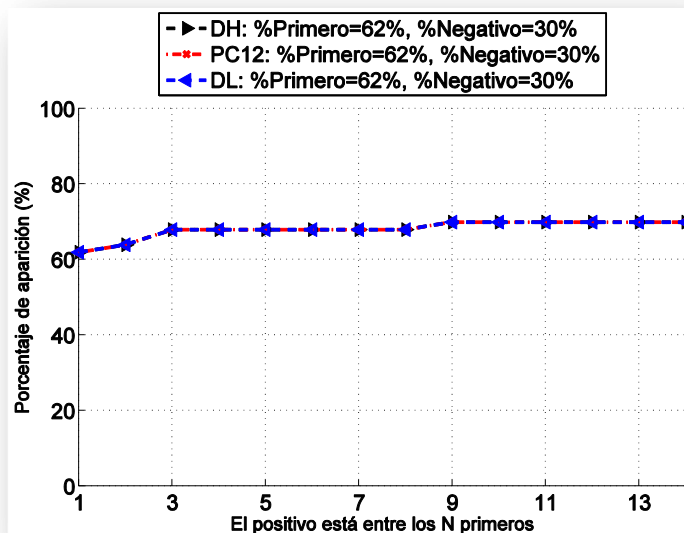


Figura 5.2.3 CMC para pruebas de variación de la distorsión con impresiones rodadas

Como puede comprobarse, en este caso tampoco existe ninguna variación entre las 3 curvas, siendo exactamente iguales para los 3 experimentos.

Podemos concluir de la realización de estos experimentos, que la barra de variación de la tolerancia a la distorsión del sistema de reconocimiento dactilar utilizado, no produce ningún efecto en los resultados de las comparaciones. Por lo tanto, a partir de este punto no se realizarán más análisis en las pruebas de tolerancia a la distorsión para ninguno de los tres tipos de huellas, ya que todos los resultados son iguales a los correspondientes a la prueba PC12.

5.3 Pruebas de variación del área de la huella

La variación en el área consiste en la utilización del lazo de selección de la aplicación del sistema utilizado. Para observar los cambios en los resultados de las búsquedas en función de la variación del área, se realizaron tres búsquedas para cada huella. En todas las búsquedas se utilizan los mismos 12 puntos característicos iguales para las tres pruebas, que son:

- PC12: todo el área de la huella.
- AR2: mitad del área de la huella
- AR3: un tercio del área de la huella.

5.3.1 Experimentos con huellas latentes

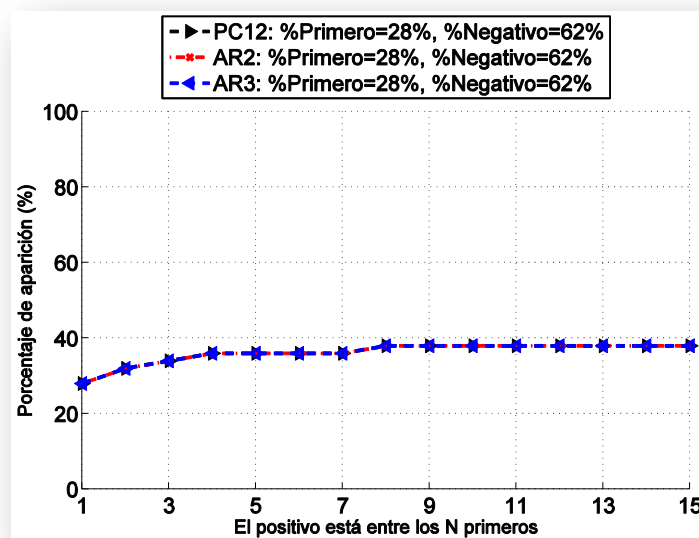


Figura 5.3.1 CMC para pruebas de variación del área con huellas latentes

Como muestra la figura 5.3.1, la variación en la selección del área de la huella buscada no produce ningún efecto en los resultados de las búsquedas, siendo la curva CMC igual para los tres experimentos. Además, puede comprobarse como las curvas son también iguales a las de las pruebas de variación de la tolerancia a la distorsión del sistema, detalladas en el apartado anterior.

En los dos siguientes apartados se muestran las curvas CMC para la variación del área en impresiones dactilares, llevándonos a la misma conclusión (figuras 5.3.2 y 5.3.3).

Por lo tanto, a partir de este punto no se analizarán más los resultados obtenidos de la variación del área de la huella utilizando el lazo de corte del sistema, ya que todos ellos se corresponden con los de la prueba PC12.

5.3.2 Experimentos con impresiones posadas

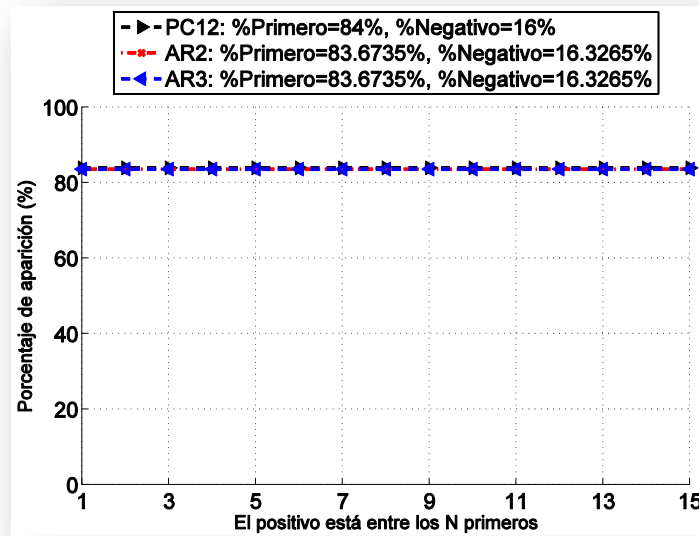


Figura 5.3.2 CMC para pruebas de variación del área en impresiones posadas

5.3.3 Experimentos con impresiones rodadas

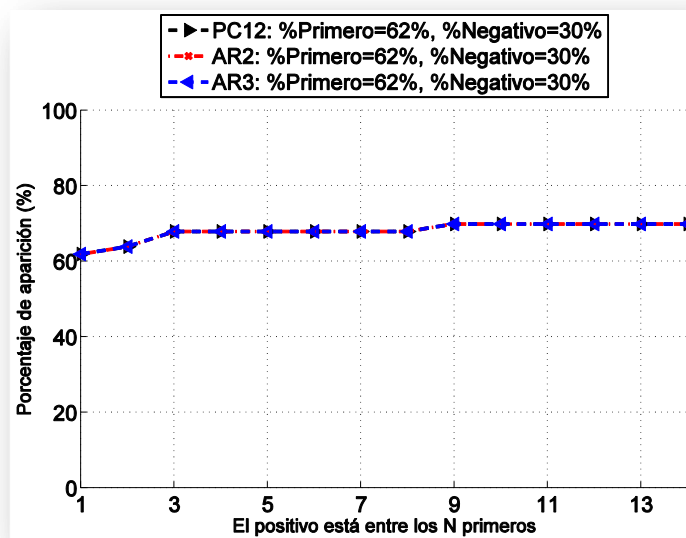


Figura 5.3.3 CMC para pruebas de variación del área en impresiones rodadas

5.4 Pruebas de variación de la calidad de las huellas

La calidad es un parámetro de gran importancia en los sistemas biométricos hoy en día [39]. En el sistema estudiado, se muestra la calidad de la huella lanzada con un parámetro llamado “Q”. Este parámetro es una variable discreta que toma valores enteros. Aunque los extremos entre los que varía se desconocen, al igual que los criterios de asignación de calidad a las huellas, se sabe que cuanto mayor es “Q” peor es la calidad de la huella.

Para estudiar el efecto de la calidad en los distintos experimentos, hemos almacenado el valor de este parámetro en cada experimento para poder analizar su relación con el porcentaje de positivos y con la posición en la lista de los mismos.

5.4.1 Experimentos con huellas latentes

En la figura 5.4.1 aparecen las curvas CMC para la prueba PCAUTO, correspondientes a distintos intervalos de calidad de la huella lanzada. En esta figura puede observarse una clara relación entre la calidad de la huella según el parámetro “Q” del sistema, y el porcentaje de aparición de positivos.

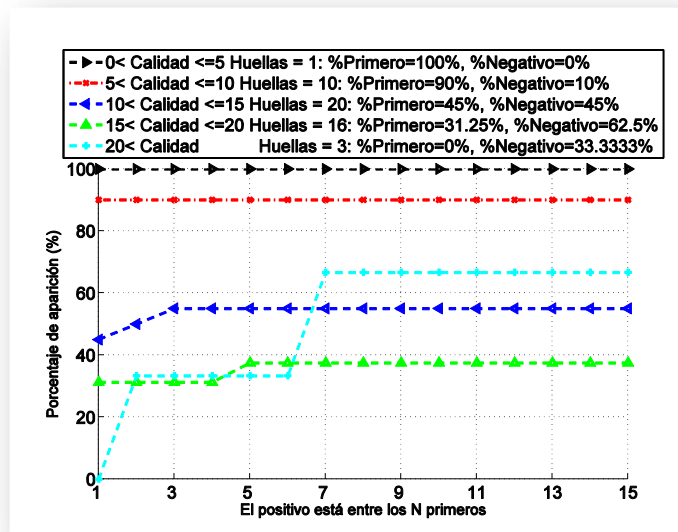


Figura 5.4.1 CMC de variación de calidad para prueba PCAUTO en latentes

Puede comprobarse que para huellas cuya calidad es mayor ($Q < 5$), existe un 100% de positivos. Esta relación va disminuyendo conforme lo hace la calidad, siendo del 90% los positivos para una “Q” entre 5 y 10, y del 45% para “Q” entre 10 y 20.

En la figura 5.4.2 se presenta la variación de resultados según la calidad de la huella para la prueba PC12.

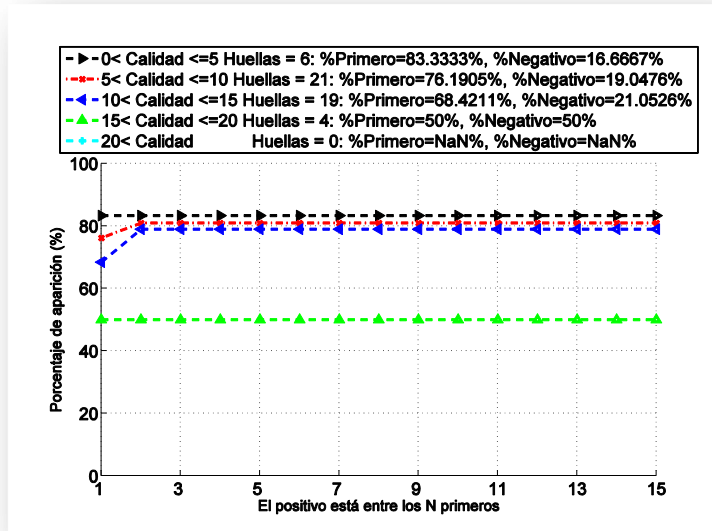


Figura 5.4.2 CMC de variación de calidad para prueba PCM en latentes

En ella se observa la estrecha relación entre la calidad de la huella lanzada y el porcentaje de positivos. Vemos que para las calidades mayores, existe un 83% de positivos, mientras que para las menores, el porcentaje de positivos es sólo del 50%.

En las figuras 5.4.3 y 5.4.4 aparecen respectivamente las curvas CMC para el desglose de la calidad de las pruebas PC12 y PC8. En estas curvas no es tan directo ver la relación entre la calidad de la huella lanzada y el porcentaje de aparición de la huella genuina en la lista de candidatos. El motivo principal de esto es la escasez de puntos característicos, que hacen que la encontrar puntos coincidentes por parte del comparador, sea más determinante que la calidad de la huella lanzada.

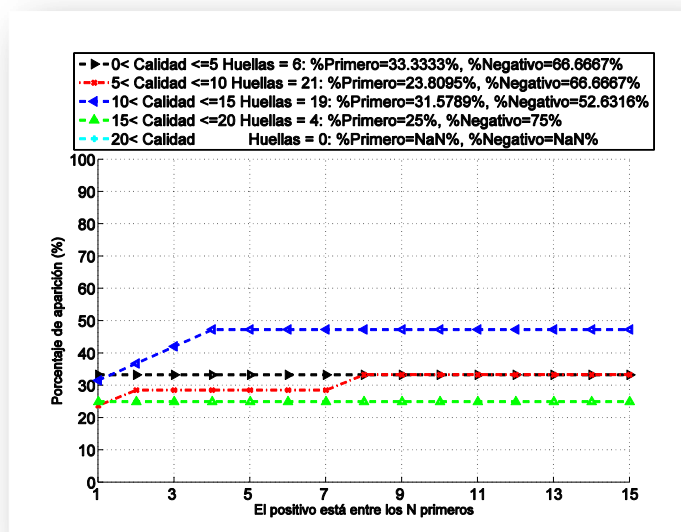


Figura 5.4.3 CMC de variación de calidad para prueba PC12 en latentes

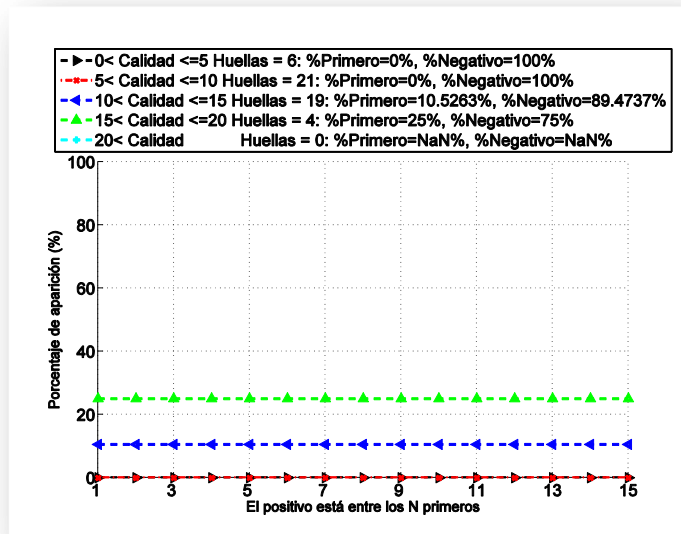


Figura 5.4.4 CMC de variación de calidad para prueba PC8 en latentes

5.4.2 Experimentos con impresiones posadas

La calidad en impresiones dactilares varía en rangos más altos (“Q” más baja) que en huellas latentes, ya que las condiciones de adquisición de aquellas son notablemente más óptimas que en el caso de las latentes. En la figura 5.4.5 aparecen las curvas CMC donde se desglosa la calidad de la huella lanzada para la prueba PCAUTO.

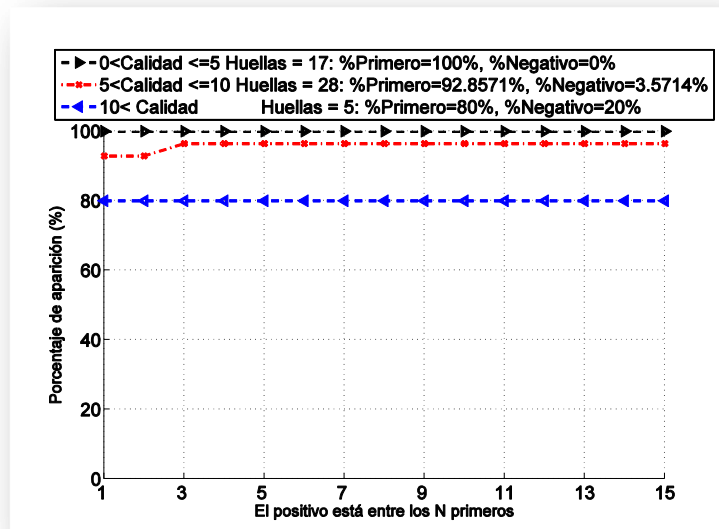


Figura 5.4.5 CMC de pruebas de variación de la calidad para PCAUTO en posadas

Ello nos permite ver la relación entre la calidad de la huella lanzada y el porcentaje de positivos. Esto se debe a que el número de puntos característicos es en general alto, siendo posible encontrar muchas coincidencias por lo que la calidad se convierte en un factor importante.

Sin embargo, en las figuras 5.4.6 y 5.4.7, donde se muestra el mismo experimento pero para las pruebas PC12 y PC8, se hace mucho más difícil ver la relación entre calidad y porcentaje de aparición de positivos, ya que al no tener todos los puntos localizados en la huella, la calidad pasa a ser un factor de menor importancia.

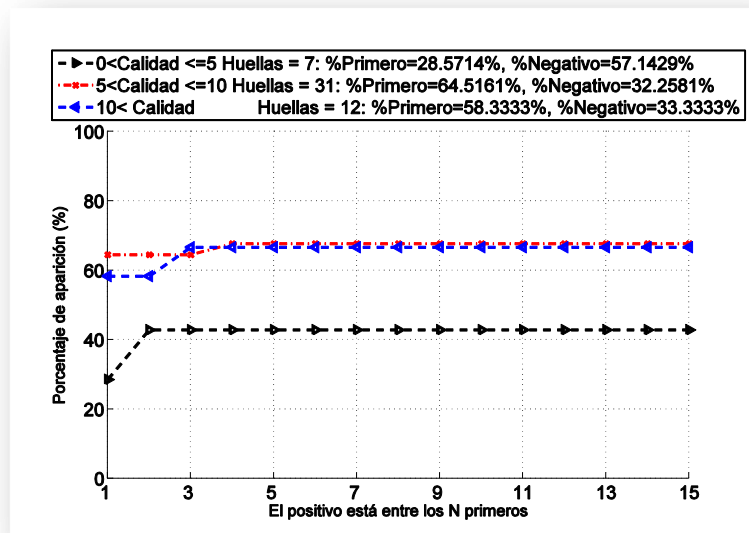


Figura 5.4.6 CMC de pruebas de variación de calidad para PC12 en posadas

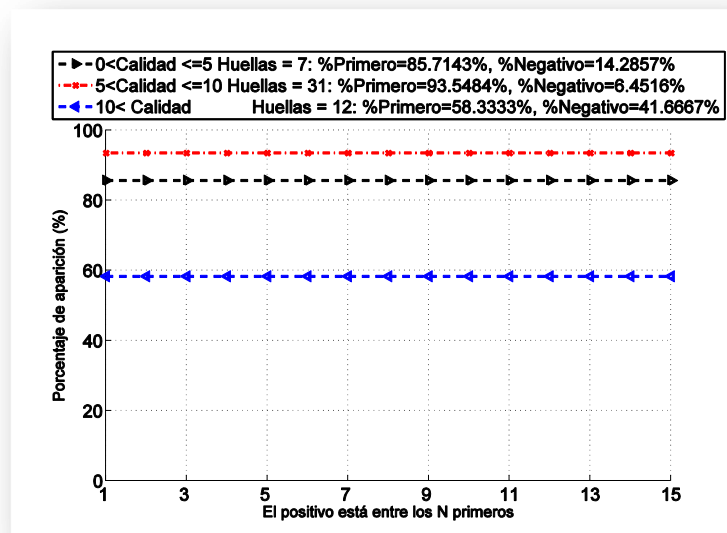


Figura 5.4.7 CMC de pruebas de variación de calidad para PC8 en posadas

5.4.3 Experimentos con impresiones rodadas

En las pruebas de variación de la calidad para impresiones rodadas, no incluimos la prueba PCAUTO porque el porcentaje de positivos en primera posición es del 100%, por lo que la calidad no mostrará variación en los resultados.

El caso de las pruebas PC12 y PC8 en rodadas se muestra en las figuras 5.4.8 y 5.4.9. En ambas se puede observar que la relación entre la calidad de la huella lanzada y el porcentaje de huellas genuinas que aparecen como primer candidato es directa.

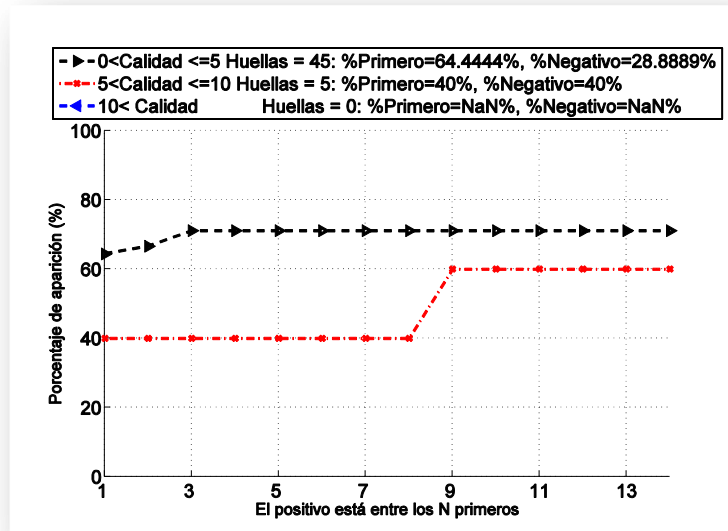


Figura 5.4.8 CMC de pruebas de variación de calidad para PC12 en rodadas

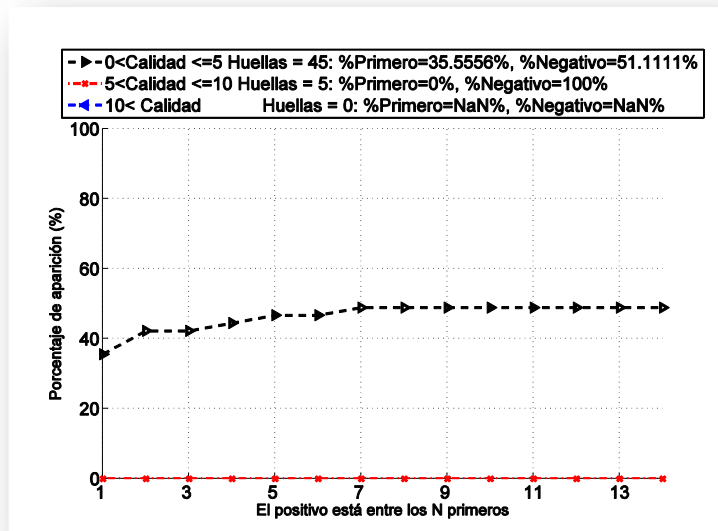


Figura 5.4.9 CMC de pruebas de variación de calidad para PC8 en rodadas

6

Rendimiento del sistema en modo verificación

En este capítulo se analizará el sistema funcionando en modo verificación. Para ello sólo es necesario variar un umbral de decisión que determina si los scores proceden de comparaciones genuinas o no genuinas. A partir de cada umbral, se calculan las tasas de falsa aceptación y falso rechazo del sistema y el punto de igual error EER, con los cuáles puede analizarse el rendimiento del sistema si funciona en modo verificación.

Además, se realizará la normalización de scores en cada conjunto o lista de candidatos, con el objetivo de mejorar el poder de discriminación global del sistema.

Para la observación del rendimiento del sistema, su poder de discriminación y las mejoras introducidas por la normalización, se representarán y analizarán distintas curvas.

En primer lugar se representarán las distribuciones de scores por medio de faunagramas. Estas figuras permiten ver fácilmente las distribuciones de scores de las distintas comparaciones y compararlas. Con esto se puede ver en qué rangos varían los scores procedentes de comparaciones genuinas y los procedentes de comparaciones no genuinas. También son muy útiles para observar el alineamiento de los scores entre las distintas comparaciones, antes y después de la normalización.

También se representan curvas *Tippett* o FA/FR, que contienen las funciones de distribución de scores *target* y *non-target*.

Finalmente se representan las curvas DET. En cada una de las figuras se muestran las curvas correspondientes al mismo conjunto de scores, antes y después de la normalización. Con esto se logra que el cambio en el poder de discriminación del sistema pueda ser fácilmente comprobado de manera visual. Además en cada una de las figuras se muestra la tasa EER antes y después de la normalización.

6.1 Distribuciones de scores

En este apartado se muestran algunos de los faunagramas obtenidos de los experimentos de variación de puntos característicos, antes y después de un proceso de normalización utilizando *T-norm*. Podrá observarse así el efecto de la normalización en el alineamiento de los scores entre distintas comparaciones. Se mostrarán también curvas *Tippett* para los scores con y sin normalizar, pudiéndose observar así la mejora del poder de discriminación del sistema.

6.1.1 Experimentos con huellas latentes

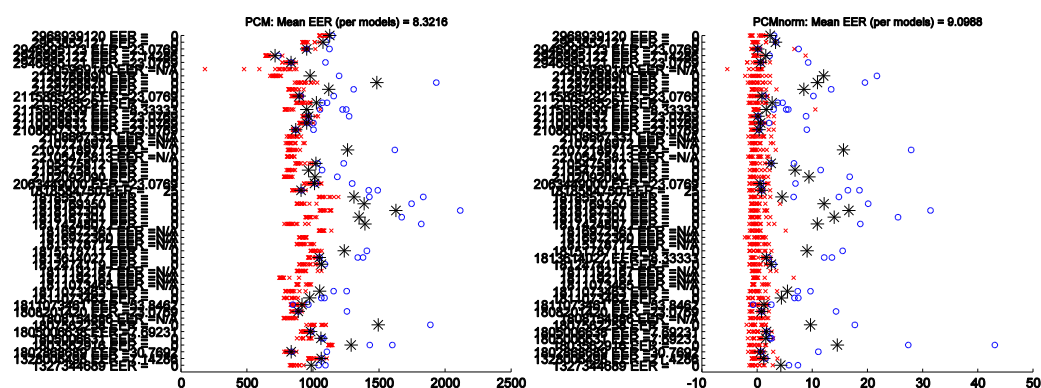


Figura 6.1.1 Faunagramas de los experimentos PCM con huellas latentes, sin y con normalización.

En la figura 6.1.1 podemos observar el efecto de la normalización para los experimentos con huellas latentes. A la izquierda se presenta la distribución de scores dados por el sistema para la prueba PCM realizada con la colección de huellas latentes. A la derecha tenemos los scores normalizados para los mismos experimentos. Como puede apreciarse, se produce un alineamiento de los scores *non-target*, marcados en rojo, en el conjunto de scores normalizados respecto al conjunto sin normalizar.

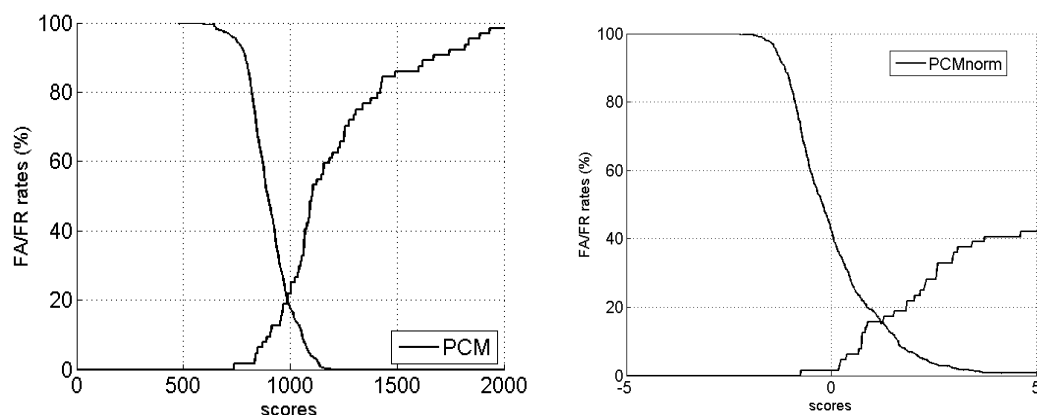


Figura 6.1.2 Curvas FA/FR sin y con normalización para experimentos PCM con latentes.

En las curvas representadas en la figura 6.1.2 puede observarse el efecto de la normalización en la distribución de las tasas de error en el sistema en función del umbral aplicado. Se puede observar sobre todo una gran disminución del error de falso rechazo (curva creciente) en el conjunto normalizado con respecto del conjunto sin normalizar.

6.1.2 Experimentos con impresiones posadas

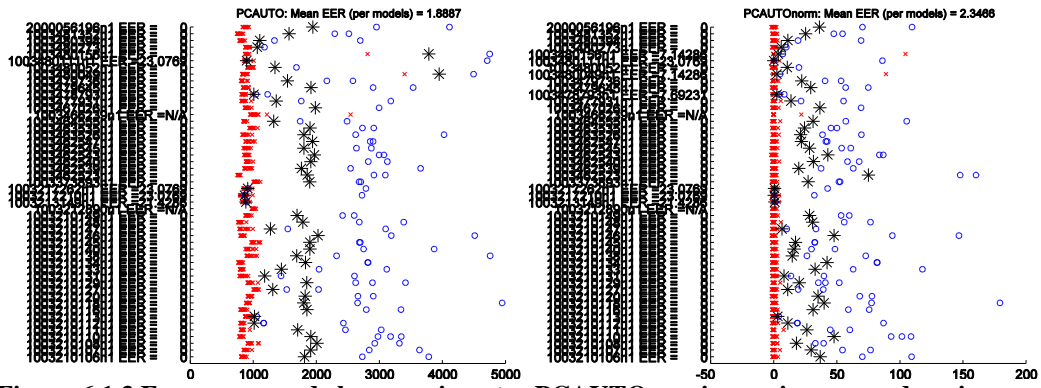


Figura 6.1.3 Faunagramas de los experimentos PCAUTO con impresiones posadas, sin y con normalización.

En el caso de las impresiones dactilares, el efecto de alineamiento causado por la normalización de scores se hace mucho más notable, tal y como puede observarse en la figura 6.1.3 donde se representan los faunagramas de la prueba PCAUTO. Gracias a eso, mejora el poder de discriminación del sistema, reduciéndose las tasas de error tal y como se observa en la figura 6.1.4, donde encontramos las curvas FA/FR para el experimento PCAUTO de estas impresiones.

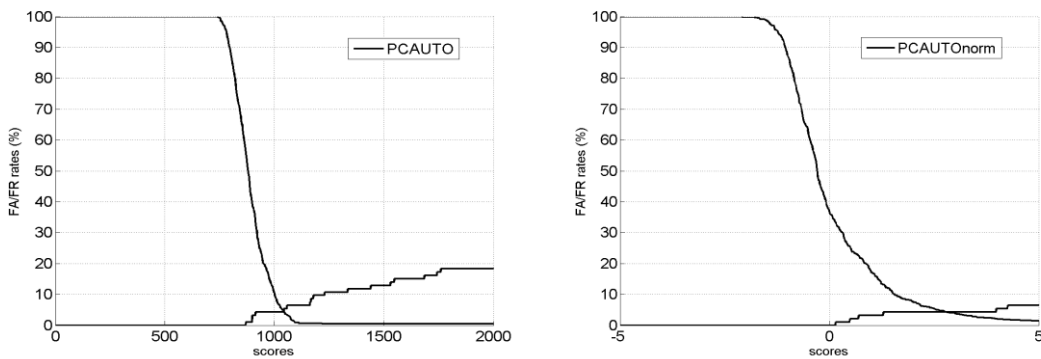


Figura 6.1.4 Curvas FA/FR sin y con normalización para experimentos PCAUTO con impresiones posadas.

6.1.3 Experimentos con impresiones rodadas

Con los faunagramas para scores con y sin normalización de la prueba PCAUTO para impresiones rodadas, representados en la figura 6.1.5, observamos los mismos efectos que en las impresiones posadas. Se produce un gran alineamiento de los scores *non-target*. Sin embargo en este caso, el alineamiento de los scores ya es bastante alto en la figura sin normalizar (izquierda).

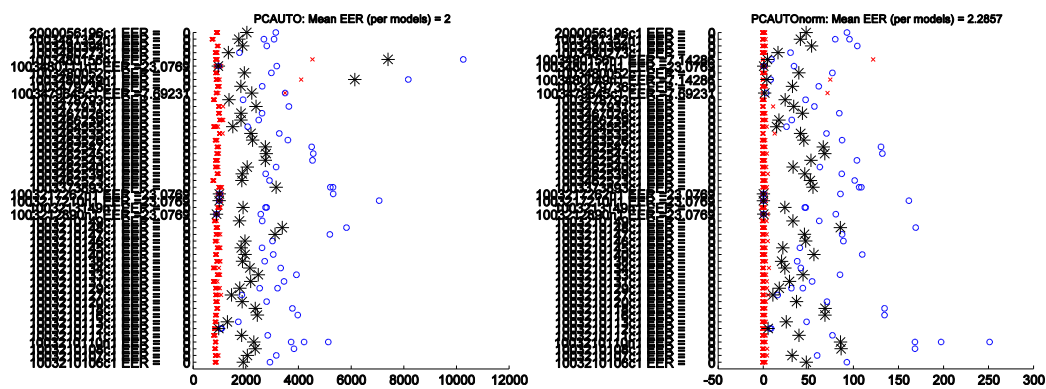


Figura 6.1.5 Faunagramas de los experimentos PCAUTO con impresiones rodadas, sin y con normalización.

En las curvas FA/FR para impresiones rodadas, que aparecen en la figura 6.1.6, se observa el efecto de la normalización (figura derecha) frente a las distribuciones de scores no normalizados (curva de la izquierda). Se observa que la tasa de falsa aceptación apenas varía entre una figura y otra, mientras que la de falso rechazo se reduce considerablemente con la normalización.

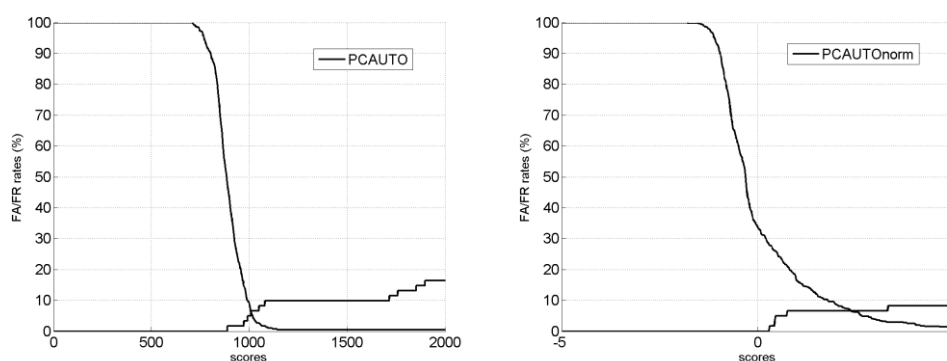


Figura 6.1.6 Curvas FA/FR sin y con normalización para experimentos PCAUTO con impresiones rodadas.

6.2 Curvas DET

Las curvas DET (*Detection Error Tradeoff*) se utilizan de manera extendida en biometría ya que ofrecen la posibilidad de comparar el poder discriminativo de distintos sistemas de forma rápida y visual. En estas curvas, los ejes x e y representan la tasa de falsa aceptación y la tasa de falso rechazo del sistema, respectivamente. Para comparar dos o más sistemas, basta con saber que cuando más cercana está la curva del origen de coordenadas, mejor es el poder de discriminación [22].

En la bisectriz de los ejes, tenemos el punto EER, en el cual la tasa de falsa aceptación y la tasa de falso rechazo son iguales. Este punto se utiliza como medida del poder de discriminación del sistema y sirve como medida para la comparación entre distintos sistemas o distintos conjuntos de scores.

En este apartado se representan las curvas DET de todos los experimentos para los scores con y sin normalizar. La curva roja punteada representa la de los scores normalizados, mientras que la curva negra sólida representa la de los scores sin normalizar, tal y como los da el sistema. Esto nos permite ver el rendimiento del sistema y la mejora de éste cuando normalizamos los scores que el sistema devuelve.

6.2.1 Experimentos con huellas latentes

En la figura 6.2.1 se representan las curvas DET de los experimentos realizados con huellas latentes. Además se muestra el valor del EER para todos ellos, para los resultados con y sin normalización. En estas curvas se puede observar un empeoramiento del rendimiento del sistema con la reducción del número de puntos característicos utilizados en las comparaciones. Como es de esperar, el EER es mayor conforme disminuimos el número de puntos característicos en las huellas lanzadas.

Además, se observa en cada una de las figuras que la normalización de scores produce un efecto positivo en el rendimiento, acercando las curvas provenientes de scores normalizados al origen de coordenadas. Se observa que para los experimentos PCAUTO, PCM y PC12, se produce una mejora del EER. Por el contrario, el EER aumenta para la prueba PC8, aunque esto es debido a una imprecisión en el cálculo de EER con pocos datos, dado que en la curva DET no se observa casi diferencia. Cabe destacar que en esta última prueba los resultados no son muy representativos debido a que en el 94% de los casos, la huella *target* no aparece en la lista.

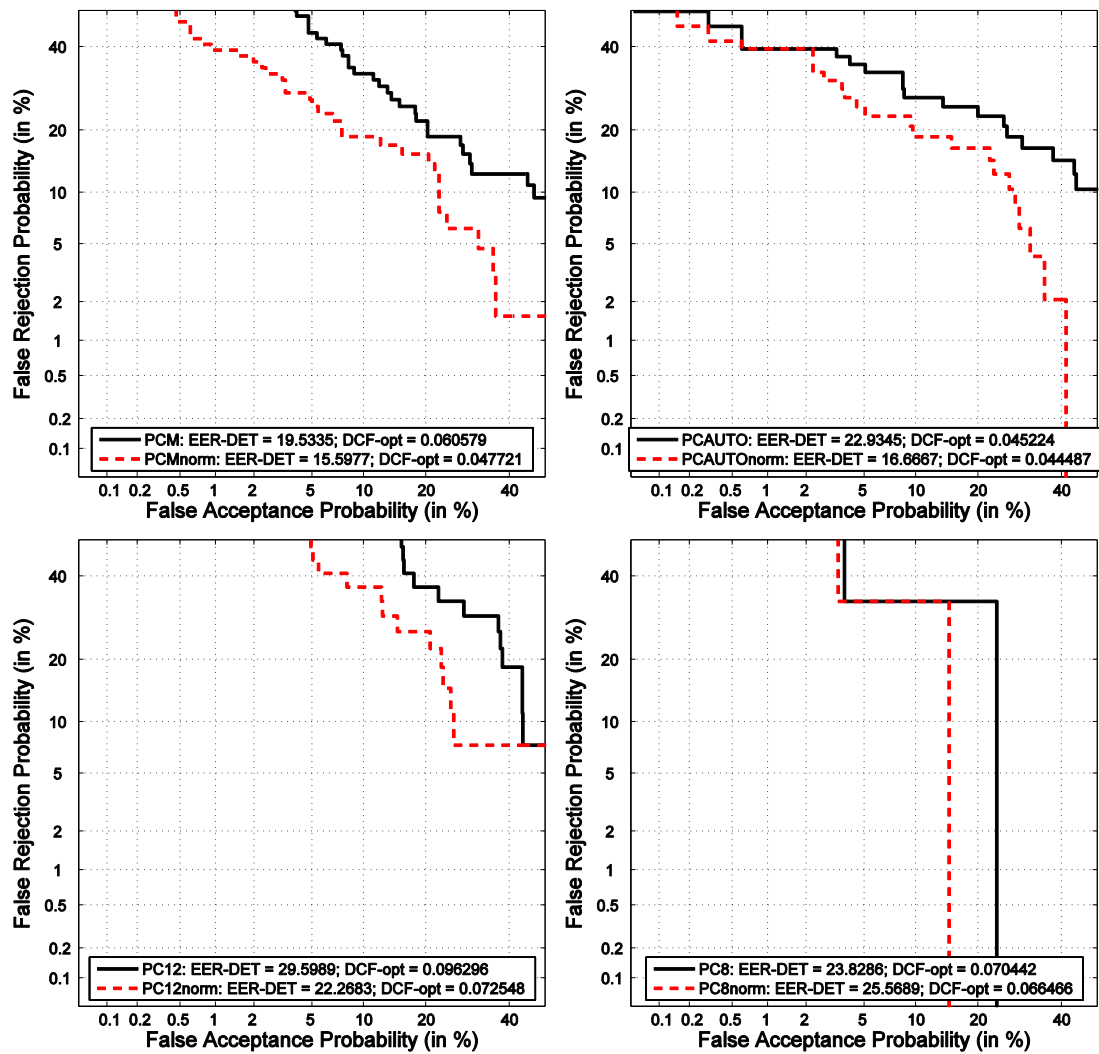


Figura 6.2.1 Curvas DET para experimentos con huellas latentes.

6.2.2 Experimentos con impresiones posadas

En la figura 6.2.2 se muestran las curvas DET calculadas con los scores obtenidos de los experimentos con impresiones posadas. En ellas se puede observar como el rendimiento empeora cuando se disminuye el número de puntos característicos de las huellas lanzadas: las curvas se alejan del origen y el EER aumenta.

Además, dentro de cada figura, se observa cómo la curva DET obtenida con scores normalizados presenta una tasa EER más baja en cada caso, siendo muy notable la mejora en el caso de PC8, donde se produce una mejora absoluta del 11 %. Esto se debe al efecto del alineamiento entre comparaciones diferentes.

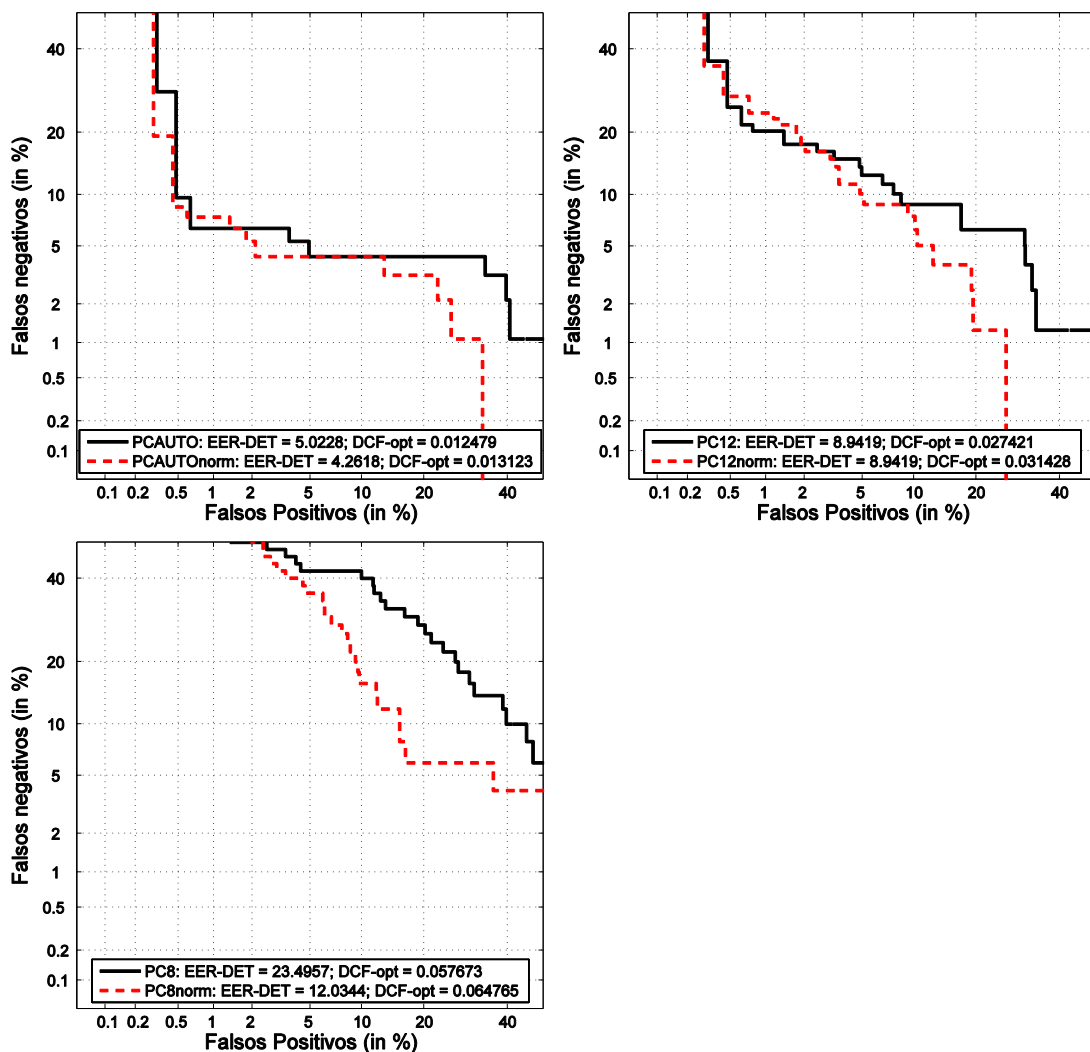


Figura 6.2.2 Curvas DET para experimentos con impresiones posadas.

6.2.3 Experimentos con impresiones rodadas

En la figura 6.2.3 se muestran las curvas DET calculadas con los scores obtenidos de los experimentos con impresiones rodadas. En ellas se puede observar como el rendimiento empeora cuando se disminuye el número de puntos característicos de las huellas lanzadas: las curvas se alejan del origen y el EER aumenta.

Además, dentro de cada figura, se observa cómo la curva DET obtenida con scores normalizados presenta una tasa EER más baja en cada caso. Sin embargo, la DET empeora en algunas regiones, debido a que los datos ya estaban muy bien normalizados para huellas posadas, y T-Norm no mejora en todos los casos.

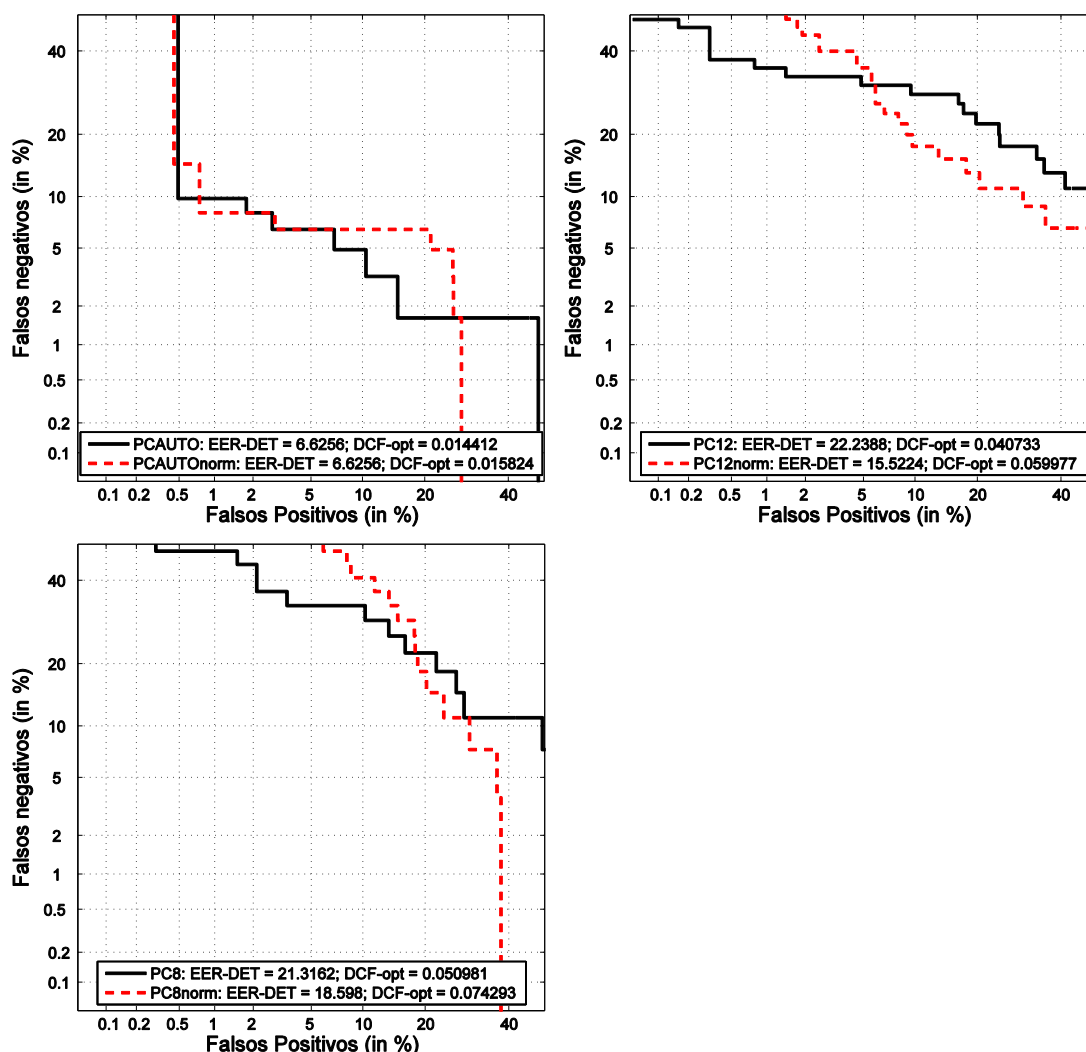


Figura 6.2.3 Curvas DET para experimentos con impresiones rodadas.

7

Cálculo de relaciones de verosimilitud (LR)

En este capítulo se presentan los resultados en el cálculo de LR para todos los experimentos realizados de variación del número de puntos característicos de todos los conjuntos distintos de huellas. Estos resultados se muestran en forma de curvas ECE. La información relevante acerca de los métodos de cálculo de LR utilizados puede encontrarse también en el capítulo 3.

Por cada tipo de experimento se muestran 4 figuras, una para cada método de cálculo de LR. Para más información sobre estos cuatro métodos (regresión lineal, gaussiano, PAV y KDF) se refiere al lector al capítulo 3 de este documento.

Para la comodidad del lector, se resume a continuación el significado de cada ECE en la figura de representación de resultados.

- La curva sólida es la entropía cruzada, es decir, la pérdida media de información de los valores de LR calculados. Cuanto más alta es esta curva, más información se necesita para saber cuál de las dos hipótesis enfrentadas es la verdadera, y por lo tanto, peor es el poder de discriminación del sistema. Además de esta curva, se representan otros dos sistemas para la comparación
- La curva de puntos azules representa el sistema calibrado, que optimiza la ECE conservando la discriminación y obtenida por el algoritmo PAV (*Pool Adjacent Violators*) [31]. Cuanto más cercanas estén la curva azul y la roja, mejor calibrado estará nuestro sistema. Además, la curva azul es una medida del poder de discriminación del sistema, y por tanto curvas DET iguales dan lugar a curvas azules iguales.
- La curva negra puntuada representa el rendimiento de un sistema con $LR=1$ siempre, conocido como sistema neutral. Esta curva nos sirve como referencia para la curva sólida, que siempre debe estar por debajo para que el sistema tenga algún poder de discriminación. Si la curva sólida estuviera por encima, el sistema estaría perdiendo más información con el cálculo de LR que si la decisión se basara solamente en la información inicial del caso.

7.1 LR en experimentos con huellas latentes

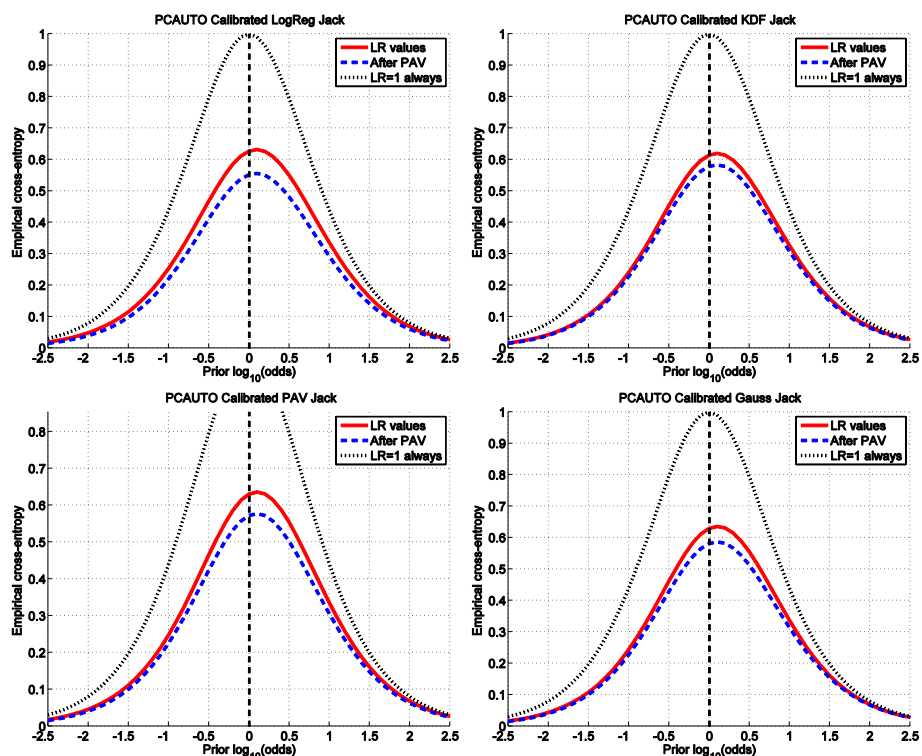


Figura 7.1.1 LR para los *scores* dados por el sistema en PCAUTO para experimentos con latentes

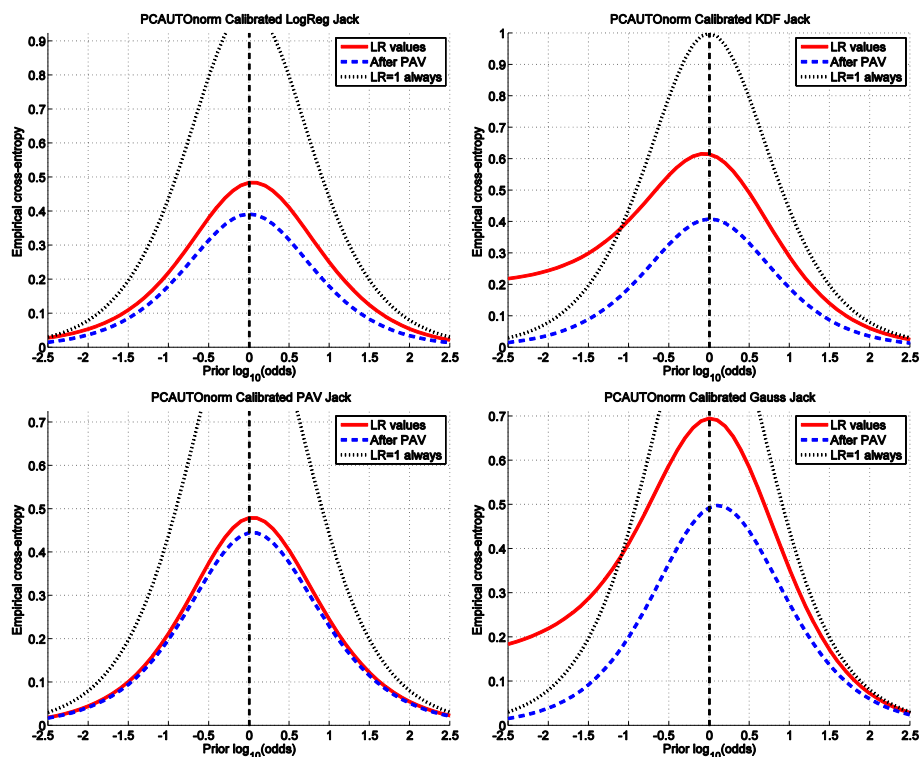


Figura 7.1.2 LR para los *scores* normalizados en PCAUTO para experimentos con latentes

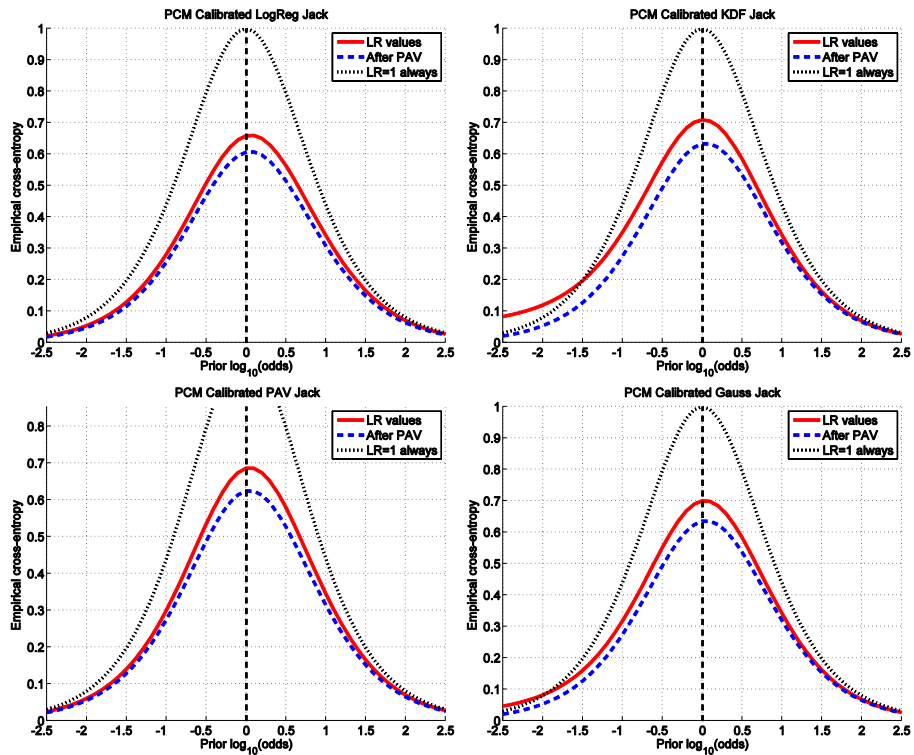


Figura 7.1.3 LR para los *scores* dados por el sistema en PCM para experimentos con latentes

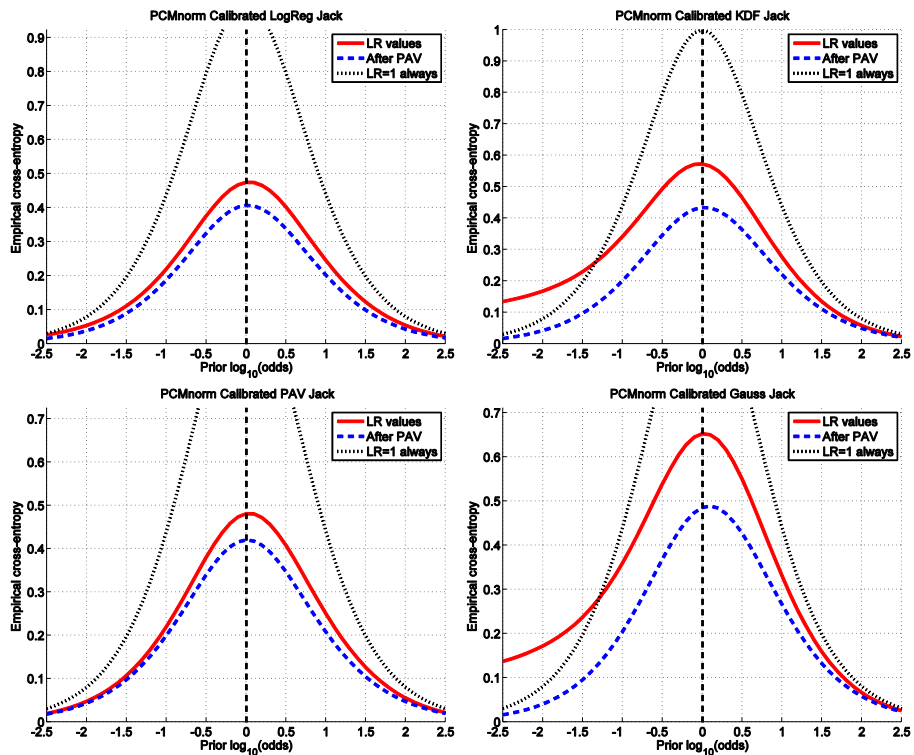


Figura 7.1.4 LR para los *scores* normalizados en PCM para experimentos con latentes

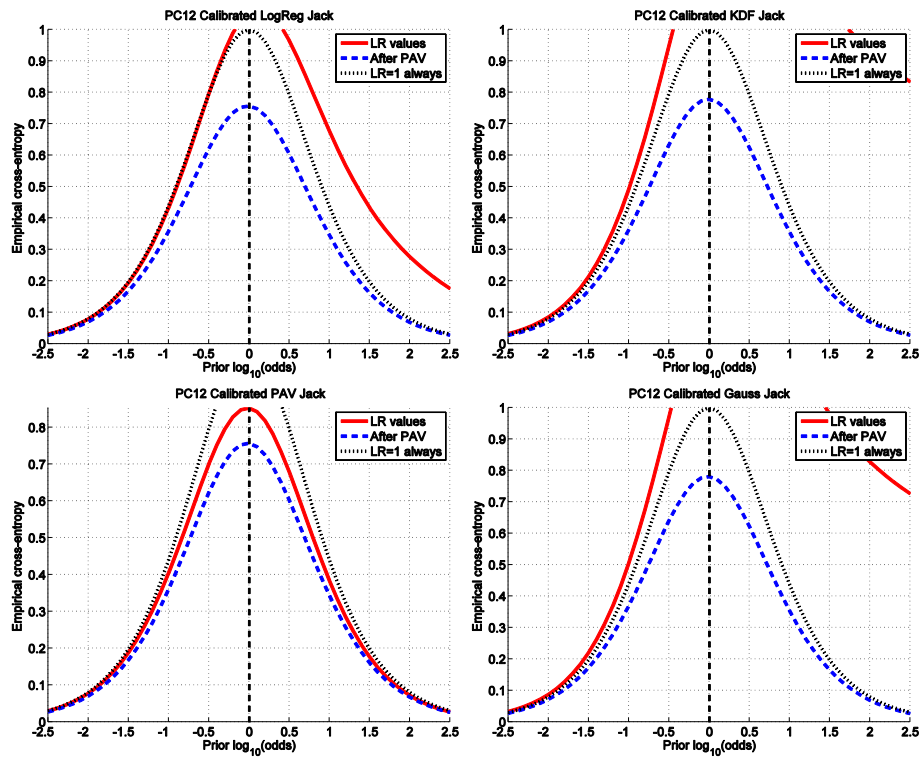


Figura 7.1.5 LR para los *scores* dados por el sistema en PC12 para experimentos con latentes

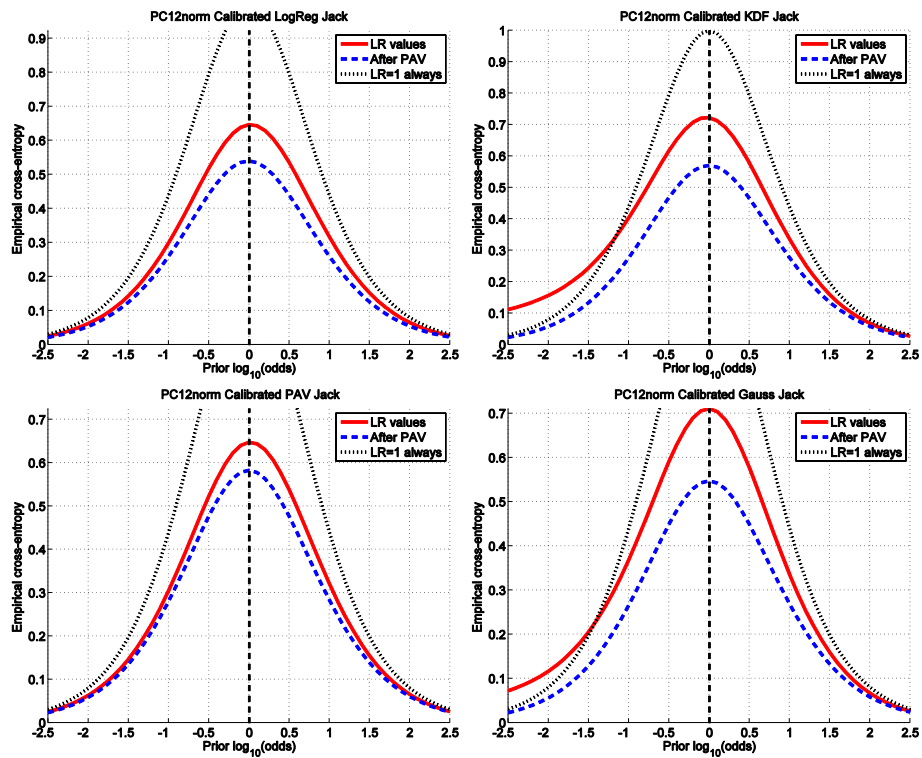


Figura 7.1.6 LR para los *scores* normalizados en PC12 para experimentos con latentes

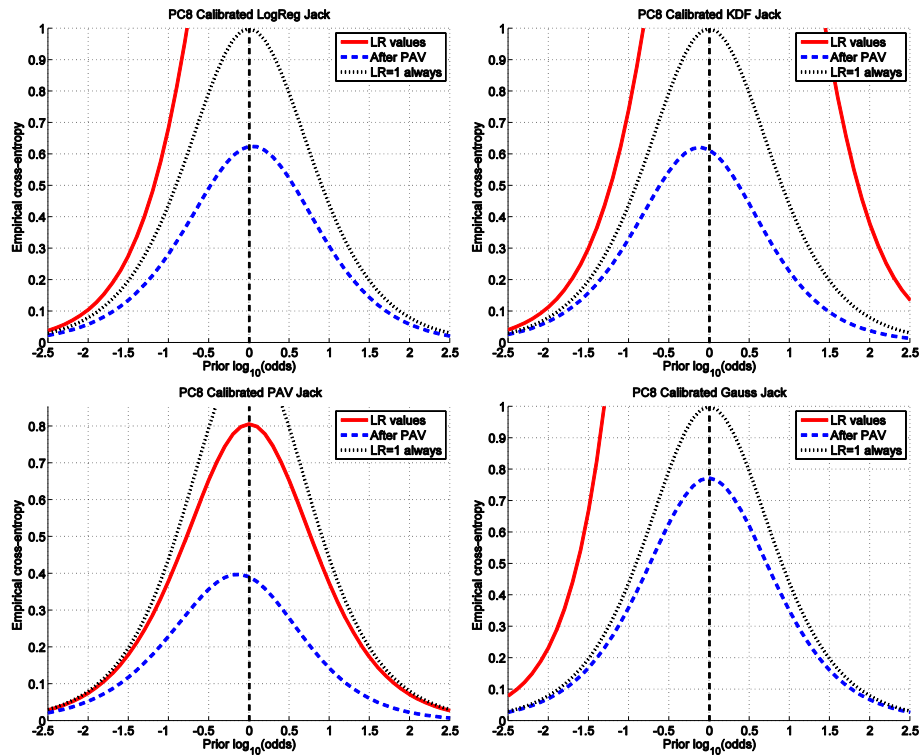


Figura 7.1.7 LR para los *scores* dados por el sistema en PC8 para experimentos con latentes

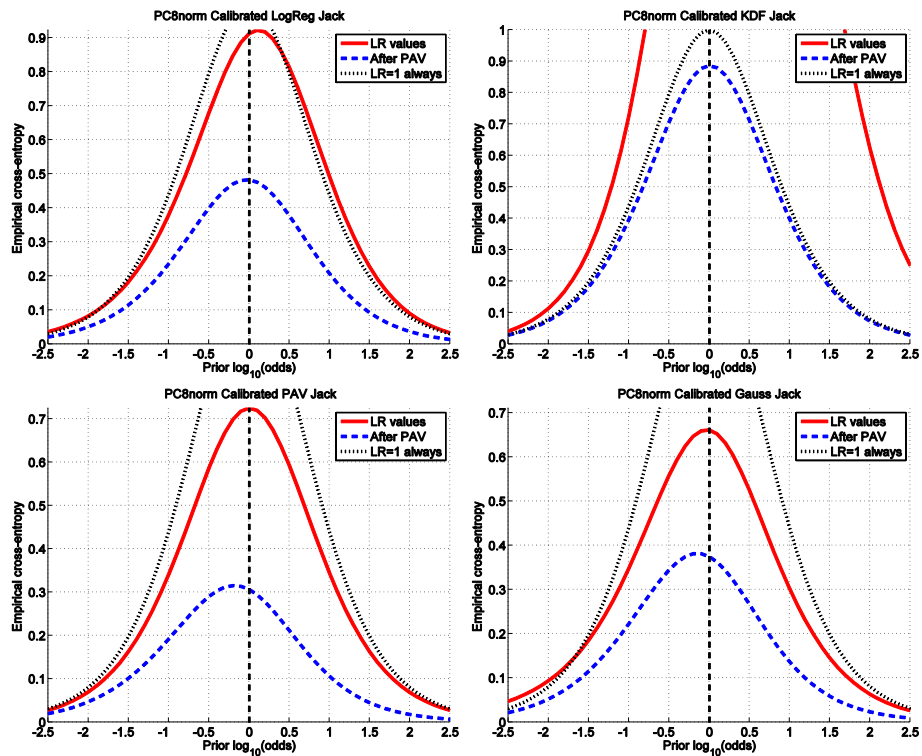


Figura 7.1.8 LR para los *scores* normalizados en PC8 para experimentos con latentes

En las figuras 7.1.1 a 7.1.8 se muestran en forma de curvas ECE los resultados en el cálculo de LR para los experimentos con huellas latentes. Se observan los siguientes efectos:

- El rendimiento de los LR para las pruebas PCAUTO y PCM es, en general, satisfactorio. Para los LR calculados a partir de los scores sin normalizar, se observa que para los cuatro métodos de cálculo la curva roja aparece bastante por debajo de la curva negra punteada, lo cual denota un buen poder de discriminación del conjunto. Además, la proximidad de la curva sólida roja a la curva rayada (azul), denota una buena calibración del sistema, lo cual permite una interpretación probabilística fiable de los resultados. En cambio, para los scores normalizados, encontramos que los distintos métodos no son tan homogéneos. EN el caso de regresión logística y PAV, el rendimiento de los LR mejora. Sin embargo, para KDF y *gaussian*, encontramos que la curva roja crece y que aparecen colas que superan a la curva negra punteada. Esto puede ser debido a que la distribución de los scores genuinos difiera mucho de una gaussiana en el caso de modelado gaussiano, o que el modelado KDF sobreentrene sobre los scores genuinos tras normalizar, que presentan una dispersión mayor que antes de ser normalizados.
- Para la prueba PC12 se observa que el único método por el cual los LR tienen un buen poder de discriminación es PAV, ya que en los otros 3 métodos la curva sólida asciende por encima de la curva del sistema neutral. Por el contrario, cuando los LR se calculan a partir de los scores normalizados, las curvas mejoran bastante para los cuatro métodos, lo cual muestra la gran utilidad de la normalización de scores. Como en el caso anterior, se observa una mejor calibración para los métodos de regresión logística y PAV, por lo motivos ya explicados.
- En el caso de PC8, los resultados se muestran en concordancia con los presentados en resultados anteriores. El escaso número de comparaciones genuinas y el gran desalineamiento de los scores hacen que el rendimiento de los LR sea muy insatisfactorio. Se puede comprobar que con la normalización se consigue una mejora en el poder de discriminación, menos para el método KDF, siendo una vez más PAV el método más apropiado para el cálculo de LR tanto si los scores están normalizados como si no lo están.

Hay que destacar que estos resultados de LR son optimistas, en el sentido de que la base de datos de scores utilizada para normalizar y calcular LR es muy parecida a los scores a transformar en LR (de hecho, son los mismos scores, pero para asegurar la generalidad de los resultados se utiliza un procedimiento de validación cruzada). En conclusión, los resultados son esperanzadores, pero es necesario confirmarlos con conjuntos experimentales más abundantes y ricos en variabilidad.

7.2 LR en experimentos con impresiones posadas

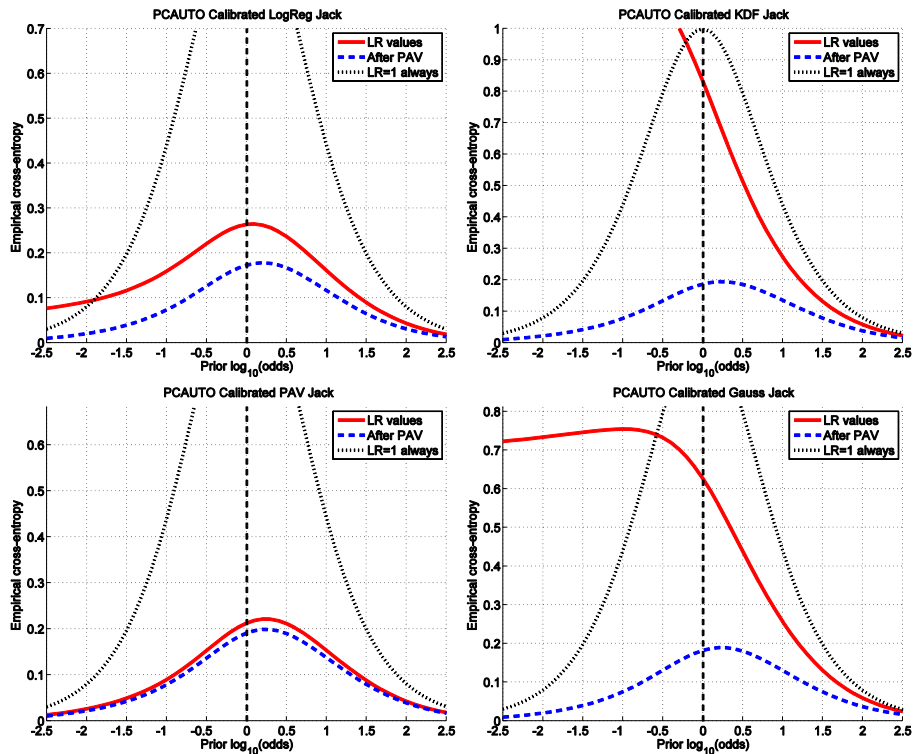


Figura 7.2.1 LR para los *scores* dados por el sistema en PCAUTO para experimentos con impresiones posadas

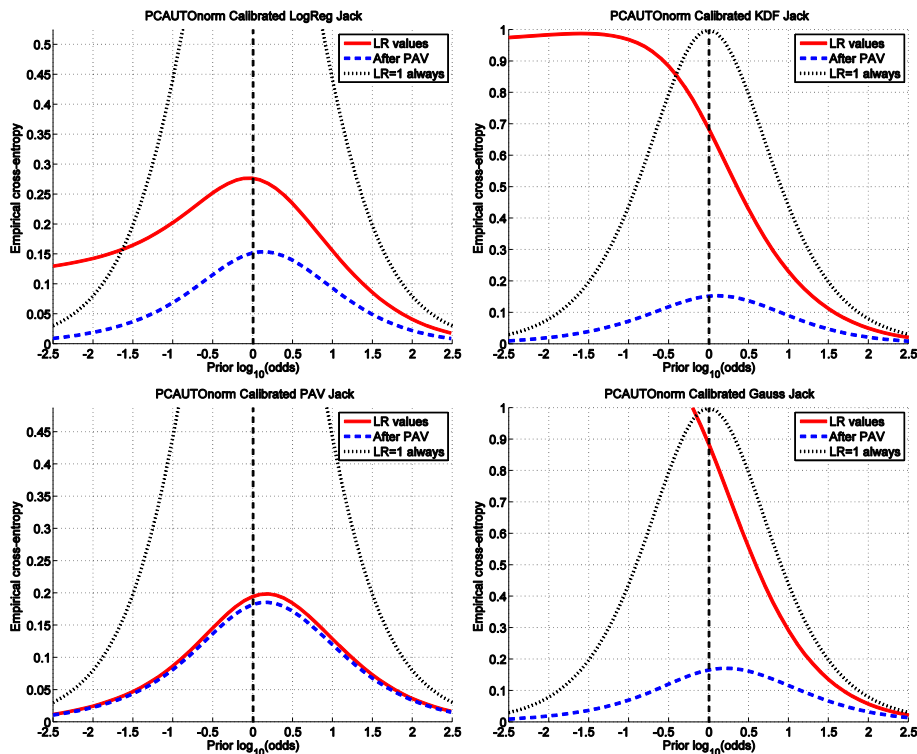


Figura 7.2.2 LR para los *scores* normalizados en PCAUTO para experimentos con impresiones posadas

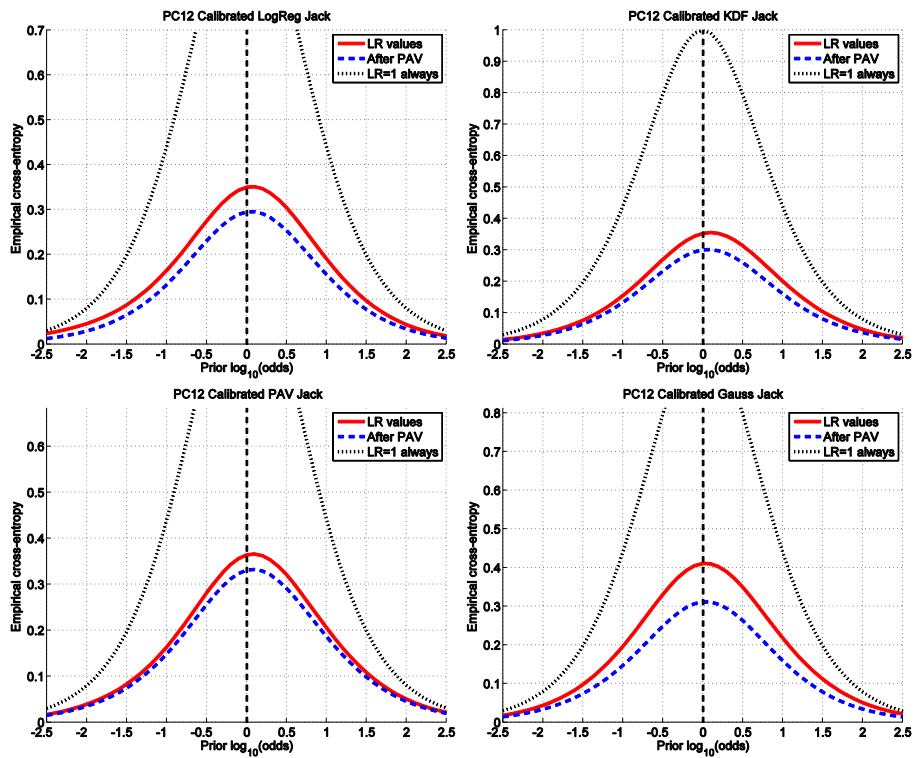


Figura 7.2.3 LR para los *scores* dados por el sistema en PC12 para experimentos con impresiones posadas

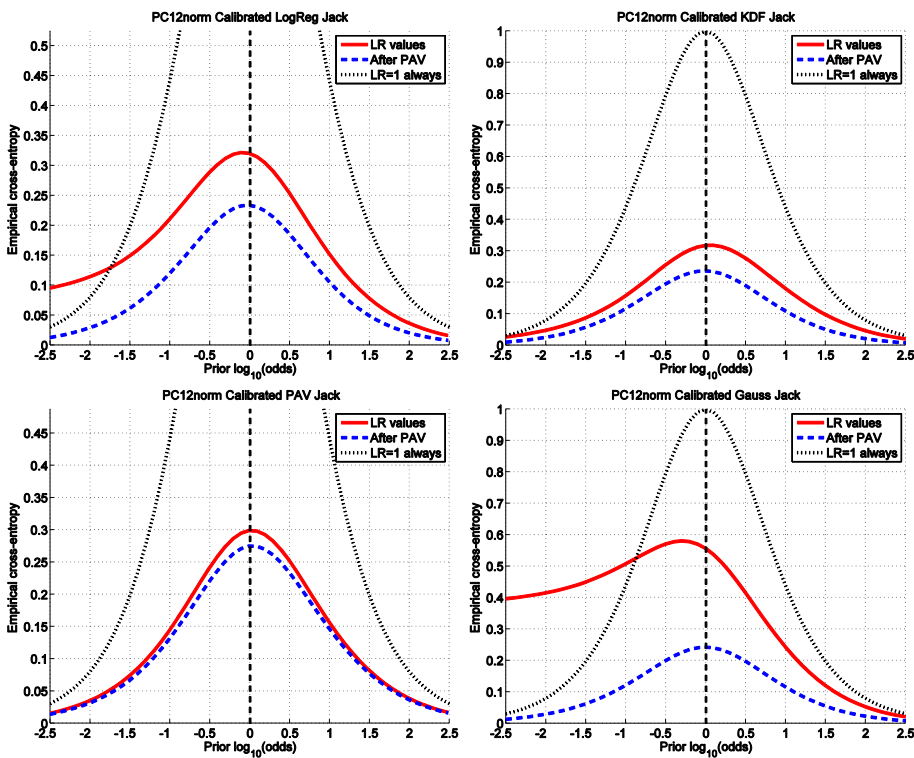


Figura 7.2.4 LR para los *scores* normalizados en PC12 para experimentos con impresiones posadas

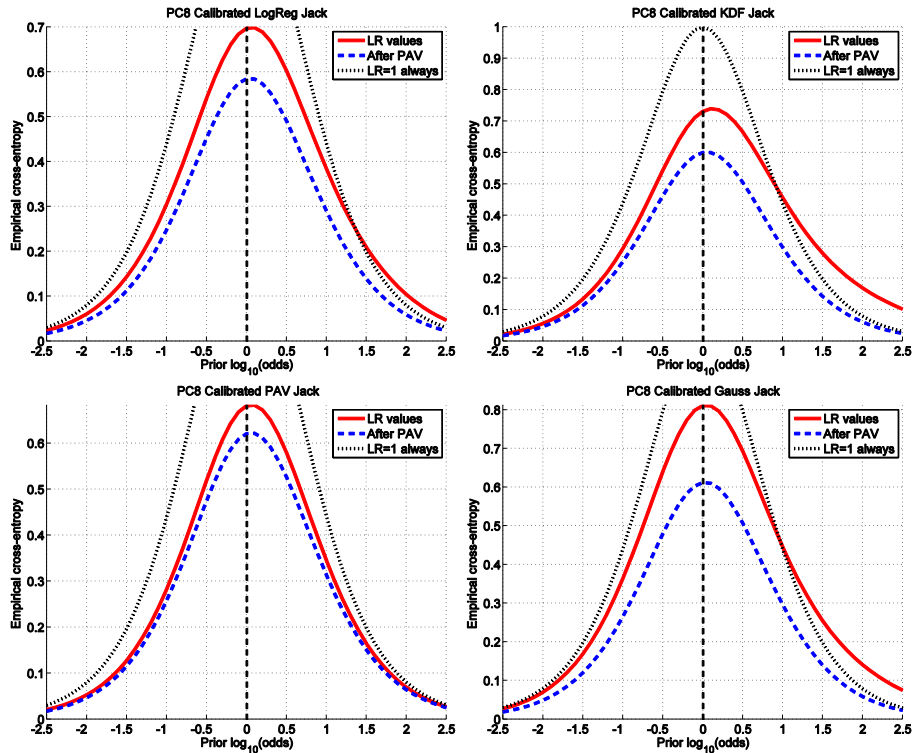


Figura 7.2.5 LR para los *scores* dados por el sistema en PC8 para experimentos con impresiones posadas

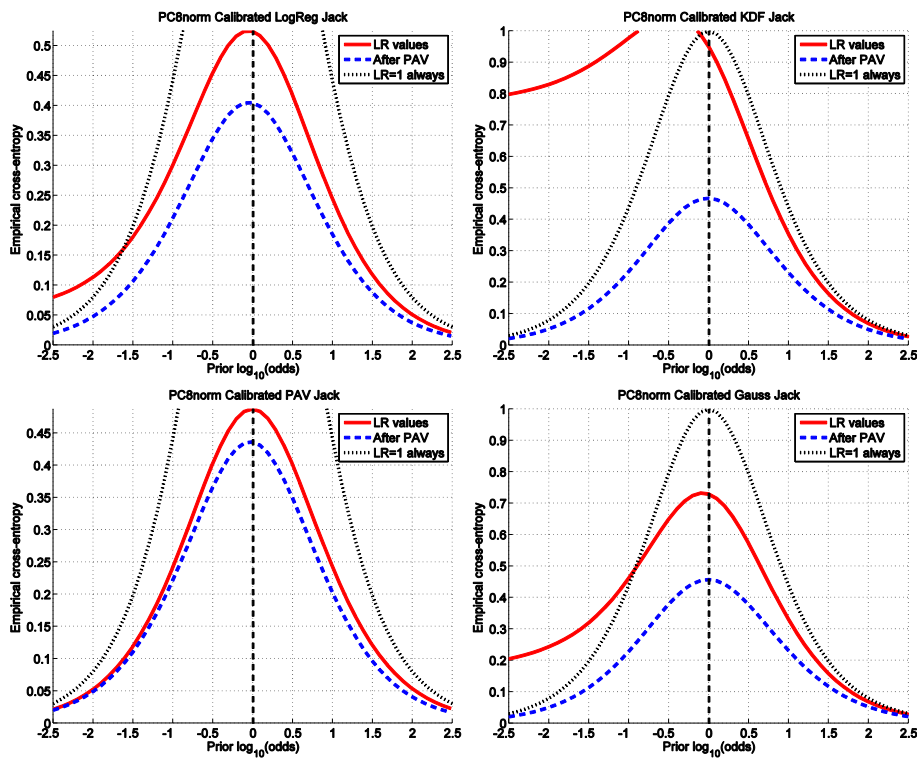


Figura 7.2.6 LR para los *scores* normalizados en PC8 para experimentos con impresiones posadas

En las figuras anteriores (7.2.1 a 7.2.6) se muestran los resultados del cálculo de LR para los experimentos con impresiones posadas en todas las pruebas. Se observa lo siguiente:

- El rendimiento de los LR para la prueba PCM difiere mucho según el método de cálculo de los mismos. Se observa que el poder de discriminación y la calibración son excelentes para el método PAV. Sin embargo, con el resto de métodos, especialmente KDF y gaussiano, aparece una gran inestabilidad tanto para scores normalizados como sin normalizar.
- En el caso de PC12, el rendimiento de los LR para todos los métodos a partir de scores sin normalizar es bueno. Sin embargo, cuando se calculan los LR a partir de scores normalizados, se producen grandes inestabilidades en los métodos de regresión logística y gaussiano, mejorando sólo en el caso de PAV.
- Para la prueba PC8 encontramos colas en los métodos gaussiano y KDF cuando los scores están sin normalizar. Cuando calculamos los LR tras la normalización de scores, encontramos que, una vez más, el método PAV mejora los resultados notablemente. Sin embargo, los otros tres métodos siguen sin aportar buenos resultados.

Con todo esto se puede seguir comprobando que el método PAV es el más adecuado para el cálculo de los LR, ya que siempre produce resultados en los que la curva sólida se sitúa por debajo de la curva negra punteada, lo cual significa que con el cálculo de LR por este método, se está ayudando a resolver el problema de atribución de fuentes.

7.3 LR en experimentos con impresiones rodadas

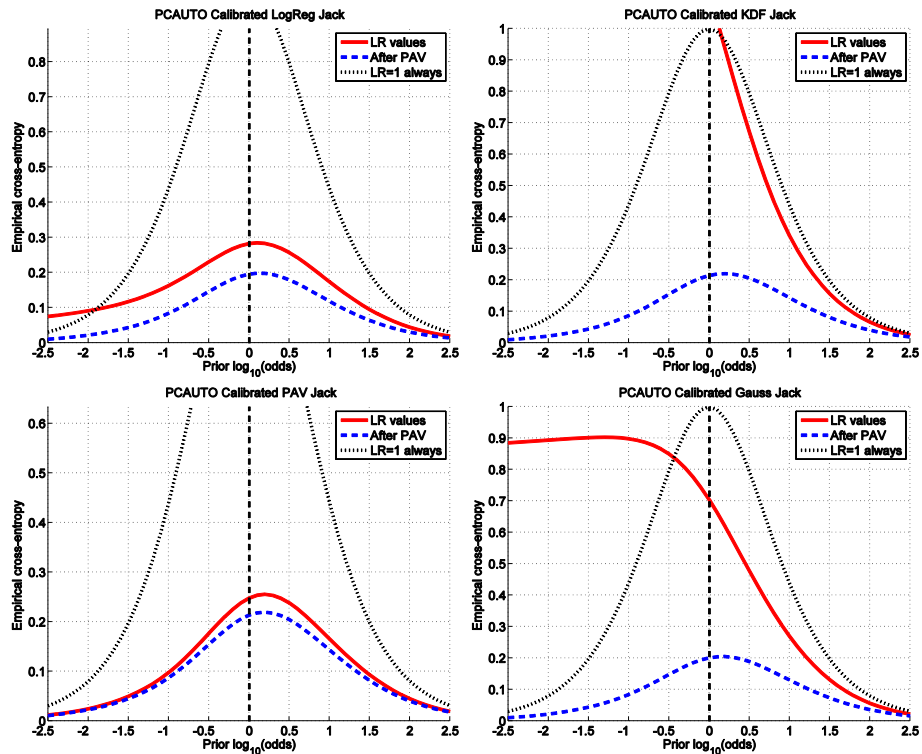


Figura 7.3.1 LR para los scores dados por el sistema en PCAUTO para experimentos con impresiones rodadas

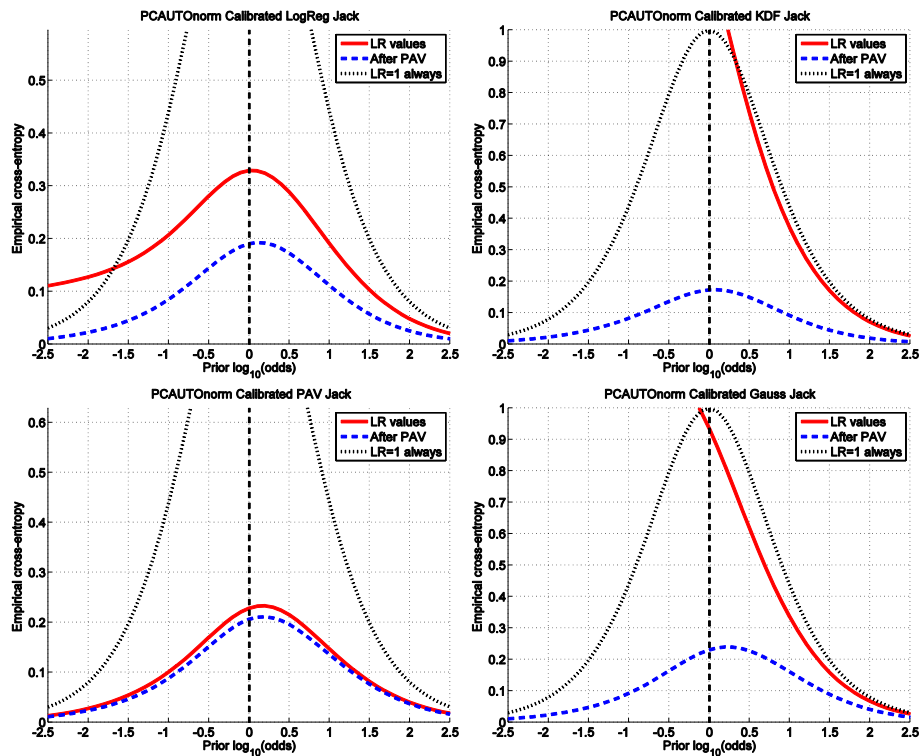


Figura 7.3.2 LR para los scores normalizados en PCAUTO para experimentos con impresiones rodadas

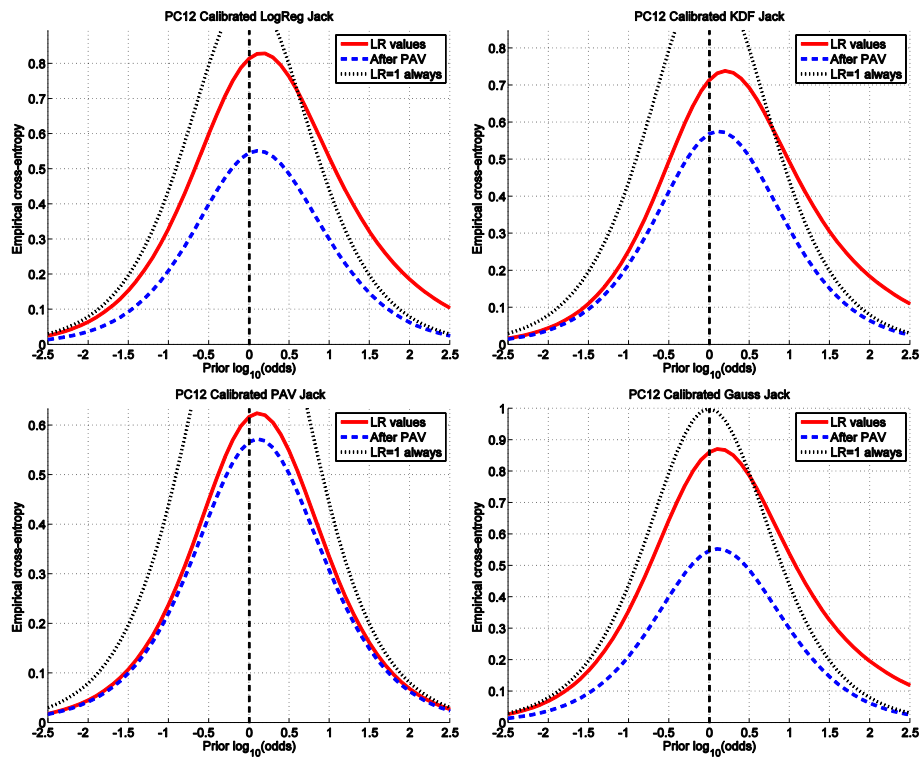


Figura 7.3.3 LR para los *scores* dados por el sistema en PC12 para experimentos con impresiones rodadas

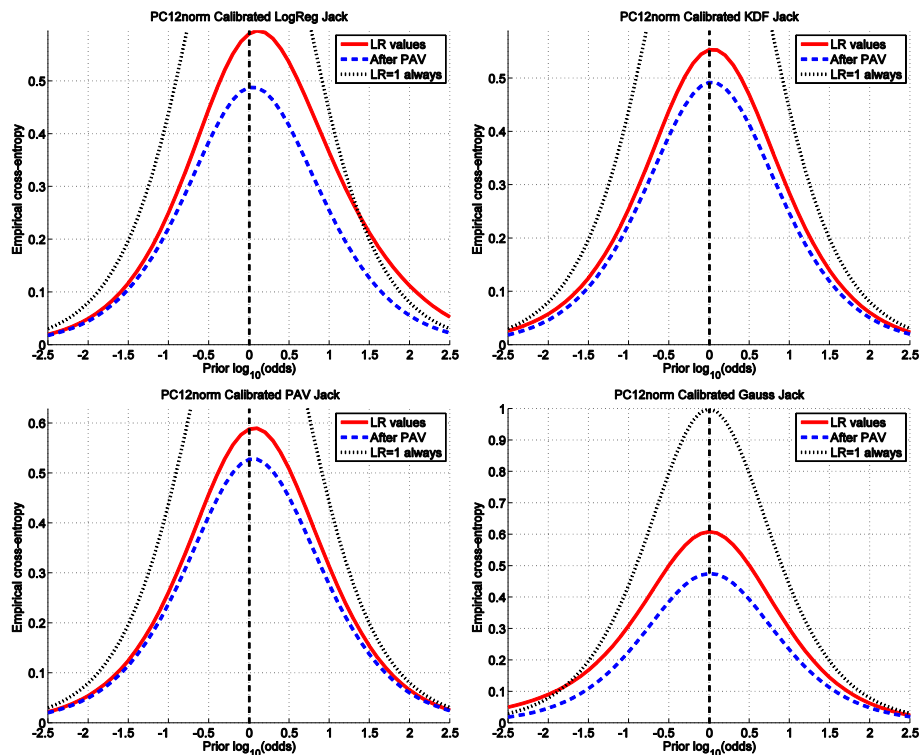


Figura 7.3.4 LR para los *scores* normalizados en PC12 para experimentos con impresiones rodadas

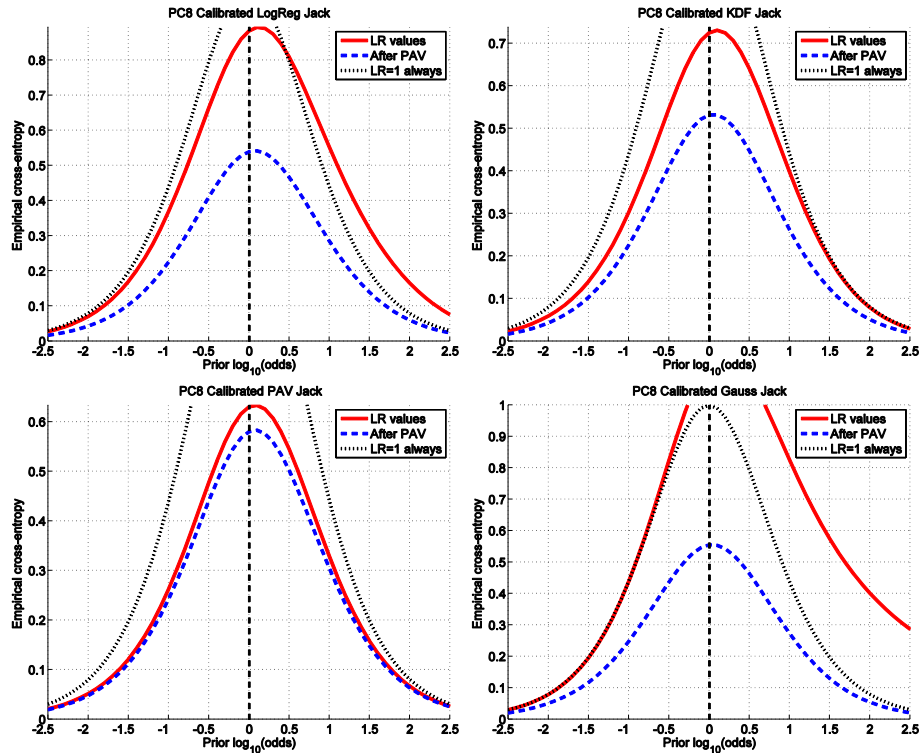


Figura 7.3.5 LR para los *scores* dados por el sistema en PC8 para experimentos con impresiones rodadas

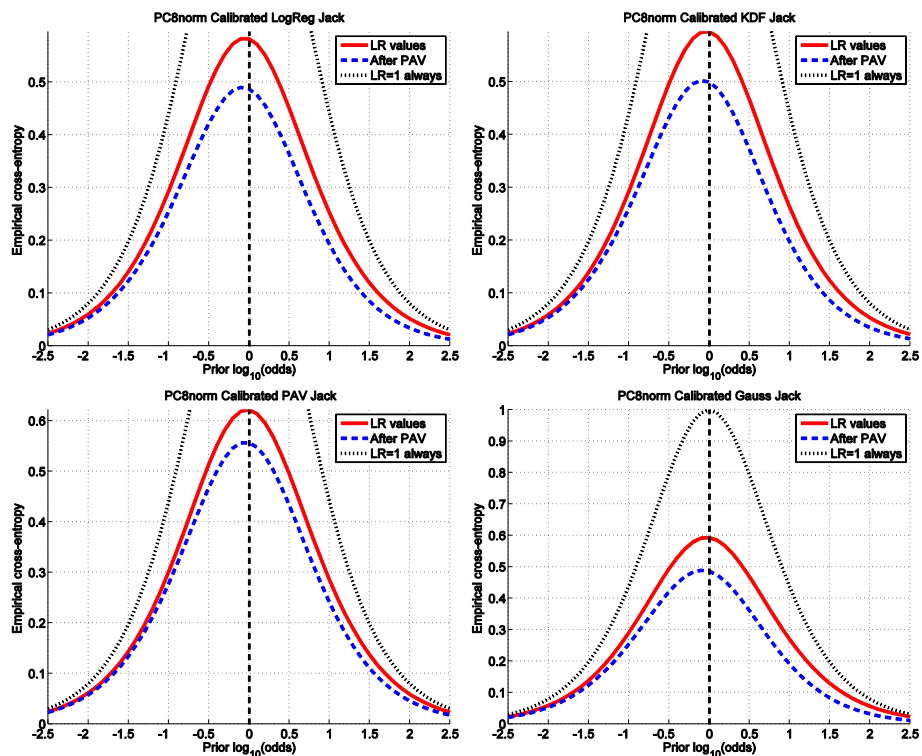


Figura 7.3.6 LR para los *scores* normalizados en PC8 para experimentos con impresiones rodadas

En las figuras anteriores (7.3.1 a 7.3.6) se muestran los resultados del cálculo de LR para los experimentos con impresiones rodadas en todas las pruebas. Se observa lo siguiente:

- En los LR calculados para la prueba PCM los resultados son muy buenos cuando se calculan por el método PAV. Para el resto de métodos se tiene una calibración nada satisfactoria. Cuando calculamos los LR con los scores normalizados, no se aprecian apenas diferencias en los resultados.
- Para la prueba PC12 se observan mejores resultados que en PCM. Para scores normalizados, la curva sólida presenta un buen poder de discriminación con los cuatro métodos con muy buena calibración, sobre todo en PAV. Cuando los scores no están normalizados, aparece inestabilidad en el poder de discriminación en todos los métodos menos en PAV.
- En el caso de PC8, análogamente con PC12, la normalización mejora mucho los resultados para todos los métodos. Antes de la normalización, las curvas para los métodos gaussiano, KDF y regresión logística son deficientes, mientras que las del método PAV muestran un buen poder de discriminación y una buena calibración, que sin embargo apenas varía con la normalización de los scores.

Una vez más se recuerda que los resultados son optimistas debido a que los scores utilizados para el entrenamiento son los mismos que los utilizados para calcular valores de LR. Además, vuelve a observarse que el método PAV aparece como el más adecuado para el cálculo de LR con todos los conjuntos de scores, ya que aporta buena calibración y buen poder de discriminación tanto para scores normalizados como para scores no normalizados.

8

Conclusiones y trabajo futuro

8.1 Conclusiones

Las conclusiones que pueden extraerse a partir de los resultados generados son muy numerosas. En este capítulo se enumeran clasificadas, así como se especifica el trabajo a futuro de continuación de este proyecto.

Conclusiones de los resultados de comparaciones entre huellas latentes y reseñas decadactilares:

- Se presentan resultados en los que se aprecia que las comparaciones entre huellas latentes y reseñas decadactilares presentan un rendimiento muy limitado. Además, son muy sensibles a la variación en el número de puntos característicos, degradándose muchísimo la fiabilidad del sistema cuando el número de puntos seleccionados es de 12 o menor. En el mejor de los casos, cuando se seleccionan todos los puntos en la latente manualmente, en un 72% de los casos la huella genuina aparece en el primer candidato. En un 22% de los casos la huella genuina no aparece en la lista de candidatos. Sin embargo, el funcionamiento del sistema es excelente a partir de 30 puntos seleccionados. Sin embargo, en huellas latentes en las que no sea posible tener más de 30 puntos el rendimiento del sistema puede degradarse sensiblemente.
- Se presentan resultados obtenidos a partir de la selección de puntos característicos de forma automática por el sistema. En ese caso, el rendimiento baja con respecto a la selección manual, pero ese descenso es mucho menos dramático que la reducción en el número de puntos, cosa que resulta muy sorprendente a priori. En casos en los que el sistema selecciona automáticamente los puntos característicos la degradación con el descenso de número de puntos es mucho más brusca que en el caso de selección manual. Sin embargo, en este último caso sorprende que a partir de 40 puntos característicos los resultados son excelentes, con un 100% de huellas objetivo en la primera posición de la lista.
- Se presentan resultados en los que se observa que la selección del área de la huella no tiene prácticamente ningún efecto en el rendimiento. Esto es sorprendente, ya que se esperaba que el sistema funcionara mejor si se eliminan los artefactos de la imagen de la huella (reglas, objetos, rótulos y escritura, etc.).

- Se presentan resultados en los que se observa que, en las pruebas realizadas, la selección de la tolerancia a la distorsión no tiene prácticamente ningún efecto en el rendimiento. Esto es sorprendente, ya que se esperaba que el sistema funcionara diferente si se tiene en cuenta distorsión que si no se tiene en cuenta. Sin embargo, el sistema parece ser suficientemente robusto a ese factor de variación.
- Se presentan resultados en los que se observa la variabilidad de los mismos en función del parámetro de medida de calidad dado por el sistema para la huella lanzada. En los experimentos realizados, se observa que la caída del rendimiento con bajas calidades medidas es muy significativa, sobre todo cuando el número de puntos característicos es alto y afecta al rendimiento del sistema en menor medida que la calidad. Esto motiva a que el uso de la calidad pueda ser indicativo de la fiabilidad que el especialista pueda darle al resultado obtenido.
- Se presentan resultados en forma de faunagramas, en los que se aprecia el desalineamiento de los scores para las distintas búsquedas. A raíz de ello se propone un método de normalización cuyo objetivo es alinear los scores y mejorar el poder de discriminación del conjunto. Se observa que tras el proceso de normalización, los scores de las distintas comparaciones aparecen alineados dentro de un mismo rango.
- Se presentan resultados en los que se mide el poder de discriminación del sistema para las bases de datos seleccionadas. Estos resultados se representan con curvas FA/FR, que muestran las distribuciones de scores, y con curvas DET, que miden el poder de discriminación. Se observa que la normalización tiene un efecto positivo en el poder de discriminación, acercando las curvas DET al origen de coordenadas, lo que supone una mejora en el EER.
- Se presentan resultados en los cuales se aprecia que varios de los métodos de cálculo de LR, seguidos de una fase de normalización de scores, presentan un rendimiento excelente. Sin embargo, estos resultados son optimistas, en el sentido de que la base de datos de scores utilizada para normalizar y calcular LR es muy parecida a los scores a transformar en LR (de hecho, son los mismos scores, pero para asegurar la generalidad de los resultados se utiliza un procedimiento de validación cruzada). En conclusión, los resultados son esperanzadores, pero es necesario confirmarlos con conjuntos experimentales más abundantes y ricos en variabilidad.

Conclusiones de los resultados en comparaciones entre huellas extraídas de reseñas decadactilares y la base de datos de reseñas decadactilares:

En general, los resultados obtenidos con huellas extraídas de reseñas decadactilares siguen una tendencia muy parecida a los obtenidos con huellas latentes. Se destacan aquí las diferencias entre ambos conjuntos de prueba:

- Se presentan resultados que muestran que, en general, los resultados con imágenes extraídas de reseñas decadactilares presentan un rendimiento mucho mayor que los obtenidos para huellas latentes. Aunque estos resultados son totalmente esperados, se observan varios efectos que han resultado muy sorprendentes:
 - El rendimiento del sistema está lejos de ser perfecto, incluso cuando se selecciona el máximo de puntos característicos en una huella (para impresiones posadas, sólo en un 94% de los casos la huella genuina aparece en el primer candidato y en un muy significativo 4% de los casos la huella genuina no aparece en la lista de candidatos).
 - El sistema sigue degradándose significativamente con la reducción del número de puntos característicos, siendo relativamente poco fiable cuando el número de puntos se reduce por debajo de 20.
 - El sistema sigue sufriendo los efectos de la pérdida de calidad en la huella a comparar, lo cual es muy significativo, porque las huellas extraídas de reseñas decadactilares se asumen como de alta calidad, y muchas veces no es así (tinta emborronada, defectos en las manos, etc.).
- Se observan efectos parecidos a los observados en comparaciones de huellas latentes en cuanto a efectos de la normalización de scores y también en cuanto a su influencia en el rendimiento del sistema.
- El cálculo de LR, se obtienen resultados parecidos a los de huellas latentes. Cabe destacar que el método PAV realizado tras la normalización de los scores es el que calibra mejor el conjunto de scores, siendo el más recomendado para este tipo de huellas.

Conclusiones Generales:

- Como se podía predecir, el número de puntos característicos es un parámetro que afecta en gran magnitud a los resultados obtenidos. El sistema funciona mejor cuanto mayor sea el número de puntos característicos utilizados en la comparación. Sin embargo, es necesario que los puntos estén bien situados en la imagen para que el funcionamiento del sistema sea correcto. Se puede comprobar que la selección automática de puntos en huellas anónimas no es muy eficaz, seguramente debido a que la calidad de las huellas no es buena, y que con la selección manual de puntos característicos se obtienen resultados notablemente mejores. Sin embargo también se observa que el número de puntos seleccionados en la huella es también crucial, ya que al reducir este parámetro a 12 puntos característicos manuales, la selección automática, que es muy errónea pero ofrece mayor número de puntos, funciona mejor que 12 puntos bien situados.
- En cuanto a los otros parámetros de variación utilizados en las búsquedas, podemos afirmar que la utilización del lazo de búsqueda y de la barra de tolerancia a la distorsión no afectan en nada a los resultados en búsquedas de huellas latentes. Con esto no queremos decir que el tamaño de la huella o la tolerancia del sistema a la distorsión no sean parámetros importantes en la identificación dactilar; únicamente parece que el sistema no tiene en cuenta la petición del usuario de variar dichos parámetros.
- El conjunto experimental analizado es de un tamaño muy pequeño como para poder sacar conclusiones estadísticamente fiables. Sin embargo, los resultados obtenidos tiene un gran valor de cara a dirigir las futuras investigaciones cuando el conjunto de resultados se amplíe.

8.2 Trabajo futuro

A partir de este trabajo se abren nuevas líneas de investigación. Las más interesantes se detallan a continuación:

- **Repetición de los experimentos con un nuevo y más rico conjunto experimental:** como ya se ha explicado, la fiabilidad estadística de los resultados obtenidos en este proyecto es pobre. Se propone la repetición de los experimentos con conjuntos de huellas mayores. Además, la limitación a 15 scores por cada lista de candidatos limita la fiabilidad en el cálculo de LR, ya que no puede modelarse de manera precisa la distribución de los scores *non-target*. Por lo tanto, se propone la repetición de los experimentos para el contraste con los presentados en este proyecto cuando las listas de candidatos sean mayores.
- **Medida de la robustez y confianza estadística:** aun disponiendo de un conjunto experimental pequeño, como del que se dispone en este proyecto, existen medidas para comprobar la robustez estadística de los resultados con el objetivo de medir la fiabilidad de los mismos. Se propone la utilización de métodos de medida de la confianza estadística con los resultados obtenidos en este proyecto.
- **Análisis de otros factores de variabilidad:** Como ya se ha explicado, el sistema dispone de más parámetros variables en las búsquedas distintos de los utilizados para este proyecto. Es interesante analizar la influencia en las comparaciones de cambios en parámetros como el margen de rotación, la utilización de la textura o la utilización de la fórmula (patrón general de la huella).
- **Uso de medidas de calidad:** se propone el uso de medidas de calidad para asistir al especialista en la predicción del rendimiento del sistema. La calidad de las huellas, tal y como las define el sistema, es un buen predictor del rendimiento del sistema sólo en el caso en el que se disponga de muchos puntos característicos en la huella latente independientemente de cómo se seleccionen. Existen otras muchas formas de medir calidad en huellas latentes aparte de la implementada por el sistema, y la combinación de las mismas puede dar lugar a una muy buena estimación del rendimiento del sistema.
- **Trabajo con huellas palmares:** el sistema utilizado ofrece también funcionalidad para trabajar con huellas palmares. Las bases de datos forenses para este tipo de huellas son todavía muy reducidas, siendo un campo no tan estudiado como el de las huellas dactilares, lo cual es una motivación adicional para su estudio.

Bibliografía

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. Fine, G.E. *A Review of the FBI's Handling of the Brandon Mayfield Case*. U.S. Department of Justice, Office of the Inspector General, 2006.
7. U.S. Supreme Court. *Daubert vs. Merrel Dow Pharmaceuticals*. Vol. [509 US. 579], 1993.
8. Saks, M.J. and Koehler, J.J. "The Coming Paradigm Shift in Forensic Identification Science". *Science* Vol. 309, pages 892-895, 2005.
9. Ramos, D. *Forensic evaluation of the evidence using automatic speaker recognition systems*. PhD. Thesis. UAM, Madrid, 2007.
10. Aitken, C.G.G. and Taroni, F. *Statistics and the Evaluation of Evidence for Forensic Science*. John Wiley & Sons, Chichester , 2004.
11. Balding, J. *Weight-of-Evidence for Forensic DNA Profiles*. Wiley, 2005.
12. Jain, A.K., Flynn, P. and Ross, A.A. editors. *Handbook of biometrics*. Springer, 2007.
13. Champod, C. "Forensic Applications, Overview". *Encyclopedia of Biometrics*. S.Z. Li, editor. Springer, 2009
14. Champod, C. "Keynote Speech". *International Conference of the European Academy of Forensic Science (EAFS)*, 2006.
15. Cook, R., Evett, I.W., Jones, P.J., Jackson, G. and Lambert, J.A. "A hierarchy of propositions: deciding which level to adress in casework". *Science and Justice*, pages 231-239. Vol. 38, 1998
16. Champod, C., Lennard, C., Margot, P., and Stoilovic, M. *Fingerprints and Other Ridge Skin Impressions*. CRC Press, 2004.
17. Margot, P. and German, E. "Fingerprint Identification Breakout Meeting". *Proc. International Symposium on Fingerprint Detection and Identification*. Israel National Police, pages 26-30, Israel, 1995.
18. Ratha, N. and Bolle, R. *Automatic Fingerprint Recognition Systems*. Springer, 2003.
19. 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

20. 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
21. Capelli, R., Maio, D., and Maltoni D. "Indexing Fingerprint databases for efficient 1:N matching". *Proc. Int. Conf. on Control Automation Robotics and Vision (6th)*, 2000.
22. Martin, A., Doddington, G., Kamm, T., Ordowski, M. and Przybocki, M. "The DET curve in assesment of detection task performance". *Proc. of Eurospeech*, pages 1895-1898, 1997
23. Asbaugh, D.R. *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. The CRC Press, Boca Raton, FL, 1999.
24. Auckenthaler, R., Carey, M. and Lloyd-Thomas, H. "Score normalization for text-independent speaker verification systems". *Digital Signal Processing*. pages 10:42-54, 2000.
25. Bimbot, F., Bonastre, J.F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska-Delacretaz, D. and Reynolds, D.A. "A tutorial on text-independent speaker verification". *Journal on Applied Signal Processing*, pages 430-451. Vol. 2004(4), 2004.
26. Scheck, B., Neufeld, P. and Dwyer, J. *Actual Innocence*. Doubleday, pages 52-53. NewYork, 2000.
27. Champod, C. and Meuwly, D. "The inference of identity in forensic speaker recognition". *Speech Communication*, pages 193-203. Vol. 31., 2000.
28. Bernardo, J.M. and Smith, M.F.A. *Bayesian Theory*. Wiley and Sons, Chichester, 1994.
29. Meuwly, D. "Forensic Evaluation from Biometric Data". *Science & Justice*. pages 205-213, Laussane, 2006,
30. Shannon, C.E. "A mathematical theory of communication". *Bell Sys. Tech. Journal*, 1948.
31. Brümmer, N. and Du Preez, J. "Application independent evaluation of speaker detection". *Computer Speech and Language*, 2006.
32. Cohen, I. and Goldsmith, M. "Properties and benefits of calibrated classifiers". In *Proc. of European Conference of Machine Learning ECML/PKDD*, 2004.
33. Brümmer, N., Burget, L., Cernocky, J., Glembek, O., Grezl, F., Karafiat, M., van Leeuwen, D.A., Matejka, P., Schwartz, P. and Strasheim, A. "Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006". *IEEE Transactions on Audio, Speech and Signal Processing*, pages 2072-2084. Vol. 15., 2007.
34. Gonzalez-Rodriguez, J., Fierrez-Aguilar, J., Ramos-Castro, D. and Ortega-Garcia, J. *Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems*. 2-3, Vol. 155, pages 126-140. December, 2005.
35. Alexander, A. *Forensic automatic speaker recognition using Bayesian interpretation and statistical compensation for mismatched conditions*. École Polytechnique Fédérale de Lausanne, 2005.
36. Meuwly, D. *Reconnaissance de Locuteurs en Sciences Forensiques: L'apport d'une aproche automatique*. PhD. Thesis, IPSC-Universite de Laussane, 2001.
37. Botti, F., Alexander, A. and Drygajlo, A. "An interpretation framework for the evaluation of evidence in automatic speaker recognition with limited suspect data". *Proc. of Oddisey*, pages 63-68, 2004.

38. Ramos, D., Gonzalez-Rodriguez, J., Champod, C., Fierrez-Aguilar, J. and Ortega-Garcia, J. "Between source modelling for likelihood ratio computation in forensic biometric recognition". *Proc. of IAPR Audio and Video-based Biometric Person Authentication 2005*, pages. 1080-1089. Vols. Springer LNCS-3546, 2005.
39. Alonso-Fernandez, F. *Biometric Sample Quality and its Application to Multimodal Authentication Systems*. PhD. Thesis. UPM, Madrid, 2008.
40. Pigeon, S., Druyts, P. and Verlinde, P. "Applying logistic regression to the fusion of the NIST'09 1-speaker submissions". *Digital Signal Processing*. pages 10(2):207-248, 2000.
41. Gonzalez-Rodriguez, J., Rose, P., Ramos, D., Toledano, D.T. and Ortega-Garcia, J. "Emulating DNA: Rigorous Quantification of Evidential Weight in Transparent and Testable Forensic Speaker Recognition". *Speech and Language Processing. IEEE Transactions on Audio*, Vol. 15, pages 2104-2115, 2007.
42. Egli, N.M., Champod, C. and Margot, P. "Exploring evaluation in fingerprint comparison and automated fingerprint identification systems". *Forensic Science International*, pages 189-195. Vol. 167, 2007.
43. de Jongh, A., Rodriguez, C. and Meuwly, D. *Exploring logical and empirical aspects of a likelihood ratio evidence system developed for forensic biometrics*. The Hague. Netherlands Forensic Institute.

A

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**
 - 640 horas a 15 € / hora 9600 €

- 5) **Material fungible**
 - Gastos de impresión 60 €
 - Encuadernación 200 €

- 6) **Subtotal del presupuesto**
 - Subtotal Presupuesto 12060 €

- 7) **I.V.A. aplicable**
 - 16% Subtotal Presupuesto 1929.6 €

- 8) **Total presupuesto**
 - Total Presupuesto..... 13989,6 €

Madrid, Febrero de 2010

El Ingeniero Jefe de Proyecto

Fdo.: María Puertas Calvo
Ingeniero Superior de Telecomunicación

B

**Pliego de
Condiciones**

Este documento contiene las condiciones legales que guiarán la realización de este proyecto. 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.

C

**Paper enviado a
ICPR 2010**

Towards a Better Understanding of the Performance of Latent Fingerprint Recognition in Realistic Forensic Conditions

Maria Puertas¹, Daniel Ramos¹, Julian Fierrez¹,

Javier Ortega-Garcia¹ and Nicomedes Exposito-Marquez²

¹ATVS - Biometric Recognition Group, Universidad Autónoma de Madrid, Spain

²Departamento de Identificación. Servicio de Criminalística de la Guardia Civil, Ministerio del Interior, Spain.

{maria.puertas, daniel.ramos}@uam.es

Abstract

This work studies the performance of a state-of-the-art fingerprint recognition technology, in several practical scenarios of interest in forensic casework. First, the differences in performance between manual and automatic minutiae extraction for latent fingerprints are presented. Then, automatic minutiae extraction is analyzed using three different types of fingerprints: latent, rolled and plain. The experiments are carried out using a database of latent fingermarks and fingerprint impressions from real forensic cases. The results show high performance degradation in automatic minutiae extraction compared to manual extraction by human experts. Moreover, high degradation in performance on latent fingermarks can be observed in comparison to fingerprint impressions.

1. Introduction

In forensics, fingerprint images can be classified in two categories: fingerprint impressions and latent fingermarks. Fingerprint impressions are obtained either by scanning the inked impressions on paper or by using scanning devices [1]. The acquisition of this kind of prints is controlled by an expert to make sure that the images have good quality. Impressions are also divided in two types: rolled and plain. Rolled impressions are obtained by rolling the finger from one side to the other. Plain impressions are those in which the finger is pressed down but not rolled. Fingerprint impressions are usually contained in ten-print cards. A ten-print card contains the rolled and plain impressions of the ten fingers of a person. The reason to include both types of impressions is that

while rolled fingerprints contain larger size and a higher number of minutiae, plain fingerprints are less distorted and often have clearer ridges [1]. The other group of fingerprints is latent fingermarks. These marks are accidentally left in a crime scene and then recovered and scanned [2].

Nowadays, state-of-the-art fingerprint recognition systems for good quality fingerprint images have an acceptable level of performance [3] [4] [5]. However, when dealing with latent fingermarks, there is still a lot of research to be done. Low quality, incompleteness and distortion are typical problems that forensic Automated Fingerprint Recognition Systems (AFIS) have to face when extracting features from latent fingermarks. As many of the available AFIS systems are mainly minutia-based, errors in minutiae extraction play a critical role in performance. Unfortunately, such kind of errors are frequent with latent fingerprints, which almost always contain less clarity, less content and less undistorted information than fingerprints taken under controlled conditions [6]. Figure 1 shows an example of the three different types of fingerprints.

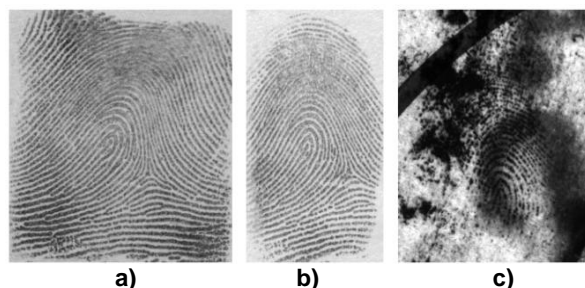


Figure 1. Three types of fingerprint images: a) Rolled fingerprint, b) Plain fingerprint, c) Latent fingerprint. (Extracted from [1])

Automated Fingerprint Identification Systems (AFIS) allow the search of the fingermarks among millions of ten-print cards usually producing a ranked list of top candidates based on similarity scores [45].

The matcher calculates these *scores* mainly based on the comparison of minutiae features and their spatial relationship [8].

As a consequence of the previously discussed quality problems in latent fingerprints, the common protocol in forensics is to manually mark the minutiae before launching a search. When the search finishes, the list of top candidates is usually manually reviewed by experts in search of a match [9].

Given the importance of the minutiae extraction process in latent fingerprints in forensic applications, in this work we compare and analyze the performance of automatic and manual latent feature extraction in forensic fingerprint recognition. Moreover, we compare and analyze the performance of automatic minutiae extraction in three types of fingerprints: latent, rolled and plain.

The selection of a proper realistic experimental set-up is critical in forensics. Therefore, we have used as query a set of 50 latent fingerprints from real, solved cases of the Identification Department of Spanish Guardia Civil. Also, the 100 impressions used as queries in the case of rolled and plain fingerprints have been taken from ten-print cards stored in the Guardia Civil database. Finally, the experiments reflect a realistic scenario, because all the queries are compared to the full database of ten-print cards in Guardia Civil database (around 2.5 million cards).

The rest of the paper is organized as follows: Section 2 presents the experimental framework used in this work. Section 3 reports and discusses the experimental results, and conclusions are finally drawn in Section 4.

2. Experimental framework

2.1 Fingerprint recognition system

A state-of-the-art fingerprint recognition system has been used (one of the top ranked in NIST FpVTE 2003 [3], and subsequent evaluation campaigns like NIST PFT¹ [10]). This system works in identification mode [8]. For each input fingerprint, the system compares it to a ten-print card database. The output consists of a list of the 15 impressions of the ten-print card database that achieve the highest matching *scores*.

2.2 Fingerprint databases

All the fingerprints used in this work belong to the Spanish Guardia Civil database. For the experiments, two different sets of fingerprints have been used. The first set consists of 50 latent fingerprints from real forensic cases. All the fingerprints in this set

have been previously manually identified with their sources by human fingerprint experts of Guardia Civil, and therefore the identity of each mark will be assumed to be known in our experiments. The second set is composed by 100 right index impressions from 50 convicted individuals. For each individual there are 2 impressions: one rolled fingerprint and one plain fingerprint.

Both sets of prints are compared to a database that contains about 2.5 million ten-print cards. For each fingerprint or fingerprint used as a query, it is guaranteed that at least one ten-print card from the individual at hand is available in the searched database.

The small size of the sets of fingerprints used for the experiments is due to the limitations in obtaining and handling fingerprint data from real cases. This limitation reduces the statistical significance of the results obtained in the experiments. However, the sample size is big enough to observe trends in the performance with different minutiae extraction techniques and types of images.

2.3 Experiments

In this work, two types of experiments have been carried out. In the first type, the matcher compares the set of latent fingerprints to the ten-print card database. In the second scenario, the set of fingerprint impressions is compared to the ten-print card database.

Four different tests have been carried out with the set of latent fingerprints:

Latent fingerprints with all the minutiae extracted automatically by the system.

A human fingerprint expert manually marks all the minutiae that he or she can find in the fingerprints.

The human expert selects the best 12 minutiae for each fingerprint according to his experience and only those are used for the matching.

The best 8 minutiae are manually selected in every fingerprint among the 12 in the previous experiment, before sending the fingerprint to the matcher.

In the case of the fingerprint impressions, we have run the experiments separating the fingerprints in two groups: rolled and plain. For both sets of impressions, only the automatic feature extraction has been used, following the forensic protocol used in casework for this kind of prints.

In all the experiments of this work, the minutiae in the ten-print card database are automatically extracted by the system.

¹ Recently updated in Dec. 2009 (<http://fingerprint.nist.gov/PFT/>)

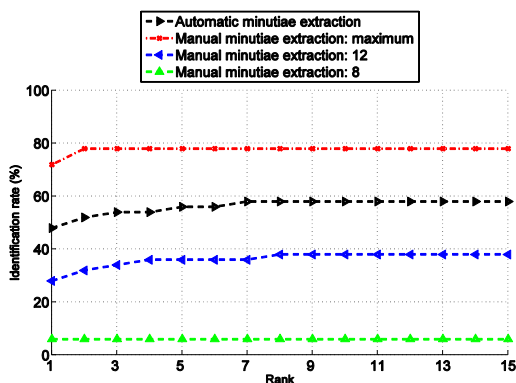


Figure 2. CMC curves for the latent queries.

3. Results

In this section, experimental results are shown using CMC (Cumulative Match Characteristic) curves, which are often used to show the matching performance of identification systems [11]. A CMC curve plots the rank- n identification rate against n , for $n=1, 2, \dots, M$, where $M=15$ in our case. The rank- n identification rate indicates the proportion of times the genuine identity appears among the top n matches.

3.1 Manual vs. automatic minutiae extraction in latent fingerprints

Results for latent fingermarks as queries against the ten-print database are shown in figure 2. It can be observed that manual minutiae selection leads to much better results than automatic selection, as it was expected. However, when the manual selection is limited to 12 minutiae, the performance dramatically drops. It is shown that automatic extraction provides 48% of first position genuine candidates, when the manually selected 12 minutiae only rates 28% first position genuine candidates. It can be also observed that matching with only 8 manually-selected minutiae shows to have the worst performance, as it was expected. We found that only 6% of the genuine fingerprints rated first on the lists of candidates and 94% of them never appeared in the lists of candidates.

These results using latent queries against the ten-print database are summarized in Table 1.

Table 1. First candidate vs. not-appearance rates for latent queries.

	Max. manual	Max. automatic	12 manual	8 manual
Rank 1	72%	48%	28%	6%
Rank>15	22%	42%	62%	94%

It is important to remark that the average number of minutiae that the system marked in the latent prints was 31.2 while the average number of minutiae marked by the experts was 25.2. This shows that due to the poor quality of latent images, the system tends to mark minutiae in points of the image where actually they are not, leading to a worse performance in the matcher.

3.2 Automatic minutiae extraction with latent, rolled and plain fingerprints

In this section, results using impressions compared to the ten-print database are shown, and then compared to latent results when minutiae are automatically extracted.

As it can be observed in figure 3, matching with rolled impressions leads to 100% identification rate for rank 1, which means that all the genuine fingerprints appeared as the first candidate in all the experiments. For the plain impressions, the results are slightly worse, as in 94% of the cases, the *target* print appeared as the first candidate. However, in 4% of the experiments with plain impressions, the genuine fingerprint did not appear in the list of candidates. The decrease in performance of plain with respect to rolled impressions is mainly due to the fact that rolled impressions tend to present a much higher number of minutiae, also presenting a much wider fingerprint area. Worth noting, while the average number of minutiae that the system found in rolled impressions is 83.8, in plain impressions it is only 44.16. Figure 3 also shows the results using latent queries with automatic minutiae selection. It is shown that the performance is much worse for latent fingerprints than for the case of plain or rolled impressions.

In table 2, the performance of the three types of matching experiments is shown in terms of percentage of first-rank appearance and no appearance of *target* identity.

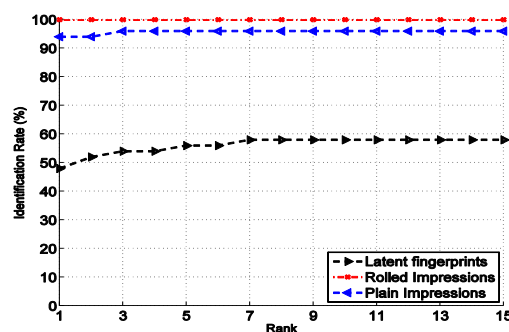


Figure 3. CMC curves for rolled, plain and latent queries.

Table 2. First candidate vs. not-appearance rates for genuine impressions and latent fingerprints.

	Rolled	Plain	Latent
Rank 1	100%	94%	48%
Rank > 15	0%	4%	42%

The big difference in performance among latent fingermarks and fingerprint impressions can be attributed not only to a decrease in the image quality, but also to the average number of minutiae for the latent set, which was 31.2, much lower than in the case of impressions (i.e. 83.8 for rolled and 44.16 for plain).

4. Conclusions and future work

This work presented a comparative study of the performance of biometric recognition with latent fingermarks and rolled and plain fingerprint impressions from real forensic cases. A database from the Identification Department of Spanish Guardia Civil has been used for that purpose. We presented and discussed the results of two types of experiments regarding fingerprint feature extraction and matching. The aim of the first experiment was to compare manual and automatic feature extraction in latent fingerprints. The second experiment consisted of comparing automatic feature extraction for latent marks and printed and plain impressions.

In the first scenario, when comparing manual and automatic minutiae extraction in latent fingerprints, it is clear that the identification accuracy increases when the feature selection is manually made by an expert, which was an expected result. However, it is significant that automatic minutiae selection works much better than manually selecting a limited amount of minutiae (12 in our experiments).

In the second scenario, we have compared the matching performance for automatic feature extraction in latent prints, rolled and plain impressions. In this case, it has been shown that automatic minutiae extraction leads to a high performance on the matching step when dealing with good quality fingerprints, such as rolled and plain impressions. However, the performance in latent fingerprints is much worse due to the wrong minutiae extraction made by the system.

Although the experimental scenario considered is very realistic (state-of-the-art AFIS and the ten-print database of Guardia Civil comprising around 2.5 million templates), it is important to remark that these results are obtained with a small set of fingerprint queries. This is mainly due to the limitations in obtaining and handling data from real forensic cases. For this reason, the statistical significance of the results must be taken with care. However, results are relevant in order to observe

trends in the performance with different minutiae extraction techniques and types of images, and pioneering in the sense of scientific understanding of this important technology for forensics.

Future work includes measuring the quality of latent marks in order to better understand its influence in performance [12]. Also, the extension of the sets of queries is planned in order to have more significant results.

Acknowledgements

This work has been partially supported by: projects Bio-Challenge (TEC2009-11186) and BBfor2 (FP7 ITN-2009-238803), "Dirección General de la Guardia Civil", and "Cátedra UAM-Telefónica". Author J. F. is supported by a Marie Curie Fellowship from the European Commission.

References

- [1] J. Feng, S.Yoon and A.K. Jain. "Latent Fingerprint Matching: Fusion of Rolled and Plain Fingerprints", *Proc. ICB 2009*, Springer LNCS-5558, pp. 695-704, 2009.
- [2] C. Champod, C. Lennard et al. *Fingerprints and Other Ridge Skin Impressions*. CRC Press, 2004.
- [3] C. Wilson, R. Austin Hicklin et al. "Fingerprint Vendor Technology Evaluation 2003", *Technical Report NISTIR 7123*, 2004.
- [4] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia and A. K. Jain. "Incorporating image quality in multi-algorithm fingerprint verification", *Proc. IAPR ICB*. Springer LNCS-3832, pp 213-220, January 2006.
- [5] H. Fronthaler, K. Kollreider, et al. "Fingerprint Image Quality Estimation and its Application to Multi-Algorithm Verification", *IEEE Trans. on Information Forensics and Security*, Vol. 3, n. 2, pp. 331-338, June 2008.
- [6] D.R. Asbaugh. *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. The CRC Press, Boca Raton, FL, 1999.
- [7] N.M. Egli, C. Champod, and P. Margot. "Evidence evaluation in fingerprint comparison and automated fingerprint identification systems: Modelling within finger variability", *Forensic Science Intl*, 167:189-195, 2007.
- [8] N. Ratha and R. Bolle. *Automatic Fingerprint Recognition Systems*. Springer, 2003.
- [9] D. Dessimoz and C. Champod. "Linkages Between Biometrics and Forensic Science", *Handbook of biometrics*. (ed.) A.K. Jain at al. Springer, 2007.
- [10] C. Watson et al. "Two Finger Matching With Vendor SDK Matchers", *Technical Report NISTIR 7249*, 2005.
- [11] R.M. Bolle, J.H. Connell et al. "The relation between the ROC curve and the CMC", *Proc. IEEE AutoID*, 2005.
- [12] F. Alonso-Fernandez, J. Fierrez et al. "A comparative study of fingerprint image-quality estimation methods", *IEEE Trans. on Information Forensics and Security*, Vol. 2, n. 4, pp. 734-743, December 2007.

