UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

DEPARTAMENTO DE TECNOLOGÍA ELECTRÓNICA Y DE LAS COMUNICACIONES

Reconocimiento Automático de Firma Manuscrita Asistida por Humanos

-- TESIS DOCTORAL --

Autor: Derlin Morocho Checa
Ingeniero en Electrónica y Telecomunicaciones,
Universidad de las Fuerzas Armadas – ESPE
Sangolquí, Ecuador

Tesis presentada para el grado de
Doctor en Filosofía.

Madrid, Mayo 2019

Departamento:	Tecnología Electrónica y de las Comunicaciones Escuela Politécnica Superior Universidad Autónoma de Madrid (UAM), ESPAÑA
Tesis PhD:	Reconocimiento Automático de Firma Manuscrita Asistida por Humanos
Autor:	Derlin Morocho Checa Ingeniero en Electrónica y Telecomunicaciones, Universidad de las Fuerzas Armadas – ESPE Sangolquí, Ecuador
Director:	Aythami Morales Moreno Doctor Ingeniero de Telecomunicación Universidad Politécnica de Madrid Universidad Autónoma de Madrid, España
Director:	Julián Fiérrez Aguilar Doctor Ingeniero de Telecomunicación Universidad Politécnica de Madrid Universidad Autónoma de Madrid, España
Año:	2019
Comité:	Presidente: Javier Ortega García Universidad Autónoma de Madrid, España Secretari@: Miguel Ángel Ferrer Ballester Universidad de Las Palmas de Gran Canaria Vocal 1: Matilde Santos Peña Universidad Complutense de Madrid

BiDA Lab

La investigación descrita en esta Tesis se llevó a cabo dentro del Grupo de Reconocimiento Biométrico - BiDA (Biometrics and Data Pattern Analytics) en el Departamento de Tecnología Eléctrica y de las Comunicaciones, Escuela Politécnica Superior, Universidad Autónoma de Madrid.

RESUMEN

Esta Tesis Doctoral explora cómo la intervención humana puede ayudar a mejorar los sistemas de autenticación automática de firma. Durante las últimas décadas, se han hecho esfuerzos importantes para mejorar el rendimiento de los sistemas automáticos de reconocimiento basados en firma. Este trabajo analiza como acciones llevadas a cabo por humanos pueden ser utilizadas para complementar las capacidades de los algoritmos automáticos. Qué acciones y en qué medidas éstas pueden ayudar a mejorar algoritmos del estado-del-arte es el objetivo perseguido en esta tesis. En este trabajo se proponen y analizan intervenciones a nivel de clasificación y de extracción de características. El análisis a nivel de clasificación comprende experimentos con más de 500 personas a través de tareas de autenticación de firma desarrolladas sobre plataformas de *crodwourcing*. Los resultados permiten establecer un rendimiento de base de reconocimiento de firma llevado a cabo por personas sin experiencia forense. La intervención a nivel de características se realiza a través de aplicaciones de etiquetación de atributos de firma desarrolladas durante la tesis e inspiradas en el trabajo de expertos forenses. Estas aplicaciones han permitido la captura de dos nuevas bases de datos. Se analiza el rendimiento humano para la etiquetación de atributos discriminantes de la firma y se analiza su complementariedad con sistemas automáticos. Los experimentos se llevan a cabo sobre una la base de datos pública BioscurID y se incluyen los dos escenarios de autenticación más populares: firma dinámica u online y firma estática u offline. A su vez, se analizan los dos tipos de ataques más habituales: aleatorios y ataques entrenados. Los resultados demuestran las limitaciones y el potencial de la intervención humana en tareas de reconocimiento de firma. Se demuestra que las capacidades humanas pueden ser explotadas para mejorar el rendimiento de los sistemas de autenticación automática de firma.

Algunas de las principales preguntas a responder durante el desarrollo e esta tesis, han sido:

1. ***¿Qué y Cómo pueden influir la asistencia del humano en los sistemas de reconocimiento automático de firma?***
2. ***¿Qué medidas o acciones tomar para ayudar en el estado del arte de los sistemas de reconocimiento automático de firma?***

Esta tesis pretende analizar la problemática a través de acciones a realizar para mejorar la validación de firmas, a través de:

1. ***Reconocimiento de firmas a través de etiquetación de atributos inspirados en el análisis de FDE, realizados por humanos.***

2. *Reconocimiento de firmas a través de Tareas de Inteligencia Humana (HIT), vía crowdsourcing en la plataforma de Amazon Mechanical Turk (MTurk).*

Para la ejecutar las acciones presentadas en esta tesis, se plantean las siguientes hipótesis (H) de investigación:

H.1. ¿Qué tan bueno es una persona sin experiencia en FDE (profano) al reconocer la autenticidad de una firma de consulta?
H.2. ¿Cuál es la consistencia de los atributos anotados por diferentes usuarios?
H.3. ¿Cuáles son los atributos más discriminantes?
H.4. ¿Qué tan escalable es el sistema cuando los experimentos involucran un conjunto de datos más grande?
H.5. ¿Cuál es el rendimiento en comparación con los sistemas de reconocimiento de firma off-line?
H.6. ¿Cuál es el rendimiento de los atributos de firma anotados manualmente?
H.7. ¿Cuál es la complementariedad (en términos de rendimiento) entre la autenticación por humanos basada en atributos y la autenticación automática de firma on-line tradicional?
H.8. ¿Cuáles son las características más empleadas por los humanos?
H.9. ¿Existe alguna relación entre las características empleadas y el rendimiento?
H.10. ¿Cuál es la diferencia entre el rendimiento de diferentes humanos?
H.11. ¿Es el humano un buen reconocedor de firmas?
H.12. ¿Cuál es la consistencia de los atributos anotados aplicados a otra base de datos?
H.13. ¿Pueden los atributos comparativos mejorar el reconocimiento automático en esquemas combinados?
H.14. ¿Cuál es la estabilidad de los anotadores para diferentes números de firmas?
H.15. ¿Es el humano estable en el proceso de etiquetado de firmas?
H.16. ¿Qué parámetros ayudan a mejorar el reconocimiento de firma en el proceso de etiquetado?

Estas hipótesis son resueltas a través de experimentos relacionados con las capacidades humanas para el reconocimiento de firma, y sus resultados son presentados a través de artículos científicos publicados en congresos y revistas de prestigio.

Las principales contribuciones de este trabajo se pueden resumir a continuación:

- Establecer y generar una línea de base del rendimiento humano.
- Generación de dos nuevas bases de datos compuesta de etiquetados de atributos de la firma realizados por los humanos (Base de datos Bio-HSL [Biometric Handwritten Signatures Labeling], y DB_Labeling_ESPE).

- Obtener y analizar el rendimiento del humano en reconocimiento de firma y sus capacidades, a través de etiquetación de atributos absolutos y comparativos de la firma inspirados en FDEs, y experimentos masivos vía crowdsourcing.
- Se ha verificado que el rendimiento humano en reconocimiento de firma por si solo no genera buenos resultados, pero al combinar sus respuestas tiene una gran mejora [FRR mejora (↓80%), de 31.4% a 7%; FAR mejora (↓40%) de 37.6% a 23.8%], en experimentos de crowdsourcing.
- El rendimiento del humano mejora con la calidad de información e interfaces interactivas en tareas de reconocimiento de firma, mediante la etiquetación de atributos característicos de la firma.
- El rendimiento en términos de EER aleatoria y simulada, permite determinar el potencial de los atributos comparativos. Los resultados muestran que el 38% de los profanos presentan un EER aleatorio por debajo del 5%, donde el mejor EER es 3.90% y el peor EER es 10.32%. Para las falsificaciones simuladas, tenemos que el 52% de los profanos tienen un EER simulado de menos del 21% con el mejor EER del 18.83% y el peor EER del 26.22%.
- El rendimiento del humano en un sistema semi-automático, en términos de EER a nivel de atributos absolutos tiene un $EER_{Simulado}$ del 24.22% y un $EER_{Aleatorio}$ del 6.89%; y a nivel de atributos comparativos tiene un $EER_{Simulado}$ del 21.20% y un $EER_{Aleatorio}$ del 5.57%, obteniendo con atributos comparativos una mejora del 13% para el $EER_{Simulado}$, y del 20% para el $EER_{Aleatorio}$.

DEDICATORIA

Esta tesis está dedicada a mi hermosa familia, Mechita, Jonithan, Ayita, Darlita, como los amo, nunca dudaron de este momento. Solo ustedes saben el sacrifico que hemos pasado todos estos años. Esto es por ustedes y para ustedes, mis amores.

A la memoria de mi Santa Madre “Rosita Elvira”, orgullosa de lo alcanzado por tu hijo, a mi Viejito Vicente, mi señor Padre siempre fuerte como un roble.

Y, a todas las personas que quieren salir adelante para superar sus propios límites.

Derlin Morocho Checa
Quito, Enero del 2019

AGRADECIMIENTOS

En primer lugar tengo que dar gracias a Dios, que con su sabiduría y sus bendiciones han iluminado y guiado para poder culminar con éxito mis estudios de doctorado.

Seguidamente mis más grandes agradecimientos a mi hermosa familia, **Mechita** mi esposa, siempre apoyándome con sus fuerzas de aliento y bendiciones para seguir siempre adelante, a mis hijitos, **Jonithan, Ayita, y Darlita** que con su cariño, amor y responsabilidad me han dado las fuerzas para tener en cuenta el sacrificio que estamos haciendo, como los amo, son mi fortaleza, me siento bendecido con mi linda familia.

Gracias Viejit@s, "Vicente", mi Padre que con tu amor, tus bendiciones y enseñanzas estoy donde estoy, a mi Madre "Rosita Elvira" que en paz descanse, que con tus bendiciones desde el cielo siempre me acompañas guiando mis pasos.

Gracias a los profesores del grupo BiDA, por darme la oportunidad de poder realizar mi tesis doctoral que con su granito de arena y motivación para llegar a culminar con éxito este momento. Un agradecimiento muy especial a mi tutor Julián por la oportunidad y la confianza para poder trabajar y desarrollar mi tesis, que con tus ideas muy innovadoras me has encaminado en cumplir este objetivo. Gracias Julián.

Un apartado muy especial de agradecimiento para ***Aythami***, más que mi co-director de tesis, un ***amigo,*** que siempre me apoyó y valoró mi trabajo y que sin su ayuda incondicional no podría alcanzar con éxito este logro. ***Aythami***, muchas gracias por toda tu ayuda y tus enseñanzas.

Derlin Morocho Checa
Quito, Enero del 2019

INDICE

LISTA DE FIGURAS

Capítulo 1

Capítulo 2

Capítulo 3

Capítulo 4

LISTA DE TABLAS

Capítulo 1

Capítulo 2

Capítulo 3

Capítulo 4

Capítulo 1

1.- Introducción

La gestión de la identidad se ha convertido en una pieza fundamental en multitud de aplicaciones en tiempo real, tales como aplicaciones forenses, migración de personas, transacciones financieras y la seguridad informática.

La identidad se comprueba tradicionalmente a través de documentos públicos generados para tal fin como es el Documento Nacional de Identidad o el Pasaporte. Otros métodos de identificación incluyen la posesión de "tokens", que se corresponden con objetos que sólo posee el sujeto a identificar (e.j. tarjetas RFID, Smartphone,…). Además los sistemas de identificación informáticos se basan en el conocimiento de contraseñas o PINs (Jain A. K. et al, 2016).

La creciente demanda de acceso a los servicios de la Sociedad de la Información, ha dado lugar en las últimas décadas a la aparición de una nueva rama de la tecnología de gestión de identidades denominada Autenticación Biométrica o simplemente Biometría.

La biometría se encarga de la autenticación de la identidad de una persona basada en rasgos biométricos asociados a sus características biológicas y de comportamiento (Jain A. K. et al, 2004). La identidad de la persona es la base sobre la que se sustentan tanto sus derechos, como sus obligaciones, de ahí radica la importancia de su validación. Los rasgos biométricos que se analizan pueden ser rasgos físicos intrínsecos o rasgos conductuales (Jain A. K., et al, 2016). La biometría excede la seguridad a las contraseñas y tarjetas, permitiendo una mayor confiabilidad en proteger su información e identidad.

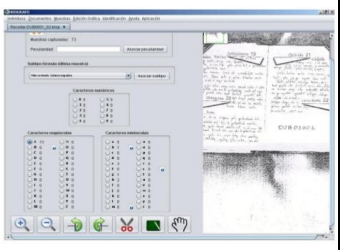
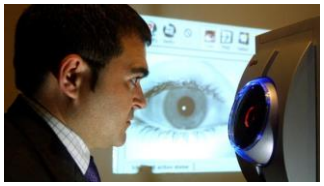
1.1.- Rasgos Biométricos

El reconocimiento biométrico, o simplemente biometría, se refiere al reconocimiento automático de los individuos en función de sus características físicas y/o conductuales (Jain A. K. et al, 2011). La Tabla 1.1, muestra una clasificación de las tecnologías de reconocimiento biométrico más populares. Entre los ejemplos de características biométricas que se han utilizado con éxito en aplicaciones prácticas se incluyen la cara, la huella dactilar, la palma de la mano, el iris, la voz, la dinámica de tecleo, o la firma manuscrita.

Los rasgos biométricos suelen ofrecer prestaciones más elevadas en términos de tasas de reconocimiento aunque su obtención tradicionalmente se percibe como más invasiva. Los rasgos de comportamiento tienden a variar en el tiempo, pero su obtención transparente no requiere de procesos de autenticación como tal

y hace que estos sistemas sean idóneos en determinadas aplicaciones como la autenticación en el dominio digital (Jain A. K. et al, 2011).

Tabla 1.1: Tecnologías de Reconocimiento Biométrico

Características Fisiológicas		**Características De Comportamiento**	
Rasgos Biométricos	**Tecnología**	**Rasgos Biométricos**	**Tecnología**
Huella Dactilar		Firma	
Geometría de la Mano		Escritura	
Rostro / Cara		Voz	
Iris / Retina		Dinámica de Teclado	

Debido a que los rasgos biométricos son generalmente inherentes a un individuo, existe un fuerte y razonable vínculo permanente entre una persona y sus características biométricas. Las tecnologías de reconocimiento biométrico han dado grandes pasos en los últimos 20 años consiguiendo tasas de reconocimiento muy elevadas en gran cantidad de rasgos biométricos (e.j. huella dactilar, cara, iris, firma, etc…) y diferentes aplicaciones (e.j. seguridad, ocio, educación, etc…) En muchos aspectos, las soluciones biométricas superan a las tecnologías tradicionales basadas en contraseñas. Sin embargo, las tecnologías de reconocimiento biométrico no están libres de errores y vulnerabilidades. Estos sistemas son vulnerables a ataques de suplantación de identidad, ataques de vinculación etc., (Gomez-Barrero M., et al, 2017).

Adicionalmente, los sistemas reconocimiento biométrico presentan errores que pueden inducir a tomar decisiones incorrectas. El proceso de adquisición y la variabilidad propia de las muestras introducen una incertidumbre que pueden conducir a dos tipos de errores: Falsa Aceptación (FA) y/o Falso Rechazos FR). La FA puede conducir a la propagación de la identidad, donde el individuo después de repetidos intentos, se las arregla para asumir la identidad de un usuario legítimo del sistema (por ejemplo, imágenes de caras en los sitios de medios sociales, imágenes de caras de gemelos idénticos). Dadas las limitaciones anteriores, un mecanismo de autenticación multi-factor que combina datos biométricos con contraseñas y/o símbolos puede ser una mejor aproximación a la seguridad en muchas aplicaciones (Jain A. K., et al, 2016).

Un problema crítico en el diseño de sistemas biométricos es la elección del rasgo biométrico. En teoría, cualquier característica anatómica, de comportamiento o fisiológica de un individuo se puede usar como un rasgo biométrico. Sin embargo, la elección de un rasgo biométrico para una aplicación particular generalmente depende de las necesidades de la aplicación y el grado en que se satisfagan los siguientes requisitos: (i) Universalidad, ii) Singularidad, (iii) Estabilidad, iv) Cuantificable, v) Aceptabilidad vi) Rendimiento, vii) Usurpación (Fierrez J. et al, 2008). El objetivo final de la utilización de estos requisitos, es el de poseer un conjunto de herramientas que permitan escoger que tecnología se adapta mejor a las necesidades de la aplicación. Los requisitos se pueden dividir en aquellos que están más estrechamente ligados al rasgo biométrico escogido:

- **Universalidad:** el rasgo se encuentra presente en un porcentaje mayoritario de la población.
- **Singularidad:** el rasgo permite hacer distinguir entre 2 individuos cualesquiera.
- **Estabilidad:** el rasgo no cambia durante un plazo de tiempo aceptable según la aplicación.
- **Cuantificación:** el rasgo se puede objetivar a través de procesos prestablecidos de caracterización de su información discriminante.

El resto de requisitos se encuadran dentro de aquellos que están más ligados al sistema automático a desarrollar y pueden variar incluso para un mismo rasgo biométrico:

- **Rendimiento:** el rasgo debe alcanzar el nivel de exactitud elevado para que sea considerada como aceptable.
- **Aceptabilidad:** el rasgo tiene un nivel de aceptación en el entorno para ser considerada como parte de un sistema de identificación biométrico.
- **Resistencia a ataques:** el rasgo debe permitir establecer el nivel de resistencia a técnicas fraudulentas.

La Tabla 1.2, presenta el cuadro comparativo de los principales rasgos biométricos evaluando sus características como rasgo identificador (A = nivel alto, M = nivel medio, y B = nivel bajo).

Tabla 1.2: Comparación cualitativa de los principales rasgos biométricos.
(Jain A. K., et al, 2014)

Características / Rasgo Biométrico	Universalidad	Singularidad	Estabilidad	Cuantificación	Rendimiento	Aceptabilidad	Resistencia a Ataques
ADN	A	A	A	B	A	B	B
Oreja	M	M	A	M	M	A	M
Cara	A	B	M	A	B	A	A
Termograma Facial	A	A	B	A	M	A	B
Venas de la Mano	M	M	M	M	M	M	B
Huella Dactilar	M	A	A	M	A	M	M
Movimiento Corporal	M	B	B	A	B	A	M
Geometría de la Mano	M	M	M	A	M	M	M
Iris	A	A	A	M	A	B	B
Huella Palmar	M	A	A	M	A	M	M
Olor	A	A	A	B	B	M	B
Retina	A	A	M	B	A	B	B
Firma	B	B	B	A	B	A	A
Forma de Teclear	B	B	B	M	B	M	M
Voz	M	B	B	M	B	A	A
Escritura	B	B	B	A	B	A	A

Para una muestra biométrica, los sistemas de reconocimiento cumplen habitualmente con una o dos funciones:

Identificación, se trata de responder a la siguiente pregunta: ¿Quién es la persona a la que pertenece dicha muestra?. Es decir tenemos información sobre una persona de la que desconocemos su identidad. Este tipo de comparación se denomina uno a muchos (1:N), puesto que para proceder a la identificación de una persona se deben comparar las características biométricas de la misma con las de todas las personas almacenadas en una base de datos.

Los requisitos de la Identificación son: 1) Una base de datos donde se almacenen las características biométricas de un amplio número de personas, 2) Un mecanismo para capturar y procesar las características biométricas de la persona a identificar, y 3) Un procedimiento para comparar las características de la persona a identificar con las almacenadas en la base de datos y que permita tomar la decisión de contestar a la pregunta anterior.

Verificación, se trata de responder a la siguiente pregunta: ¿Pertenece esta muestra a una persona determinada?. Es decir, una persona reclama tener una determinada identidad y el sistema debe verificar que es cierto. Este tipo de comparación se denomina uno a uno (1:1) puesto que únicamente se comparan las características biométricas de una persona con las de la persona que reclama ser. Los requisitos de la Verificación son: 1) Un sistema de autenticación tipo usuario más password al que se le puede añadir un sistema tipo carné de identidad, 2) Un mecanismo para capturar y procesar las características biométricas de la persona a identificar, y 3) Un procedimiento para comparar con un umbral las características de la persona a identificar con la previamente almacenada para esa persona y que permita tomar la decisión de contestar a la pregunta anterior.

Habitualmente se entiende por sistema de reconocimiento biométrico el conjunto de recursos software y hardware que permiten realizar la identificación o verificación de un individuo a partir de un rasgo biométrico.

Los sistemas de reconocimiento tradicional se pueden dividir en los siguientes bloques: 1) Adquisición, 2) Extracción de características, 3) Comparación entre características del usuario almacenadas, y 4) Toma de decisión. La Figura 1.1, muestra el diagrama de bloques de un sistema de entrenamiento conjuntamente con un sistema de Identificación/Verificación.

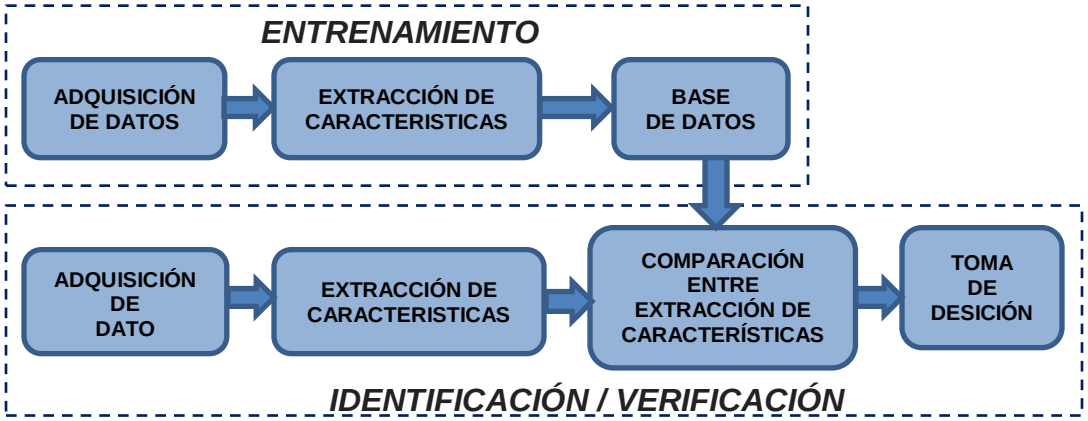


Figura 1.1: Sistema de reconocimiento tradicional

Entre los rasgos biométricos más populares se encuentran: la huella dactilar, el rostro y el iris. La huella dactilar fue el primer rasgo biométrico en ser utilizado a nivel masivo dado su conveniencia en aplicaciones forenses (ej. escenas de crimen) así como de control de fronteras (ej. Programa de visitas a USA). La cara se ha popularizado por ser un rasgo biométrico de fácil adquisición que permite un reconocimiento transparente sin necesidad de interaccionar con el usuario. Finalmente, el iris, ofrece tasas de discriminación tan elevadas como la huella dactilar, pero sufre del rechazo por parte de los usuarios del que adolece esta. Existen otros rasgos más (Jain A. K., et al, 2016; Jain A. K., et al, 2004; Martinez-

Diaz M., 2015) y en los últimos años se ha disparado el uso de estos sistemas. Algunos ejemplos recientes de este despliegue son:

- Programa Aadhaar de la India (Khanna T., Raina A., 2012): en el que se ha capturado la huella dactilar, el rostro y el iris de más de 600 millones de personas como parte del programa más ambicioso en la historia de las tecnologías biométricas. Estos rasgos serán utilizados para identificar a los ciudadanos de la India y asignarles un identificador único.

- FaceID de Apple: es un sistema de desbloqueo por reconocimiento facial para el iPhone X, desarrollado por Apple. La tasa de error del FaceID es de 1 entre un millón. Esto es posible al conjunto de cámaras que forman la tecnología TrueDepth que cuenta con tres nuevas características: una cámara de infrarrojos, un iluminador IR y un proyector de puntos. TrueDepth es capaz de componer un mapa en detalle del rostro y realizar fotos en 3D. Este nuevo sistema tiene como objetivo, reemplazar el sistema de reconocimiento de huellas, Touch ID. En el mundo, existen 16 millones de móviles vendidos con este sistema.

Dependiendo del grado de automatización y la necesidad o no de supervisión humana, los sistemas biométricos se pueden dividir en dos tipos: automáticos y semi-automáticos (Ver Figura 1.2).

Sistemas automáticos: sistemas que requieren respuesta en tiempo real y no hay intervención o supervisión del humano para la toma de decisión, por ejemplo: los móviles, el control de acceso, control de horario, pagos de compras, etc.

Sistemas semi-automáticos: sistemas donde su respuesta no necesariamente es en tiempo real y normalmente interviene el humano en el proceso, por ejemplo: Transacciones bancarias, análisis forense, entrega de mensajería y paquetes, control migratorio, etc.

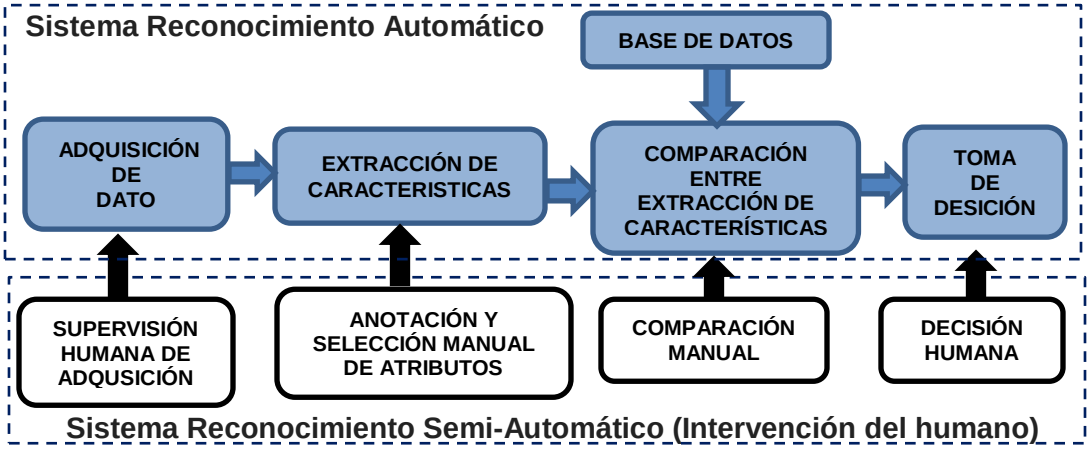


Figura 1.2: Sistema de reconocimiento Automático y Semi-Automático

1.2.- Reconocimiento de Firma

La firma es un rasgo biométrico de comportamiento mundialmente aceptado como un método de autenticación de identidad, y ha sido utilizada durante siglos por diferentes culturas en transacciones comerciales. La firma incorpora características neuromotoras del firmante definidas por sus habilidades cognitivas (cerebro) y motóricas (músculos, huesos, tendones) (Diaz M., et al, 2018). Además, hay que considerar la influencia sociocultural como son los estilos occidental y asiático. En los países occidentales y gran parte de los orientales, todos los ciudadanos tienen una firma con la que legitiman documentos y transacciones.

La firma manifiesta rasgos característicos como: de la mano que firma, del brazo que controla la mano, del corazón y el cerebro que articulan ese brazo, en fin todos estos aspectos manifiestan el comportamiento de la persona en el momento de realizar una firma. Es por ello que muchas veces se sitúa a la firma en la frontera entre los rasgos fisiológicos y de comportamiento. La Figura 1.7, muestra varias formas de la firma, que pueden ser representadas de diferentes maneras tales como: solo el nombre, o solo el apellido, el nombre y apellido, iniciales del nombre y apellido, y a esto le añaden lazos o círculos o algún tipo de adorno (Occidentales), o simplemente signos característicos (Orientales) que hacen que la firma de cada individuo sea un rasgo personal y único.

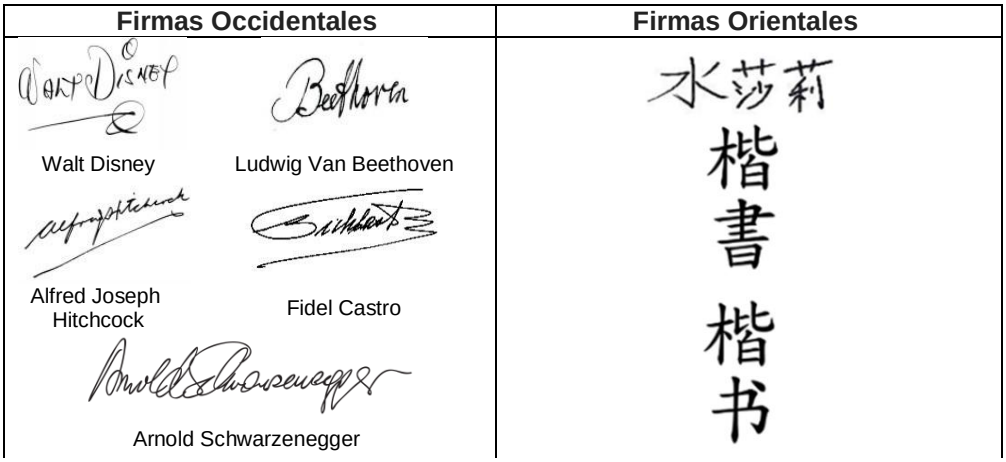


Figura 1.3: Ejemplos de Firmas manuscritas Occidental (izquierda) y Oriental (derecha)

Desde mucho tiempo atrás, el análisis forense ha permitido examinar y evaluar una firma, a través de expertos en análisis forense que determinan su autenticidad.

En la actualidad existen sistemas automáticos de verificación de firma, como ayuda a los Examinadores de Documentos Forenses (FDE) (Malik M. I., et al, 2013; Oliveira L. S., et al, 2005; Coetzer J., et al, 2006; Malik M. I., et al, 2013; Morocho D., et al, 2016; Morales A., et al, 2017). La firma es reconocida dentro de

la legislación como medio para dar veracidad a un documento y tiene aplicaciones de legalidad laboral como por ejemplo la designación de un ascenso laboral o designación y reubicación de actividades en una empresa, la aprobación o declinación de un proyecto, por ende la firma del contrato del mismo. En conclusión una firma ratifica una decisión que nos ata o libera de responsabilidades.

Actualmente, el reconocimiento automático de firma es un área de investigación activa que aglutina a numerosos grupos de investigación de todo el mundo trabajando en el desarrollo de sistemas cada vez más fiables (Plamondon R., et al, 1989; Impedovo D. et al. 2008; Diaz M., et al, 2018).

La firma manuscrita a pesar de ser el rasgo o método de verificación e identificación más aceptado a nivel mundial, aún su investigación y difusión en sistemas automáticos continúa siendo compleja debido a factores como la variabilidad y la permanencia de la firma. La alta variabilidad intra clases, la baja variabilidad inter clases y la baja permanencia son factores que conllevan a seguir en la investigación de mejoras de los sistemas de reconocimiento de firmas.

Alta variabilidad intra-clase: este factor se genera cuando una misma persona puede realizar diferentes versiones de su propia firma en diferentes sesiones, en este caso consideradas todas las firmas como genuinas, pero con alta variabilidad, pero como es la misma persona esta variabilidad es conocida como intra-clases, este factor también es considerado en la verificación de identidad del usuario. La variabilidad intra-clases se debe a factores como el estado anímico (emociones, salud), físico (dolencias musculares) o la posición del individuo al momento de firmar.

Baja variabilidad inter-clase: este factor se genera cuando diferentes personas puede realizar diferentes versiones de una misma firma en diferentes sesiones, en este caso consideradas todas las firmas como falsificadas, pero con baja variabilidad, pero como son diferentes personas esta variabilidad es conocida como inter-clases, las reproducciones de firmas por parte de falsificadores pueden ser muy similares a las firmas genuinas, superando el umbral determinado para la verificación.

Baja permanencia: la firma de un individuo tiende a variar con el paso del tiempo.

Los sistemas automáticos de verificación de firmas (ASV por sus siglas en inglés *Automatic Signature Verification*) tienen por objetivo diferenciar entre firmas genuinas y falsificaciones. Los sistemas ASV se dividen tradicionalmente en dos dependiendo de la naturaleza de la información utilizada: firma off-line o estática y firma online o dinámica.

Autenticación de firmas off-line o estáticas: Las firmas son desarrolladas usando un lápiz de tinta y la información es digitalizada por escáneres ópticos. La

autenticación es ejecutada a través del análisis de características visuales de la firma incluyendo la morfología, la textura y la geometría. Las aplicaciones potenciales son relacionadas mayormente con análisis de documentos (Morales A., et al, 2017).

La verificación de una firma off-line es más difícil debido a la falta de valiosa información sobre el comportamiento de cómo la persona creó la firma en términos de velocidad y aceleraciones de la escritura, presión de escritura y secuencia de trazos.

El proceso de digitalización de una firma manuscrita off-line a partir de una muestra (firma) obtenida en papel, se puede realizar mediante el uso de un escáner o una cámara digital. Este tipo de adquisición también es conocido como firma estática (ver Figura 1.4).

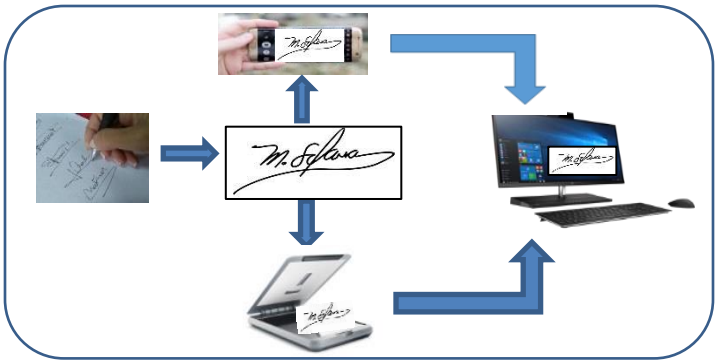


Figura 1.4: Adquisición de la firma Off-line a través de escáneres y cámara digital.

Acondicionamiento para la firma off-line: La adquisición off-line radica en la digitalización de la imagen de una firma, por lo tanto se aplican técnicas de procesamiento digital de imágenes:

- **Binarización,** es una técnica de reducción de la escala de grises a dos únicos valores (Negro=0 y Blanco=255), estableciendo un valor de umbral.

- **Eliminación de Ruido,** este proceso emplea filtros pasabajos y se pueden realizar antes o después del proceso de binarización.

- **Segmentación,** consiste en aislar los trazos que contienen la información necesaria para caracterizar una firma. Se puede extraer toda la firma o solamente el cuerpo de la misma.

- **Normalización en Posición y Tamaño,** este proceso ubica la imagen en un nivel de posición inicial y un tamaño normalizado para análisis de la imagen.

Autenticación de firmas on-line o dinámicas: Las firmas son adquiridas con dispositivos digitales que capturan las secuencias temporales del proceso de firmar. La autenticación es ejecutada que contiene parámetros globales (ej. tiempo total y número de correcciones) o funciones temporales derivadas de secuencias adquiridas (ej. velocidad y aceleración). Las aplicaciones de este tipo incluyen aquellos relacionados con sistemas de autenticación automática en tiempo real (ej. puntos de venta, entrega de servicios y autenticación de móviles) (Morales A., et al, 2017).

La figura 1.5, muestra las últimas innovaciones en tecnología en cuanto a pantallas táctiles que han proporcionado un entorno factible para la verificación de firmas on-line, como son los teléfonos inteligentes, tabletas y las tabletas graficas digitalizadoras (wacon), que es uno de los equipos de mayor precisión para la captura de firmas, que contienen información adicional de la firma como: la presión ejercida (P), trayectoria (Azimuth y Altitud), inclinación del lápiz (Ángulo), entre otros (ver Figura 1.6). Este modo de adquisición es conocido como adquisición dinámica.

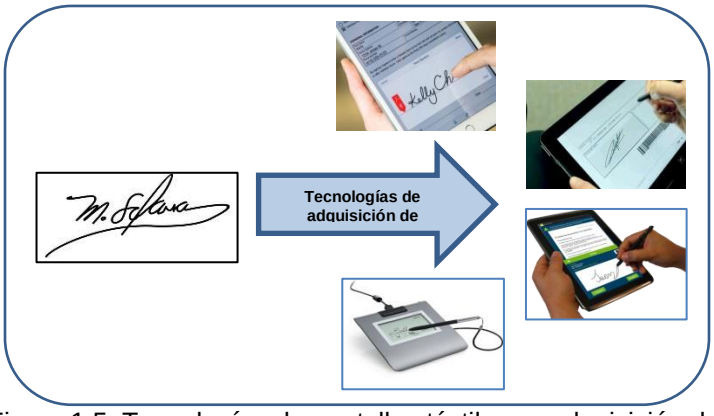


Figura 1.5: Tecnologías de pantallas táctil para adquisición de firmas

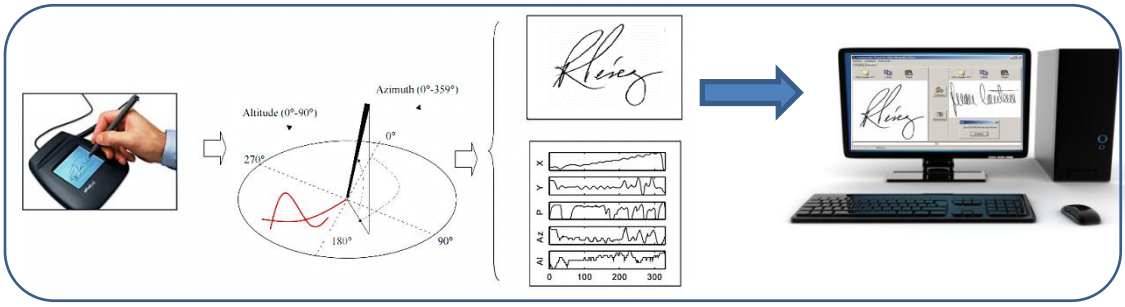


Figura 1.6: Tableta grafica digitalizadora (wacon)

La verificación de firmas on-line continua siendo una tarea compleja en el campo de la biometría, esto se debe a su naturaleza conductual, las firmas presentan una alta variabilidad incluso entre realizaciones sucesivas. Esto conduce a tasas de error más altas que otros rasgos, como el iris o las huellas dactilares.

Acondicionamiento para la firma on-line: En el pre-procesado de firmas capturadas on-line se busca obtener una representación robusta respecto a las tres variaciones geométricas básicas: rotación, traslación y el escalamiento de los diferentes trazos de las firmas. El pre-procesado que se aplica para una firma on-line son los siguientes:

- **Alineamiento Respecto a la Posición:** se realiza respecto referencial al punto inicial o el centro de masas.

- **Normalización en rotación:** se realiza la alineación con respecto al ángulo de la trayectoria media, obtener una representación de la firma en coordenadas polares y normalizar respecto al ángulo medio, o normalizar respecto al eje de mínimo momento de inercia.

- **Normalización del Tamaño:** se realiza con respecto a valores extremos de las coordenadas, rangos de variación o estadísticos de primer y segundo orden.

Grado de cumplimiento de la firma como rasgo biométrico: La firma manuscrita es un rasgo con un alto grado de aceptabilidad, que permite verificar a una persona de otra.

Tabla 1.3: Grado de cumplimiento de la firma como rasgo biométrico.

Características / Rasgo Biométrico	Universalidad	Singularidad	Estabilidad	Cuantificación	Rendimiento	Aceptabilidad	Resistencia a Ataques
Firma	B	B	B	A	B	A	A

La Tabla 1.3, muestra el grado de cumplimiento de las características de la firma como rasgo biométrico:

- **Universalidad:** Cumple un grado de cumplimiento Bueno (B). Esto se debe a que un 84% de la población no sabe leer ni escribir, pero el 16% de la población si tiene una firma como un rasgo de distinción.

- **Singularidad:** Cumple un grado de cumplimiento Bueno (B). Esto se debe a que la firma en su mayoría de aplicaciones cumple con la función de solo la verificación de la persona y no con la identificación.

- **Estabilidad:** Cumple un grado de cumplimiento Bueno (B). Esto se debe a factores comportamiento del humano que conlleva al envejecimiento de la firma.

- **Cuantificación**: Cumple un grado de cumplimiento Alto (A). Esto se debe a que sus datos son medibles. En la actualidad con la tecnología de última generación se pueden recoger firmas con cualquier dispositivo táctil (Smartphone, Tablet, PDS), o a través de una firma sobre un papel y llevando a un escáner y ser digitalizado.

- **Rendimiento:** Cumple un grado de cumplimiento Bueno (B). Esto se debido a que los niveles de identificación deben alcanzar un alto grado de exactitud para que sea considerada como aceptable.

- **Aceptabilidad:** Cumple un grado de cumplimiento Alto (A). Esto se debe al alto grado de aceptabilidad de las persona como un método para validar documentos financieros y legales utilizados por cientos de años.

- **Resistencia a ataques:** Cumple un grado de cumplimiento Alto (A). Esto se debe a la complejidad que tiene cada firma y a las cualidades caligráficas de cada firmante que generan gran cantidad de información discriminatoria que se puede extraer de la firma de un usuario, generando una resistencia a ser eludidos.

El análisis de las características de la firma como rasgo biométrico que se reflejan en la Tabla 1.3 y las características de otros rasgos presentados en la Tabla 1.2, desprenden ventajas y desventajas que ningún rasgo biométrico cumple con éxito todos los atributos.

Sistemas automáticos de reconocimiento de firma: sistemas que requieren respuesta en tiempo real, mientras la persona está realizando la firma sobre un dispositivo de adquisición de datos, el sistema automático de reconocimiento está extrayendo características, comparando con la base de datos y finalmente toma la decisión de si es o no la firma genuina (Verificación / Identificación).

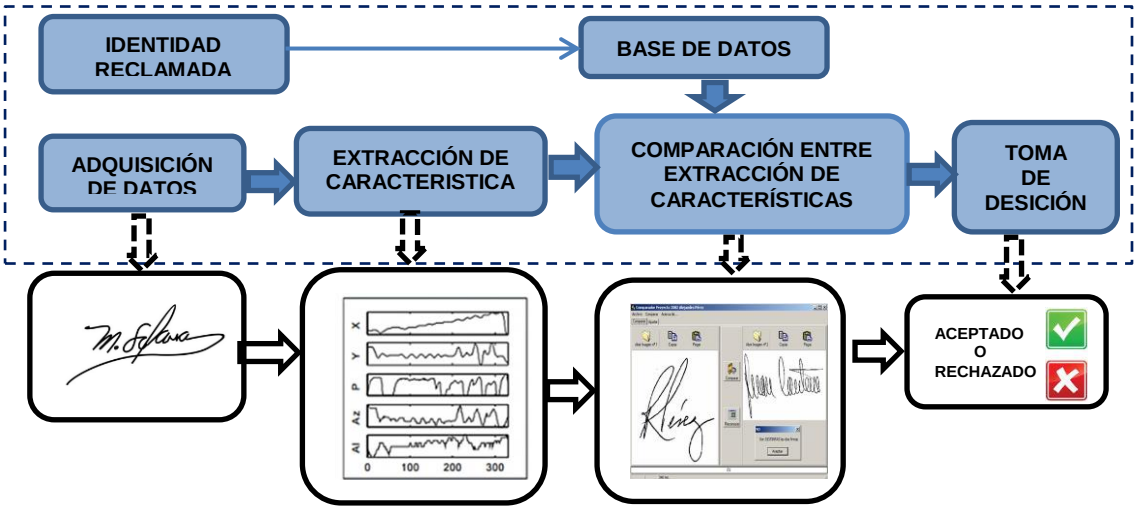


Figura 1.7: Arquitectura típica de un sistema automático de reconocimiento de firma

La Figura 1.7, muestra la arquitectura típica de un sistema automático de reconocimiento de firma con sus diferentes etapas:

1. **Adquisición de datos:** los datos de una firma son capturadas a través de dispositivos táctiles y tabletas digitales (posición X, Y, presión p). La cantidad de información capturada depende del dispositivo utilizado. Las tabletas digitales proporcionan información relevante como se refleja en investigaciones de (Martinez-Diaz M., et al, 2014; Houmani N., et al, 2011). Esta información adicional capturada son: la presión, el ángulo de inclinación y la trayectoria del bolígrafo durante el proceso de firmado (penup). Los datos adquiridos son muestreadas en el tiempo y posteriormente almacenadas en forma de series discretas. La frecuencia de muestreo utilizada por los dispositivos de captura suele ser entre 100-200 Hz. Esta frecuencia de muestreo cumple con el criterio de Nyquist, ya que la mayor frecuencia observada a la hora de realizar una firma oscila entre 20-30 Hz (Martinez-Diaz M., et al, 2009). Después de la adquisición de datos normalmente existe una etapa de pre-procesamiento. En esta etapa se realiza el procesamiento de la señal como: filtraje de ruido, técnicas de diezmado y/o interpolación para eliminar o recuperar muestras, logrando datos más confiables.

2. **Extracción de características:** en esta etapa se obtiene la información discriminatoria de un firmante a partir de los datos de la firma On-line. Se distinguen dos tipos de aproximaciones:
   - Sistemas basados en características: trabaja en la extracción de características globales de la firma por ejemplo: el tiempo de realización de la firma, número de pens-ups, y velocidad media, esto con el objetivo de obtener un vector de características discriminatorio de cada usuario.
   - Sistemas basados en funciones: trabaja en la comparación de secuencias en el tiempo de propiedades de la firma por ejemplo: la presión, la trayectoria.

3. **Comparación entre extracción de características:** en esta etapa se realiza la comparación de similitud. En sistemas basados en características se aplican técnicas con métricas de distancias (Mahalanobis, Euclidea) entre las características extraídas del firmante y la característica extraída almacena en la base datos. En sistemas basados en funciones se utilizan otro tipo de técnicas como DTW (Dynamic Time Warping) o HMM (Hidden Markov Models) para comparar entre modelos de firmas.

4. **Toma de decisión:** en esta etapa antes de tomar de decisión si es una firma genuina o impostora, las puntuaciones generadas después de realizar la etapa de comparación son normalizadas. Este proceso es muy importante cuando se utilizan técnicas de fusión de sistemas. En (Jain A. K., et al, 2005), se presentan algunas técnicas de normalización más aplicadas en sistemas biométricos.

Sistemas semi-automáticos de reconocimiento de firma: sistemas donde su respuesta no necesariamente es en tiempo real, la persona realiza procesos de etiquetación o verificación manual de una firma en proceso toma la decisión de una firma si es genuina o falsificada, por ejemplo: Transacciones bancarias, análisis forense, entrega de mensajería y paquetes, control migratorio, etc.

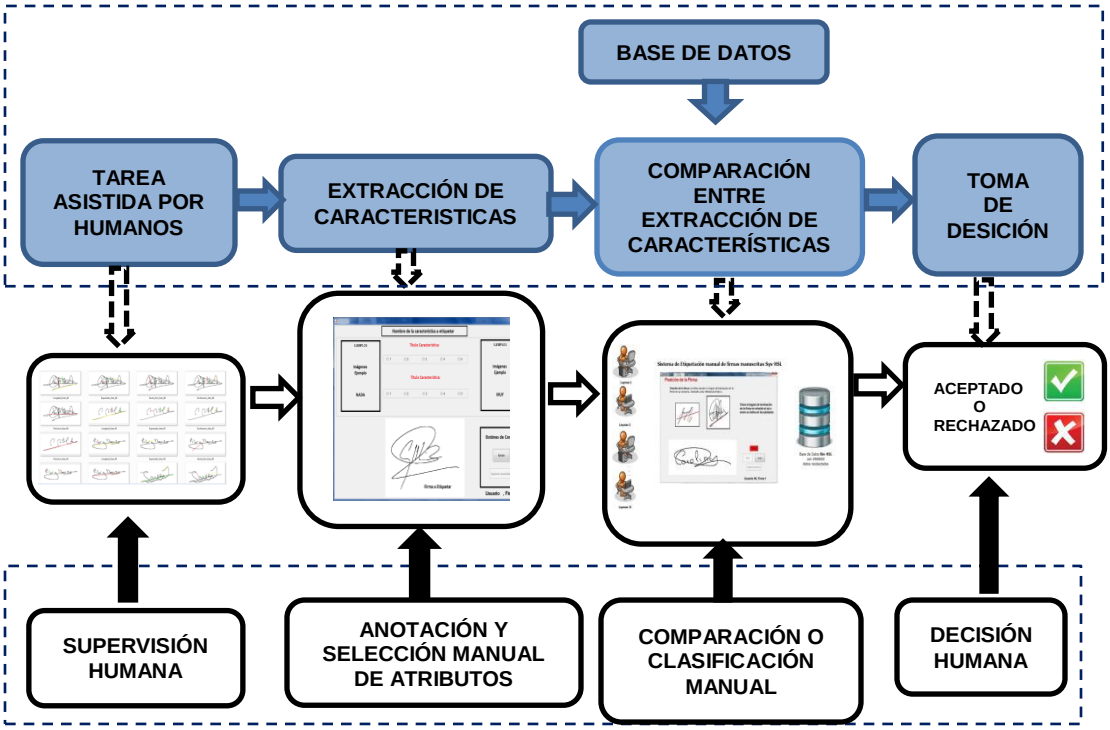


Figura 1.8: Arquitectura típica de un sistema semi-automático de reconocimiento de firma

La Figura 1.8, muestra la arquitectura de un sistema semi-automático de reconocimiento de firma con sus diferentes niveles. En biometría, los esquemas asistidos por humanos aprovechan tanto las habilidades humanas como las capacidades de un sistema automatico (Kumar N., et al. 2011; Reid D., et al, 2014; Klare B.F., et al, 2014; Samangouei P., et al, 2016; Tome P., et al, 2014). La Figura 1.8, muestra la intervención humana en un sistema biométrico, donde, se pueden realizar a diferentes niveles de acuerdo con las diferentes tareas o etapas a realizar:

1. **A nivel de imagen:** la evaluación y supervisión humana de calidad para descartar muestras con grandes distorsiones
2. **A nivel de característica:** la anotación manual de atributos discriminativos
3. **A nivel de comparación o clasificación:** las calificaciones humanas en forma de valores escalares
4. **A nivel de decisión:** la toma de decisión binaria, genuina o falsa.

Existen varias aplicaciones, donde el humano realiza la supervisión de una firma, pero solo se dedican a tomar un registro de verificación de datos. Estos supervisores son personas sin experiencia en análisis de FDEs, en (Malik M. I., et

al, 2013; Oliveira L. S., et al, 2005; Coetzer J., et al, 2006; Malik M. I., et al, 2013; Morocho D., et al, 2016; Morales A., et al, 2017) se refieren a estas personas como profanos. Los sistemas automatizados están minimizando la intervención humana en muchas aplicaciones de reconocimiento. Sin embargo, su capacidad analítica y su percepción de identificar más objetivos, podrían permitir al humano mejorar un proceso de reconocimiento (Malik M. I., et al, 2013; Oliveira L. S., et al, 2005; Coetzer J., et al, 2016).

Si exceptuamos el trabajo de los peritos caligráficos y demás profesionales. Las tareas humanas relacionadas con el reconocimiento de firma se reducen casi exclusivamente a comprobar la correcta adquisición de la muestra. Existen escenarios donde un profano puede contribuir a una verificación de firma, como en la banca, la notaría pública, entrega de paquetes, negocios, control de migración, en recolección de firmas para consulta popular, suplantación de identidad, etc. En estos escenarios la interacción humana para la evaluación de la firma es muy importante por la toma de decisión en corto tiempo, este rasgo biométrico posee una alta intra-variabilidad propia del humano (Morales A. et al, 2017; Morocho D. et al, 2017; Morocho D. et al, 2016).

Las aplicaciones del reconocimiento de firma automático han generado inquietudes que conllevan a una investigación de ver las potencialidades, el rendimiento y las influencias del humano en este tipo de sistemas:

- ¿Puede la acción de una persona ayudar a mejorar el rendimiento de un sistema automático de reconocimiento de firma?
- ¿Qué tan bueno es una persona en reconocer la autenticidad de una firma?
- ¿Cómo puede influir el comportamiento conductual de la persona en el reconocimiento de firma?
- ¿El rendimiento de la verificación de firmas realizada por humanos es fiable?

Estos temas de interés, se generan, debido a que en los últimos años se han hecho grandes esfuerzos por automatizar muchos procesos. Se han propuesto grandes mejoras a nivel algorítmico, de sensores y de procesos. En este contexto, en determinadas aplicaciones, la intervención humana se ha ido diluyendo y sus capacidades infravalorándose. La Figura 1.9, muestra un típico sistema de verificación automático de firma, donde existe dos problemas fundamentales en el reconocimiento biométrico como son: 1) Encontrar una representación de característica invariable y 2) diseñar un comparador robusto. Es aquí donde el humano podría intervenir y combinar para mejorar el reconocimiento de firmas y poder ver sus potencialidades como observación, intuición, comprensión, razonamiento y otras cualidades que le permitan ayudar a mejorar los Sistemas Automáticos de Reconocimiento.

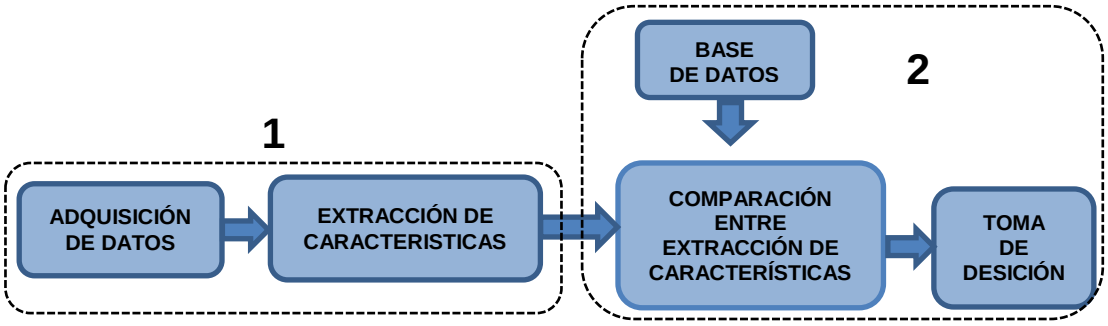


Figura 1.9: Operación de un sistema Automático de Verificación de Firma (AVS).

El reconocimiento automático de firma se enfrenta a grandes retos como: Alta variabilidad de firmante, Alta variabilidad de estilo, y Usurpación de identidad:

Alta variabilidad de firmante: una persona nunca firma igual dos veces. Esto hace que exista incertidumbre a la hora de generar las plantillas que caracterizan a un firmante. Para el caso particular de la firma, el desarrollo de un sistema de procesamiento enfrenta un gran obstáculo: la variabilidad de la escritura y de los estilos de la firma. Es por esta razón que los investigadores se enfocan en obtener un mejor conocimiento de la escritura. Desde, esta perspectiva, hay que destacar la existencia de numerosos procesos fisiológicos complejos en los que se han detectado variaciones periódicas regulares e irregulares, algunas de los cuales han sido interpretadas como el reflejo de estados de ánimo (Morales A., et al, 2017).

Alta variabilidad de estilo: existen usuarios con firmas complejas que incluyen nombres, apellidos y florituras. Mientras que existen usuarios con firmas sencillas que contienen muy poca información (Tolosana R., et al, 2017).

Usurpación de identidad: falsificar un rasgo como la firma es más sencillo que falsificar cualquier otro rasgo

Los retos en el reconocimiento de firma se han logrado minimizar poco a poco, a través de bases de datos robustas que permitan hacer más y mejor entrenamiento de los sistemas de reconocimiento de firma. Los parámetros más utilizados en la literatura para medir la eficiencia o el rendimiento de un sistema ASV son: la tasa de falsa aceptación (porcentaje de falsificaciones que el sistema reconoce como firmas genuinas) o *False Acceptance Rate* (FAR), y la tasa de falsos rechazos (porcentaje de firmas genuinas que el sistema rechaza por considerarlas falsificaciones) o *False Rejection Rate* (FRR). El punto en el que la FAR y la FRR son iguales se conoce como *Equal Error Rate* (EER).

Las bases de datos de firmas son conjuntos estructurados de firmas manuscritas recopiladas de un grupo de individuos, que se utilizan para la

evaluación de algoritmos de reconocimiento. (Martinez-Diaz M, et al., 2009). Uno de los mayores problemas era la carencia de bases de datos públicas, debido en gran medida a los problemas legales y a la privacidad de los usuarios, ya que a diferencia de otros rasgos biométricos, las firmas se pueden falsificar con relativa facilidad. En la actualidad se dispone de varios conjuntos de bases de datos de firmas manuscritas dinámicas, en gran medida al esfuerzo realizado por las instituciones educativas y grupos de investigación biométrica. (Tolosana R., et al, 2014).

Existen varias bases de datos generadas por grupos de investigación en el área de reconocimiento de firma manuscrita. Estas bases de datos han permito generar un punto de partida o de inicio (base line), para el desarrollo de diversos sistemas de reconocimiento de firma:

Base de Datos PHILIPS

La base de datos de firmas “Philips” fue capturada con una Tablet de digitalización interactiva avanzada (PAID) de la misma marca, con una frecuencia de muestreo de 200 Hz. En cada punto muestreado, el digitalizador captura las coordenadas de posición, presión. Azimut y altitud. (Martinez-Diaz M, et al., 2009). Esta base de datos contiene 1.530 firmas auténticas, 1.470 falsificaciones "over the shoulder" (30 firmas/usuario a excepción de 2 usuarios), 1.530 falsificaciones "home improved" (30 firmas/ usuario) y 200 falsificaciones "profesionales" (10 firmas para 20 usuarios).

Base de Datos MCYT (Ministerio De Ciencia Y Tecnología)

La base de datos MCYT fue adquirida por una iniciativa coordinada entre cuatro universidades españolas en el marco de un proyecto financiado llamado MCyT2000. Es una base de datos bimodal ya que se compone de firmas y huellas dactilares de 330 usuarios. Las firmas se adquirieron usando una tablet Wacom Intuos A6 con una frecuencia de muestreo de 100 Hz, con este instrumento se capturan las siguientes secuencias de tiempo: coordenadas de posición, presión, azimut y altitud. (Martinez-Diaz M, et al., 2009).

Base de Datos BIOMET

La base de datos BIOMET es multimodal y está compuesta por cinco rasgos: Imagen facial (en 2D y 3D), huella dactilar, geometría de la mano, firma y voz. (Gaytán S., et al, 2014) Las firmas se capturaron usando una Pen Tablet Wacom Intuos 2, con una frecuencia de muestreo de 100 Hz. Se capturaron las coordenadas de posición, presión, azimut y altitud. La base de datos contiene un total de 84 usuarios, con 15 firmas genuinas y 12 falsificaciones por usuario. Las firmas fueron capturadas en 2 sesiones con un espacio de tiempo entre ellas de 5 meses.

Base de Datos BIOSECURE

La Base de Datos Multimodal BioSecure tiene datos de rasgos de rostro, huella dactilar, mano, iris, firma y voz. Incluye tres conjuntos de datos DS1 y dos subcorpus de firmas correspondientes a los conjuntos de datos DS2 y DS3.

- DS1 fue capturado remotamente a través de Internet,
- DS2 se adquirió en un entorno de escritorio
- DS3 en condiciones móviles.

Los conjuntos de datos de firmas fueron producidos por un grupo de 667 usuarios. El conjunto de datos DS2 se capturó utilizando un digitalizador Wacom Intuos3 A6 a 100 Hz y el conjunto de datos DS3 se capturó con un PDA. Se solicitó a los usuarios que firmaran mientras estaban de pie y sostenían el PDA en una mano, emulando condiciones de operación realistas. Las señales de posición, presión, azimut y altitud de la pluma están disponibles en DS2, mientras que sólo la posición está disponible en DS3 debido a la naturaleza de la pantalla táctil del PDA (Martinez-Diaz M, et al., 2009).

Las firmas fueron capturadas en dos sesiones y en bloques de 5. Se dejó un promedio de dos meses entre cada sesión. Durante cada sesión, se pidió a los usuarios realizar 3 series de 5 firmas genuinas y 5 falsificaciones entre cada conjunto. Siguiendo este protocolo, cada usuario realizó 5 falsificaciones para los 4 usuarios anteriores en la base de datos. Así, 30 firmas genuinas y 20 falsificaciones están disponibles para cada usuario.

Base de Datos BiosecurID

Esta base de datos fue recogida por 6 diferentes instituciones de investigación española. Incluye los siguientes rasgos biométricos: habla, iris, rostro, firma, escritura a mano, huellas dactilares, mano y pulsación de teclas. Los datos fueron capturados en 4 sesiones distribuidas en un período de 4 meses a un total de 400 usuarios.

Las señales de posición, presión, azimut y altitud de firma se adquirieron usando un digitalizador Wacom Intuos3 A4 a 100 Hz. Durante cada sesión, dos firmas fueron capturadas al principio y dos al final, dando lugar a 16 firmas genuinas por usuario. Cada usuario realizó una falsificación por sesión de firmas de otros tres usuarios en la base de datos, en total se obtuvieron 12 firmas falsificadas (Galbally J., et al, 2010).

Base de Datos BiosecurID-SONOF DB

La base de datos más recientes es la BiosecurID-SONOF DB (Fierrez J., et al, 2010), fue generada por el Grupo de Reconocimiento Biométrico ATVS de la Universidad Autónoma de Madrid, los datos recopilados comprenden: firmas on-

line y off-line. Estos datos fueron adquiridos en 4 sesiones en un lapso de tiempo de 4 meses. Está compuesta por 132 firmas distintas, cada usuario posee un total de 16 firmas genuinas y 12 firmas impostoras, dando un total 2112 firmas genuinas y 1584 firmas falsificadas, dando un total de 3696 firmas.

La tabla 1.4, muestra un resumen de las firmas almacenadas en cada una de las bases de datos mencionadas anteriormente.

Tabla 1.4: Resumen de la composición de las Bases de Datos de Firmas

Base de Datos	Firmas Genuinas por Usuario	Firmas Falsificadas por Usuario	Firmas por usuario	Número de Usuarios	Total de Firmas
PHILIPS	16	70	86	51	4730
MCyT	25	25	50	330	16500
BIOMET	15	17	32	91	2912
BIOSECURE	30	20	50	667	33350
BIOSECUREID	16	16	32	400	12800
BIOSECUREID-SONOF DB	16	12	28	132	3696

1.3 Human in the loop: estrategias de colaboración hombre-máquina para reforzar el aprendizaje automático

Los sistemas de crowdsourcing reclutan a una multitud de seres humanos para ayudar a resolver una amplia variedad de problemas. En la actualidad se generan actividades de soporte para la investigación de áreas de interés científico, a través de los sistemas de crowdsourcing en la Web. Es importante tener en cuenta que ahora se cuenta con varias plataformas de crowdsourcing, tales como: Mechanical Turk (utilizada en esta tesis), Turkit, Mob4hire, uTest, Freelancer, eLance, oDesk, Guru, Topcoder, Trada, 99design, Innocentive, CloudCrowd y Cloud-Flower. Estas plataformas permiten construir rápidamente sistemas de crowdsourcing en varias líneas de investigación. Además es un método de resolución de problemas de propósito general, que enlista a una multitud de personas para ayudar a resolver un problema definido por los propietarios del sistema.

La revista, “Aprendizaje Activo del Mundo Real: Aplicaciones y Estrategias para el Aprendizaje de maquina Human-in-the-loop”, examina el campo del "aprendizaje activo". Este campo está explotando con aplicaciones prácticas que demuestran la eficiencia de la combinación de inteligencia humana y los sistemas automáticos.

Stitch Fix, Eric Colson, experto en algoritmos, codifica los atributos de una prenda de vestir (color, talla, estilo, material, marca, precio y tendencia), atributos

que se comparan con el perfil de un cliente y el sistema automático genera recomendaciones basadas en el modelo. En este momento es posible que sistema automático no pueda realizar las recomendaciones para el cliente. Aquí es donde intervienen los estilistas (humanos) de Stitch Fix, que entrega una selección de recomendaciones a uno de los aproximadamente 1.000 estilistas humanos, cada uno de los cuales atiende a un conjunto de clientes.

En el informe, **Cuzzillo** nos lleva desde las recomendaciones de moda hasta el mapeo de ubicaciones off-road en Google. Los algoritmos recopilan datos de imágenes satelitales y aéreas de calles vistas por Google, y extraen datos como números de calles, límites de velocidad y puntos de interés. Sin embargo, incluso en Google, los algoritmos solo lo llevan a cierto punto, y luego los seres humanos deben intervenir para verificar y corregir manualmente los datos. Además se sumerge en la tendencia estrechamente relacionada del crowdsourcing, una forma crítica de etiquetar rápidamente cientos o incluso miles de elementos que finalmente se incorporan a un algoritmo para mejorar su rendimiento.

Adrian Bridgwater, menciona que la Inteligencia Artificial (IA) tiene un problema: es artificial, debido a que AI y sus disciplinas relacionadas de aprendizaje automático, computación cognitiva, análisis de sentimientos y redes neuronales tienen un problema, se crean artificialmente a través del poder de los algoritmos de los desarrolladores de software. Además manifiesta la regla 80:20, permitirá asegurarnos de que siempre haya un factor humano en el bucle.

Lukas Biewald, manifiesta que debido a la gran cantidad de sistemas realizados con inteligencia artificial y aprendizaje automático, tales como: autos autónomos, dispositivos inteligentes y hasta la función de etiquetado de fotos de Facebook. Todas estas tecnologías necesitan un grado de humano en ellas.

CrowdFlower, se define a sí mismo como una "empresa de enriquecimiento de datos". Esto se adentra en el área llamada "aprendizaje activo", donde el algoritmo de aprendizaje es un programa de computadora que puede hacer preguntas periódicas e interactivas a usuarios para recopilar la información deseada. Las plataformas de enriquecimiento de datos se han convertido en un recurso valioso para los científicos, que buscan datos para automatizar y mejorar el etiquetado y el enriquecimiento de datos utilizando la inteligencia humana para el aprendizaje automático, es decir, el aprendizaje activo mejora y fortalece su inteligencia artificial por la calidad de los datos de capacitación que recibe de los colaboradores de las plataformas de enriquecimiento de datos.

Esta tendencia es cada vez más pronunciada e incluso Google aún utiliza a los seres humanos para desarrollar su "inteligencia" y su capacidad de búsqueda. Pinterest también ha utilizado este tipo de sistema para ayudar a filtrar ciertos tipos de contenido en línea como: pines sexualmente explícitos, lenguaje soez, contenido engañoso. Existen tareas donde los seres humanos tienden a hacer mejor la evaluación de contenido, debido que pueden analizar con relevancia los resultados de búsqueda y filtrar ciertos tipos de contenido. Para este proceso se

utiliza el crowdsourcing para hacer todo, desde evaluar la relevancia de la búsqueda hasta comparar grupos de tratamientos de experimentos.

En conclusión nos hacemos esta pregunta: **¿Por qué todavía necesitamos humanos?**, Hay todo tipo de razones por las que todavía necesitamos a los seres humanos por ejemplo en aplicaciones con Inteligencia Artificial. Desde los matices del lenguaje hablado hasta los errores tipográficos inesperados, ni una sola computadora ni un solo sistema robusto pueden ser perfectos. **Biewald** argumenta que los modelos de IA que no tienen algún tipo de elemento humano en el bucle son defectuosos. Además, las computadoras son excelentes para analizar situaciones tácticas difíciles, pero aún no son tan buenos como los humanos para entender la estrategia a largo plazo.

1.3.1.- Analizando y explotando las capacidades humanas en el reconocimiento biométrico

Los esquemas asistidos por personas en biometría aprovechan tanto las habilidades humanas como las capacidades del sistema automatizado (Kumar N., et al, 2011; Reid D. et a, 2014; Klare B.F., et al, 2014; Samangouei P., et al, 2016; Tome P., et al, 2014). La intervención humana en sistemas biométricos se puede realizar a diferentes niveles de acuerdo con las diferentes tareas a realizar: a nivel de imagen (evaluación de calidad para descartar muestras con grandes distorsiones); a nivel de característica (anotación manual de atributos discriminativos); en el nivel de comparación o clasificación (calificaciones humanas en forma de valores escalares); y finalmente a nivel de decisión (decisión binaria, genuina o falsa).

El estudio del rendimiento humano en aplicaciones biométricas ayuda a comprender mejor el potencial y las capacidades de los sistemas automáticos (Best-Rowden L., et al, 2014; Han H., et al, 2015). Las habilidades humanas se utilizan comúnmente como punto de referencia para la evaluación de algoritmos automáticos (Coetzer J., et al, 2006; Phillips P.J., et al, 2015). En general, los seres humanos reconocerán mejor los rasgos biométricos como la cara, la firma o la voz que otras características como la huella digital, el iris y la huella de la palma. Estudios revelan que los humanos pueden ser muy inexactos en el reconocimiento de características biométricas como las caras de personas desconocidas (Best-Rowden L., et al, 2014; Phillips P.J., et al, 2015). La intervención humana en la autenticación de firmas se relaciona con las ciencias forenses. Sobre la base de su formación y experiencia, las FDE analizan la autenticidad de una firma determinada de acuerdo con un conjunto de evidencias. La anotación de atributos de firmas es una tarea común en el análisis de FDE y consiste en etiquetas discretas (por ejemplo, la firma tiene la puntuación adecuada) o medidas escalares de características específicas (por ejemplo, una longitud de trazo de 6 mm).

1.4 Motivación de la tesis

La firma manuscrita es uno de los métodos de autenticación personal más aceptados y se ha utilizado durante los últimos 2000 años. Como rasgo biométrico conductual, una de la característica clave de la firma es su alta variabilidad intrapersonal. La alta variabilidad entre las muestras de la firma de la misma persona junto con la habilidad de las personas para realizar falsificaciones especializadas hace que el reconocimiento de firmas sea un gran desafío. Históricamente, el reconocimiento de firmas lo realizan Examinadores de Documentos Forenses (FDE) que han desarrollado protocolos y métodos bien establecidos para analizar la autenticidad de una firma de consulta. Esta es una tarea manual que requiere mucho tiempo y depende de la capacitación y experiencia de FDE. Por lo tanto, las aplicaciones se limitan a autenticaciones sin requisitos de respuesta en tiempo real (análisis forense y escenarios fuera de línea). Los sistemas automáticos de verificación de firmas (ASV) surgieron como una forma viable de automatizar el método tradicional de verificación de firmas hecho por FDE's (Impedovo D., et al, 2008; Martinez-Diaz M, et al., 2009) y ampliar las aplicaciones potenciales.

La motivación que conllevan a esta investigación trasciende a varias preguntas de las potencialidades del humano en el reconocimiento de firma, relacionadas con factores como: la evaluación de calidad para eliminar muestras de mala calidad, anotación de características, clasificación de muestras o soporte de decisiones, entre otros. Sin embargo, el rendimiento del humano en tareas de verificación de firmas permanece sin explorar y sus capacidades minimizadas (Kumar N., et al, 2011):

¿Cuál es el rendimiento humano (no experto) en el reconocimiento de firma?: La variedad de aplicaciones basadas en sistemas automáticos de reconocimiento de firmas es grande (la banca, puntos de venta, entrega de paquetes, notaría pública). En la mayoría de estas aplicaciones, los humanos supervisan el proceso de firma, pero sus responsabilidades se limitan principalmente a garantizar el correcto registro de los datos (sin ningún impacto en el análisis de la autenticidad). Estos supervisores no tienen la experiencia específica de FDE.

El despliegue de sistemas automatizados está reduciendo la confiabilidad de las habilidades humanas. Sin embargo, la percepción y la capacidad analítica de los humanos no deberían estar infravaloradas y hay un amplio margen para las mejoras que explotan tanto la eficiencia de las computadoras como las capacidades humanas. La anotación de atributos realizada por humanos ha surgido como una forma de mejorar los sistemas de reconocimiento automático en la cara (Best-Rowden L., et al, 2014; Kumar N., et al, 2011, Best-Rowden L., et al, 2014), la marcha (Reid D., et al, 2014; Martinho, D., et al, 2014) o la evaluación de seguridad (Kumar N., et al, 2011).

¿Qué tan bueno es un profano en el reconocimiento de la autenticidad de una firma de consulta?: El uso de nuestra firma en nuestra vida cotidiana nos convierte en buenos detectores de falsificaciones de imitación (hechas por otros) de nuestra propia firma. Somos capaces de diferenciar nuestra variabilidad intrapersonal de la variabilidad de un falsificador (nuestros modelos cerebrales están entrenados con cientos de muestras hechas durante años de práctica). Esta capacidad puede extenderse a las firmas de otras personas, pero se espera una disminución del rendimiento debido a la falta de información sobre la variabilidad del propietario. Además, debemos considerar la motivación como un factor importante a considerar. Sin capacitación específica y teniendo en cuenta que el reconocimiento de la firma no es la principal tarea de los profanos, su desempeño es una pregunta abierta. La Figura 1 intenta ilustrar las dificultades relacionadas con esta tarea.

Figura 1.10: Firma genuina (de dos diferentes firmantes) y falsificaciones (hechas por otras personas después de practicar durante 2 minutos). ¿Qué firmas son genuinas?. (Morocho D., et al, 2016)

¿Cómo pueden ayudar estos humanos no expertos a los sistemas automáticos?: En la mayoría de las aplicaciones de verificación de firmas, usualmente los humanos (profanos) supervisan el proceso de firmas, pero sus responsabilidades son mayormente limitadas a garantizar una adquisición valida sin ninguna contribución a la autenticación, considerando que la autenticación de firma no es la tarea principal de su trabajo.

Esquemas asistidos por humanos en biometría toman ventaja de las habilidades humanas y las capacidades de sistemas automatizados (Kumar N., et al, 2011; Reid D., et al, 2014; Klare B.F., et al, 2014; Samangouei P., et al, 2016; Tome P., et al, 2014). Las intervenciones humanas en sistemas biométricos pueden ser hechos a diferentes niveles (ver Fig. 2) de acuerdo a diferentes tareas a ser realizadas: a nivel de imagen; a nivel de características ; a nivel de clasificación o emparejamiento ; y a nivel de decisión. En esta tesis, se enfocará en dos tipos específicos de intervenciones: (a) calificaciones humanas que miden la autenticidad percibida (intervención a nivel de clasificación) y (b) manual de

anotación de atributos (intervención a nivel de características) usado como entrada de un sistema de clasificación automático.

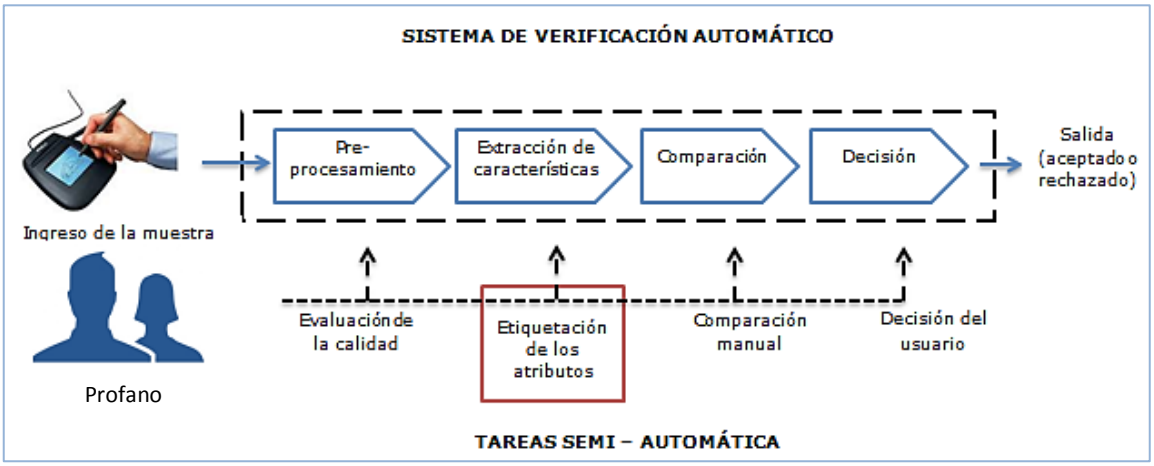


Figura 1.11: Esquema básico del reconocimiento de firmas asistidas por humanos. (Morales A., et al, 2017).

El estudio del rendimiento humano en aplicaciones biométricas no es nuevo, y ayuda a entender mejor el potencial y la capacidad de sistemas automáticos (Coetzer J., et al, 2006; Han H., et al, 2015). Las habilidades humanas son usadas comúnmente como punto de referencia para la evaluación de algoritmos automáticos (Coetzer J., et al, 2006; Phillips, P.J, et al., 2015).

Escenario de Aplicación: La presente tesis de investigación genera varias aplicaciones donde el humano entra a participar, ayudar y colaborar de forma oportuna en la verificación de firma del cliente a través de la toma de decisiones con ayuda de una interface de etiquetación de atributos característicos que permite verificar la identidad de la persona. Las figuras 1.12 y 1.13, muestran los procesos de diferentes escenarios de aplicación de la colaboración del humano en la vida real:

Figura 1.12: Proceso de entrega de paquetes

Figura 1.13: Proceso Pago de cheque bancario

1.5 Contribuciones de la tesis

Las contribuciones de la tesis se basa en el análisis del potencial del humano en el reconocimiento de firma manual basado en 2 procesos fundamentales: i) El reconocimiento de firmas a través de evaluaciones realizadas por humanos (workers) evaluadores sin experiencia en FDEs, mediante la metodología del Crowdsourcing utilizando la plataforma de Amazon Mechanical Turk, y ii) El Rendimiento y comportamiento el humano en el reconocimiento de firma a través de la etiquetación de diferentes tipos de atributos tales como absolutos, de medida y comparativos (Ejemplo: inclinación, distancia entre caracteres, entre otros), con la colaboración de un gran grupo de personas de la Universidad de las Fuerzas Armadas – ESPE, Ecuador.

Esta investigación permite establecer la base del rendimiento humano en tareas de reconocimiento de firmas, además se utiliza un novedoso sistema de verificación de firmas semiautomática basado en atributos obtenidos del análisis de personas no expertas en documentos forenses. Este trabajo impulsa la investigación sobre la capacidad de intervención humana para mejorar el rendimiento de los sistemas automáticos de reconocimiento de firmas.

La contribución de esta investigación es generar información relevante y robusta para mejorar los sistemas automáticos a través de las capacidades de los humanos que pueden ser utilizadas para mejorar el rendimiento de los sistemas automáticos de reconocimiento de firma. Para ello, se deben diseñar esquemas de interacción hombre-máquina dirigidos a explotar de forma eficiente estas capacidades.

1.6 Organización de la Tesis

La presente Tesis Doctoral se enfoca en el estudio de la problemática del reconocimiento de firma manuscrita en aspectos como: Variabilidad de las firmas, Usabilidad y Universalidad. Los principales objetivos de esta tesis doctoral son los siguientes:

- Revisar y estudiar el problema del reconocimiento de firma en sistemas semi-automáticos.
- Analizar la influencia de la asistencia del humano en los sistemas de reconocimiento automático de firma.
- Estudiar y analizar las medidas o acciones para ayudar a los sistemas de reconocimiento automático de firma.
- Analizar el comportamiento y rendimiento del humano a través de experimentos de crowdsourcing y etiquetación de atributos.

Esta disertación está estructurada de acuerdo con un tipo de complejo tradicional que incluye teoría de fondo, métodos prácticos y una serie de estudios experimentales independientes. La estructura del capítulo es la siguiente:

- **El Capítulo 1,** presenta temas relacionados con los rasgos biométricos, el reconocimiento de firma, y como el humano hay intervenido en los sistemas automáticos.

- **El Capítulo 2,** resume los trabajos relacionados que han motivado esta Tesis.

- **El Capítulo 3,** describe el rendimiento del humano en reconocimiento de firma manuscrita, mediante tareas inteligentes humanas (HIT) a través del crowdsourcing en la plataforma de Amazon Mechanical Turk.

- **El Capítulo 4,** estudia el Reconocimiento de firmas asistida por humanos, a través de etiquetación de atributos o características de la firma.

- **El Capítulo 5,** concluye esta Disertación. Se discuten los principales resultados y se proponen futuras áreas de investigación.

1.7 Contribuciones Científicas

En esta sección se presentan las contribuciones de investigación de esta tesis doctoral (ver Tabla 1.5):

Tabla 1.5: Resumen de las publicaciones en congresos y revista

Publicación de 7 artículos en Congresos Internacionales		
AÑO	**ARTÍCULO**	**PAÍS**
2016	D. Morocho, A. Morales, J. Fierrez and R. Tolosana. Signature recognition: establishing human baseline performance via crowdsourcing. Proc. 4th Int. Workshop on Biometrics and Forensics (IWBF), pp. 1-6, 2016.	CHIPRE
	D. Morocho, A. Morales, J. Fierrez and R. Vera-Rodriguez. Towards human-assisted signature recognition: improving biometric systems through attribute-based recognition. IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), pp. 1-6, 2016.	JAPÓN
	D. Morocho, M. Proaño, D. Alulema, A. Morales and J. Fierrez. Signature Recognition: Human performance analysis vs. automatic system and feature extraction via crowdsourcing. Proc. Mexican Conference on Pattern Recognition (MCPR), Springer International Publishing, pp. 324-334, 2016.	MÉXICO
	D. Morocho, J. Hernandez-Ortega, A. Morales, J. Fierrez, & J. Ortega-Garcia. On the evaluation of human ratings for signature recognition. IEEE International Carnahan Conference on Security Technology (ICCST), pp. 1-5, 2016.	ESTADOS UNIDOS
2017	D. Morocho, A. Morales, J. Fierrez, et al. . State of the art: Humans performance contributions in signature recognition via crowdsourcing and manual annotation. IEEE International Conference on eDemocracy & eGovernment (ICEDEG), pp. 221-225, 2017.	ECUADOR
	D. Morocho, A. Morales, J. Fierrez, & J. Ortega-Garcia. Humans in the Loop: Study of Semi-Automatic Signature Recognition Based on Attributes. IEEE International Carnahan Conference on Security Technology (ICCST), pp 1-5, 2017.	ESPAÑA
	D. Morocho, A. Morales, J. Fierrez and R. Vera-Rodriguez. Human-Assisted Signature Recognition based on Comparative Attributes. International Conference on Document Analysis and Recognition (ICDAR), vol. 8, pp 5-9, 2017	JAPÓN
Publicación de artículos de Revistas JCR Internacionales		
AÑO	**ARTÍCULO**	**Observación**
2017	A. Morales, D. Morocho, J. Fierrez, R. Vera-Rodríguez. Signature authentication based on human intervention: performance and complementarity with automatic systems. IET Biometrics, vol. 6(4), pp. 307-315, 2017.	Publicado
2018	D. Morocho, A. Morales, J. Fierrez and R. Vera-Rodríguez. Signature Recognition based on Comparative Attributes. IEEE Access.	En Preparación

Capítulo 2

2.-Trabajos Previos

Este capítulo se enfoca, como los humanos pueden ayudar a los sistemas biométricos. Además se presentan trabajos previos que demuestran que la ayuda del humano en el etiquetado de atributos mejora los sistemas automáticos. En el reconocimiento de firma, se presentan trabajos previos de la mejora el rendimiento de los forenses. Además se expone como un humano bien entrenado, es capaz de alcanzar tasas de reconocimiento elevadas.

2.1.- Analizando y explotando las capacidades humanas en el reconocimiento biométrico

El uso del etiquetado humano para mejorar el rendimiento de sistemas de reconocimiento automático no es nuevo. En campos como la huella dactilar, el trabajo de expertos forenses y sistemas automáticos coexiste desde hace más de 20 años. Más recientemente, se ha propuesto extender estos modelos semi-automáticos a otras biometrías como la cara (Han H., et al, 2015). Los humanos pueden obtener una gran variedad de información de una imagen del rostro, que incluye identidad, edad, sexo, raza y más características como se puede ver en la Figura 2.1. Las características específicas de identificación de las imágenes faciales se han explorado bien en la investigación de reconocimiento facial y en diversas aplicaciones donde permite generar investigación de comparación del rendimiento humano versus los sistemas automáticos.

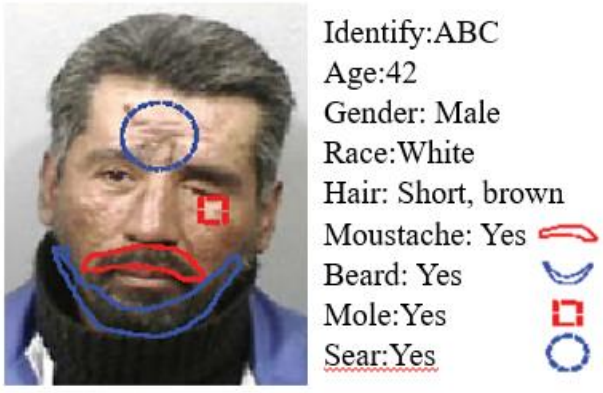


Figura. 2.1: Ejemplo de atributos extraídos de una imagen de la cara (Han H., et al, 2015)

El etiquetado de información demográfica a partir de imágenes faciales ha atraído a la comunidad científica (Han H., et al, 2015; Ricanek K., et al, 2006; Minear M., et al, 2004; Choi S., et al, 2011) . Así como el cotejo de la identidad, puede entrañar una mayor dificultad, el etiquetado de atributos tales como la

etnia, género, edad, están al alcance de cualquier persona sin formación específica en la tarea (Han H., et al, 2015; Semaj L., et al, 1981; Burt D. M., et al, 1995; Rhodes M., et al, 2009).

En (Han H., et al, 2015), se presenta una investigación sobre la capacidad de percepción humana para la etiquetación de atributos faciales a través del *crowdsourcing* utilizando un conjunto de bases de datos de imágenes faciales (1002 imágenes de FG-NET, 2000 imágenes de MORPH II y 4200 imágenes de la Oficina del Alguacil del Condado de Pinellas (PCSO). El objetivo del estudio era establecer un rendimiento de base o baseline de las capacidades humanas en esta tarea, y así poder comparar con el rendimiento de los sistemas automáticos. Esto permite una comparación de las habilidades de la máquina y el humano para estimar los datos demográficos. Además, se genera una línea de base, de la recopilación de estimaciones demográficas hechas por trabajadores (*workers*) del mundo utilizando el servicio de crowdsourcing de Amazon Mechanical Turk (MTurk). En la Figura 2.2, se muestran las Tareas de Inteligencia Humana (HIT),que consistieron en el etiquetado de: (a) descripción general del proceso de las tres HIT, (b) estimación de edad, (c) clasificación de género y (d) clasificación de raza, a partir de una imagen facial, con ayuda del *crowdsourcing*, donde las imágenes que se muestran a los trabajadores de MTurk son exactamente las mismas que procesaron en el algoritmo de (Han H., et al, 2015).

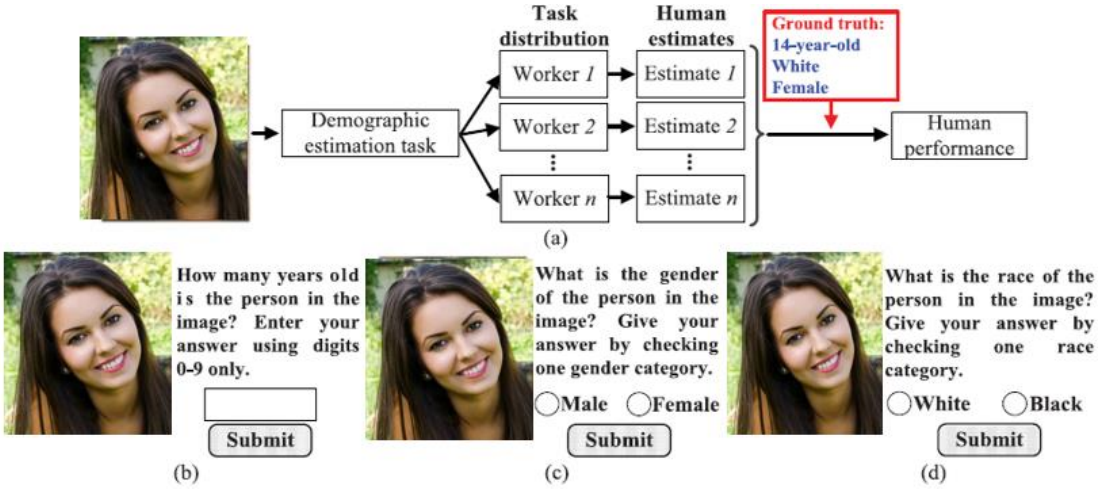


Figura 2.2: Estimación demográfica de una imagen facial (Han H., et al, 2015).

Otros investigadores han explorado el rendimiento de etiquetación de atributos por humanos en tareas como el reconocimiento facial (Best-Rowden L., et al, 2014; O'Toole A., et al, 2007; Kumar N., et al, 2011; Sun Y., et al, 2014; Taigman M., et al, 2014; O'Toole A., et al, 2002; Sinha P., et al, 2006). Estos trabajos han servido para establecer un rendimiento de base de la capacidad humana en una tarea compleja como el reconocimiento facial en condiciones no controladas. Para ello, se desarrollaron interfaces similares a los propuestos en (Best-Rowden L., et al, 2014), pero enfocados a estimar el rendimiento humano en tareas de reconocimiento (ver Figura 2.3). Además presenta un análisis de la precisión humana en el reconocimiento de caras sin restricciones de imágenes y

vídeos a través de *crowdsourcing* en *Amazon Mechanical Turk* Este trabajo presenta la actuación del humano aplicado en la base da datos de cara de LFW (Labeled Faces in the Wild) (Huang G. B., et al, 2007) y YTF (YouTube Faces) (L. Wolf L., et al, 2011), demostrando que los seres humanos son superiores a las máquinas, sobre todo cuando los vídeos contienen señales contextuales a más de la imagen de la cara. Además una fusión de reconocimiento de cara realizada por humanos y un comparador automático facial mejora el rendimiento en la etiquetación de atributos.

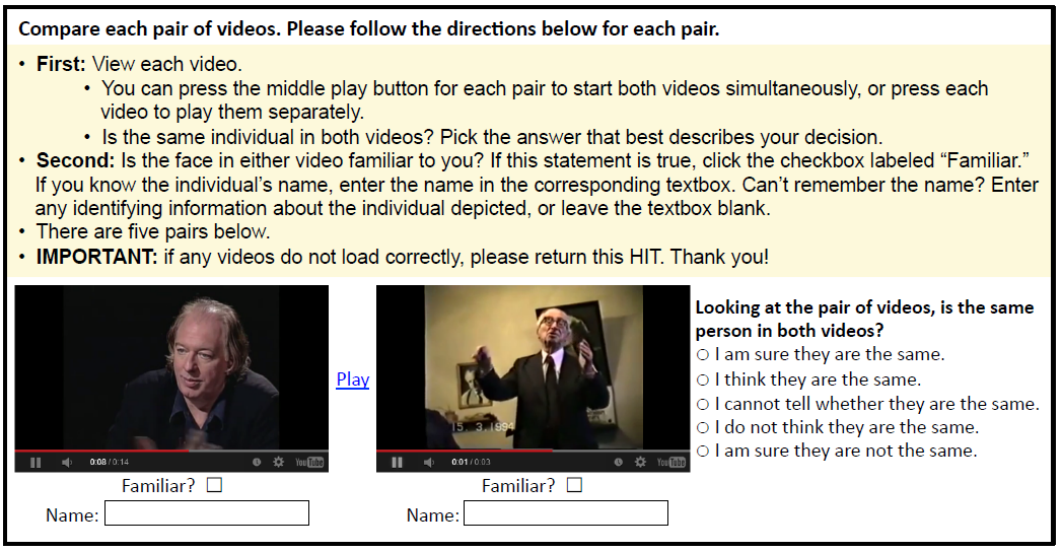


Figura 2.3: Interfaz para el reconocimiento de rostros humanos (Best-Rowden L., et al, 2014).

En (Best-Rowden L., et al, 2014), se manifiesta que los workers de los EEUU superan a los de la India en la precisión de reconocimiento facial del 1.1%. Esto manifiesta que los resultados de investigación aplicados por crowdsourcing, puede afectar a la precisión del rendimiento del humano debido a factores como la nacionalidad de los workers, motivación del pago a los workers, mayor cantidad de workers, calidad de datos, y otros factores.

Los resultados del rendimiento humano aplicados a la base de datos LFW, presentan que Best-Rowden L., et al, tiene 98.3% de precisión y es ligeramente menor a Kumar N., et al, que tiene un 99.2% de precisión, los mismos que son comparados con Sun Y., et al, y Taigman M., et al, que presentan precisiones del 97%, donde se refleja una mejora en el rendimiento del humano.

Investigadores como (O'Toole et al., 2017; Kumar et al., 2011; Sinha P., et al, 2006), estudian el rendimiento del humano frente al algoritmo de verificación facial, donde se ha demostrado que las señales dinámicas (imágenes de movimiento) mejoran la precisión humana en tareas de reconocimiento facial. Además imágenes estáticas que involucran rostros familiares, podrían mejorar el rendimiento en el reconocimiento. Además de la biometría facial, los investigadores han intentado explotar las etiquetas humanas en otros rasgos biométricos.

En (Jaha E., et al, 2014), los autores muestran una investigación relacionada con la biometría *soft* basada en atributos para la descripción corporal de una persona. Los humanos por naturaleza pueden describir a una persona por sus características físicas y de vestimenta. Además, muestra cómo los rasgos de la ropa pueden ser explotados para fines de identificación, a través de atributos semánticos. La biometría soft es forma de rasgo biométrico para la identificación de personas, donde utiliza descripciones humanas convencionales y las traduce a las formas biométricas de la máquina. Las técnicas biométricas soft dependen principalmente de la definición de una serie de atributos semánticos y la asignación de un conjunto de etiquetas descriptivas (rasgos) para cada atributo. Un atributo semántico puede ser cualquier propiedad observable que tenga un nombre designado o descripción por parte de los humanos. Dichos atributos pueden ser atributos binarios asociados con rasgos categóricos o atributos relativos, que pueden asociarse con etiquetas categóricas o comparativas.

Existen otros autores que realizan su trabajo de investigación relacionado con la biometría soft relacionado con la forma de vestir, el tipo de ropa, colores de ropa, la categoría de la ropa, estilo de vestimenta, etc. (Liu S., et al, 2012; Chen H, et al. , 2012; Zhu J., et al, 2013; Bossard L., et al, 2013; Vaquero D. A., et al, 2009), con el objetivo de identificar atributos para ser etiquetados y puedan identificar a la persona por su forma de vestir. En la vida diaria, las personas usan una ropa representativa con el objetivo de identificarse, y poder ser visto a distancia o cuando no se puede visualizar su cara. Se ha demostrado que los atributos de la ropa están naturalmente correlacionados y son mutuamente dependientes el uno del otro, esto puede explotarse al componer un rasgo biométrico. La Figura 2.4, expone un caso práctico de la aplicación, donde en la imagen están un grupo de personas escandalosas y destaca a un sospechoso con cara y cabeza cubierta, pero no se observan rasgos soft, excepto los atributos de la ropa. La CCTV publica una foto del posible sospechoso más buscado y que viste con el mismo estilo y color de ropa, lo que evidencia un vínculo de identificación con el sospechoso. La imagen proporciona un ejemplo real de cómo los atributos de la ropa podrían ser beneficiosos en la identificación y también demuestra que, en algunos casos, los atributos de la ropa pueden ser los únicos rasgos soft observables que se explotan.

Figura 2.4: Imagen que resalta un sospechoso con la cara cubierta y la ropa distintiva. (Jaha E., et al, 2014)

Las pruebas realizadas en (Jaha E., et al, 2014), se utiliza la base de datos de (Shutler J., et al, 2002), que es una base de datos estándar utilizados para esta investigación, y que comprende de un subconjunto de imágenes fijas de vista frontal y lateral del cuerpo entero de las personas, a través de una interface para anotación de vestimenta basado en la web para obtener etiquetas de ropa y comparaciones, como se muestra en la Figura 2.5. La interface gráfica permite obtener etiquetas categóricas y comparativas.

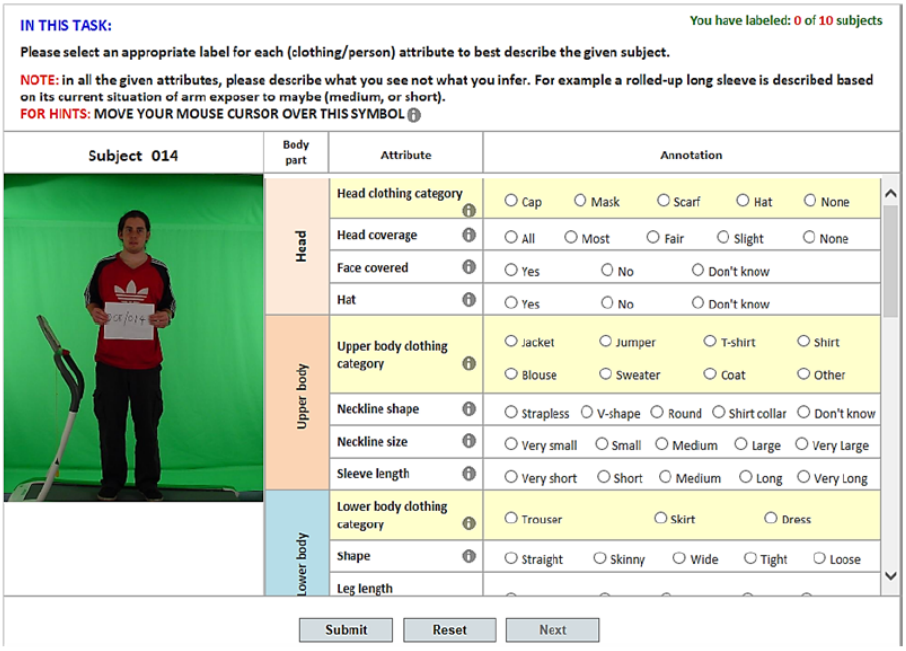


Figura 2.5: Sitio web desarrollado para obtener datos de anotación (Jaha E., et al, 2014)

En (Jaha E., et al, 2014), se concluye que los atributos semánticos a través de características categóricas y comparativas influyen en la descripción de la persona para ser identificada. Además con las características de la ropa se pueden utilizar para transmitir descripciones afectivas, a través de combinaciones de biometría soft. Una buena correlación entre dos atributos permite generar la predicción de un nuevo atributo. Los resultados obtenidos del rendimiento, al utilizar atributos semánticos de ropa, indican una motivación para explotar los rasgos derivados de la vestimenta, que permita mejorar la identificación de la persona.

En (Martinho-Corbishley D., et al, 2018), el reconocimiento basada en atributos de imágenes de vigilancia de los humanos es ampliamente estudiado para la re-identificación de las personas. Además, el crowdsourcing facilita una multitud de etiquetas y generar eficientes etiquetas refinadas como un medio para describir la variación visual percibida, la ambigüedad y la incertidumbre en imágenes desafiantes. Además la incorporación de técnicas de aprendizaje no supervisadas con reconocimiento de imágenes supervisado de vanguardia no solo supera los enfoques convencionales, sino que excede el máximo rendimiento

de recuperación posible de las descripciones binarias de género y edad con etiquetas refinadas en un escenario de múltiples tomas.

En (Jaha E.S., et al, 2016), expone que la biometría soft, avanza cada vez más el interés de la investigación. En la vida cotidiana, varios incidentes y escenarios forenses destacan la utilidad y la capacidad de la información de identidad que se puede deducir de la ropa. Los atributos semánticos de la ropa se han introducido recientemente como una nueva forma de biometría soft. Aunque los rasgos de la indumentaria pueden describirse y compararse de forma natural por los humanos para un uso operable y exitoso, es conveniente explotar la visión de la computadora para enriquecer las descripciones de la indumentaria como información más objetiva y discriminatoria. Además propone un conjunto de atributos soft de ropa, utilizando pequeños grupos de etiquetas semánticas de alto nivel, y extraídos automáticamente mediante técnicas de visión por computadora. Los rasgos soft de ropa categóricos y comparativos se derivan y se usan para la identificación / re-identificación.

En (Jaha E.S., et al, 2016), propone atributos de alto nivel de ropa que están más cerca de la percepción humana, la comprensión y el juicio; más genérico y menos detallado y también más separable, dejando un espacio más pequeño para la ambigüedad y la falta de coincidencia. Los atributos también son menos sensibles a los cambios en la iluminación, el punto de vista y la postura; y puede usarse en biometría, especialmente en vigilancia, mientras que cada rasgo único puede tener sus propios desafíos, que están relacionados con él pero no con los demás. Se emplean dos grupos diferentes de biometría soft de ropa con el propósito de identificación y recuperación de personas. El primer grupo comprende rasgos derivados manualmente a través de anotaciones basadas en humanos, mientras que el segundo grupo comprende rasgos derivados automáticamente a través de técnicas de visión por computadora. Los atributos soft de ropa pueden ser etiquetados manualmente por humanos para la identificación y re-identificación de personas.

2.2.- Reconocimiento de firma: expertos forenses

El uso de etiquetas humanas para mejorar el rendimiento en sistemas biométricos ha sido probado con éxito en rasgos como la huella dactilar, cara y cuerpo. Antes de saber si los humanos pueden ayudar a mejorar los sistemas de reconocimiento de firma, es necesario analizar la literatura para conocer cuál es el desempeño humano en una tarea como esta.

En (Coetzer J., et al, 2006), se realiza el experimento de verificación de firma por el humano sin experiencia en FDEs (prueba subjetiva realizada por el profano). Cada verificador humano no debe reflexionar sobre la toma de decisión, a fin de simular lo que probablemente haría un empleado del banco. El resultado de la verificación de firmas realizado por las personas fue de 3.5 a 4.7 segundos por firma. Las pruebas de verificación por sistemas automáticos (a través de

HMM), se utilizaron con las mismas firmas de entrenamiento y prueba que la verificación humana. La figura 2.6, muestra, los resultados de tasas de error (FRR vs FAR) de la verificación de firma, para los 22 verificadores humanos indicados mediante círculos.

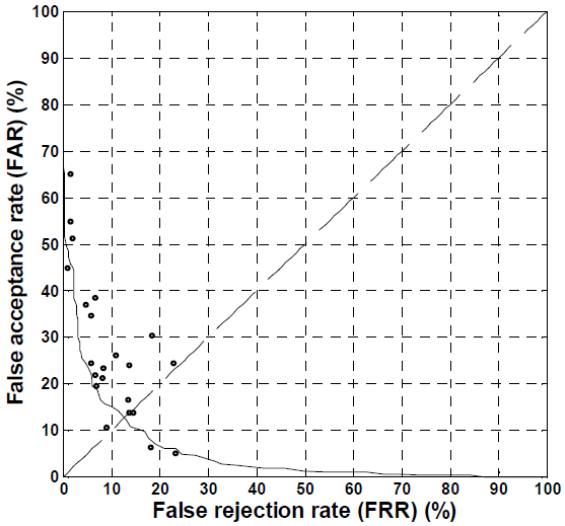


Figura 2.6: Curva ROC (FRR vs FAR) del sistema basado en HMM (Coetzer J., et al, 2006).

La Figura 2.6, muestra claramente que solo 4 verificadores humanos obtuvieron tasas de acierto superior al sistema basado en HMM, y de ellos, solo uno se desempeñó significativamente mejor. Es importante tener en cuenta que los círculos (verificación manual) más alejados de la diagonal (FAR = FRR) son menos significativos que los círculos más cercanos a la diagonal. Después de todo, un verificador humano puede garantizar que rinda tan bien como el sistema simplemente aceptando o rechazando todas las firmas. Además se puede observar que un verificador humano, con una FRR de 9.0% y una FAR de 10.5%, tuvo un rendimiento significativamente mejor que el sistema basado en HMM. La tasa de error igual (EER) para el sistema basado en HMM es del 12.6%. Pero dado el reducido número de muestras y verificadores, los resultados deben ser tomados con precaución.

En (Malik M. I., et al, 2013), se enfoca en comparar el desempeño de Examinadores de Escritura Forense (FHE) y sistemas automáticos, aplicados en la misma base de datos con el objetivo de realizar un análisis comparativo de los dos sobre la base de la precisión y la tasa de error. Los Examinadores de Escritura Forense, hacen un uso muy limitado de herramientas automatizadas, destacando, CEDAR-FOX (Srihari S. N., et al, 2003), FISH (Philipp M., et al, 1996), Proyecto WANDA (Franke K., et al, 2004). Si un sistema encuentra suficiente evidencia auténtica del vector de características de la firma cuestionada, toma la decisión de que la firma es genuina; caso contrario, toma la decisión que la firma es falsificada. En cambio los FHE toman la verificación de la firma como un problema de clasificación de clase múltiple, donde a más de las firmas genuinas y falsificadas, también buscan la posibilidad de firmas encubiertas.

Los resultados de varios sistemas de verificación de firmas de última generación, son analizados en los concursos de verificación de firmas, 4NSigComp2010 y 4NSigComp2012 organizados con la 12th y 13th Conferencias Internacionales sobre Fronteras en Reconocimiento de Escritura a Mano (ICFHR). En la competencia 4NSigComp2010 se evaluaron siete sistemas y en la competencia 4NSigComp2012 se evaluaron cinco sistemas. Más detalle de los sistemas de verificación automática en (Liwicki M., et al, 2010).

La precisión y la tasa de error son los parámetros de comparación entre los sistemas automatizados versus los FHE's, aplicando condiciones similares de evaluación (imágenes de firmas escaneadas para los sistemas y fotocopias de firmas para expertos humanos). La evaluación de las opiniones de FHE ha sido realizada por dos expertos forenses, validando sus opiniones y participando en las pruebas de competencia, donde es la única forma en que los FHE pueden verificar sus opiniones con puntajes reales (Ver Tabla 2.1). La metodología, las evaluaciones y el análisis de datos obtenidos se pueden ver con profundidad en (Malik M. I., et al, 2013).

Tabla 2.1. Resultados de las pruebas de competencia de FHE de diferentes años. Firmas: G: genuino, D: encubierta, y F: falsificadas (Malik M. I., et al, 2013).

Año / Resultado	2001			2004		
	G	D	F	G	D	F
Correcto	1628	571	2840	587	69	343
Errores	30	461	265	1	13	9
Poco Concluyente	105	895	3455	9	78	488
Sin Opinión	1763	1927	6560	1000	160	840
Año / Resultado	**2005**			**2006**		
	G	**D**	**F**	**G**	**D**	**F**
Correcto	587	73	1263	93	10	1151
Errores	1	52	174	0	111	113
Poco Concluyente	32	154	764	0	96	1526
Sin Opinión	620	279	2201	93	217	2790

La Tabla 2.2, muestra el promedio y el mejor de los resultados de evaluación entre el humano y los sistemas automáticos, en base a la precisión. El rendimiento, en muchos casos los sistemas automáticos están al mismo nivel de los FHE. Cabe recalcar que hay casos donde los sistemas automáticos al igual que los FHE son mejores en diferentes datos. Los FHE y los sistemas automáticos tiene dificultades para clasificar correctamente las firmas falsificadas, debido a la limitada disponibilidad de datos de entrenamiento falsificados.

Tabla 2.2: Resultados de evaluación: Humano vs. Sistema automático (Malik M. I., et al, 2013).

Datos del año	Precisión (%)			
	Promedio Humano	Promedio Sistema Automático	Mejor Humano	Mejor Sistema Automático
2001	44.8	70.8	100	93.6
2004	66.2	70.4	97	87
2005	62	59.8	100	68
2006	38.8	71.7	91	92

En (Malik M. I., et al, 2013BBBB), se presenta un novedoso sistema automático basado en características locales para la verificación de firmas forenses. Se presenta un escenario de la comparación de rendimiento de varios sistemas automatizados para la tarea de verificación de firmas, donde tienen que calificar la probabilidad de autoría y no autoría de firmas y detectar falsificaciones hábiles (firmas simuladas y encubiertas) a partir de firmas genuinas de un escritor de referencia, donde los resultados se comparan con las opiniones de los FHE sobre las mismas tareas. El sistema propuesto alcanza una tasa de error igual de 3.36% en la clasificación de firmas encubiertas y genuinas, y compara el rendimiento del sistema propuesto con varios sistemas de verificación de firmas de última generación utilizando la misma base de datos de competencia 4NSigComp20103. Además se presenta una comparación del rendimiento del sistema propuesto con los examinadores de escritura forense (FHE), y se destaca el potencial del sistema propuesto ya que permite ayudar a los humanos a resolver casos de verificación de firmas forenses del mundo real, donde los FHE están interesados en clasificar firmas genuinas, disfrazadas y falsificadas al mismo tiempo.

Los protocolos y metodologías de experimentación para la comparación entre FHEs y Sistemas automáticos se pueden ver en detalle en (Malik M. I., et al, 2013). Esto se traduce en una tasa de error del 15.2% en las decisiones (Precisión del 84.8%) al ignorar los casos que no fueron concluyentes (ver Tabla 2.3).

Tabla 2.3: Resultados de las pruebas de competencia realizadas con FHEs (Malik M. I., et al, 2013)

Clasificación	Genuino	Encubierta	Falsificada	Total
Correctamente Clasificado	93	10	1151	1254
Mal Clasificado(Error)	0	111	113	224
Reporte no concluyente	0	96	1526	1622
Total Opiniones de Autoría	93	217	2790	3100

La Tabla 2.4, proporciona los resultados comparativos generales de esta comparación Hombre vs. Sistemas automáticos, en términos de la precisión.

Tabla 2.4: Resultados comparativos generales (Malik M. I., et al, 2013).

Precisión %			
Promedio Humano	Promedio Sistema Automático	El Mejor Humano	El Mejor Sistema Automático
38.8	71.7	91	92.8

En (Liwicki M., et al, 2010), presentan una comparación entre las opiniones de FHEs sobre la autoría de firmas y los desempeños de sistemas para detectar falsificaciones hábiles (firmas simuladas y encubiertas) a partir de firmas genuinas de un escritor de referencia. La verificación de la firma forense se realiza por comparación visual por FHE. La autenticidad de la firma cuestionada se estima al ponderar las similitudes y diferencias particulares observadas entre las características de la firma cuestionada y las características de varias firmas

conocidas de un escritor de referencia. La interpretación de las similitudes o diferencias observadas en el análisis de firmas no es tan directa como en otras disciplinas forenses como el ADN o la evidencia de huellas dactilares, porque las firmas son producto de un proceso conductual que puede ser manipulado por el escritor de referencia o por una persona que no sea el escritor de referencia.

En la investigación de verificación de firmas, una coincidencia 100% perfecta no necesariamente es compatible con la Hipótesis H1: La firma cuestionada es una firma auténtica utilizada normalmente por el escritor de referencia, porque una coincidencia perfecta puede ocurrir si se traza una firma. Además, las diferencias entre las firmas no necesariamente respaldan la Hipótesis H2: La firma cuestionada es el producto de un proceso de falsificación (a: es simulado por otro escritor que no sea el escritor de referencia; b: está encubierto por el escritor de referencia), ya que el escritor de referencia puede poner cambios leves en una imagen de la firma cuando disfraza su firma con el propósito de denegación, o puede ocurrir debido a una variación dentro del escritor (Liwicki M., et al, 2010).

La verificación de la firma forense se realiza de una manera altamente subjetiva, la disciplina necesita una base científica y objetiva. El uso de herramientas automáticas de verificación de firmas puede ser objetiva la opinión de los FHE sobre la autenticidad de una firma cuestionada. Es de conocimiento que los algoritmos de verificación de firmas no son ampliamente aceptados por los FHE (Liwicki M., et al, 2010).

Para la competencia se utilizaron 6 sistemas: Grupo de reconocimiento biométrico-ATVS EPS–UAM, LISIC, NifiSoft, Parascript, Sabanci, Anónimo. Los protocolos y el corpus para las pruebas realizadas pueden ser revisadas en (Liwicki M., et al, 2010). El objetivo de los experimentos es comparar el rendimiento de los sistemas automatizados con los pronunciamientos de expertos forenses de escritura a mano (FHE). Es necesario considerar que es complejo calificar el EER de los expertos humanos, ya que no existe un umbral que pueda ser equilibrado. En cuanto a la evaluación de los sistemas automáticos, se obtuvieron resultados bastante buenos, donde se rescata la importancia que el rendimiento de los sistemas automatizados no está tan lejos de las decisiones humanas. En un total de 3100 opiniones de autoría de firma. De estas opiniones, 1254 (40.5%) fueron correctas, 224 (7.2%) fueron erróneas y 1622 (52.3%) fueron no concluyentes. Esto se traduce en una tasa de error del 15,2% en las decisiones (precisión del 84,8%), ver tabla 2.5.

Tabla 2.5: Resultados de Opiniones de FHEs (Liwicki M., et al, 2010)

Resultados	Genuina	Encubierta	Simuladas
Correctas	93	10	1151
Errónea	0	111	113
No concluyente	0	96	1526

2.2.1.- Resumen del estado del arte

El análisis del estado del arte en el campo de la biometría arroja los aportes más importantes dentro del reconocimiento de firma a través de sistemas automáticos y Expertos Forense en Escritura (FHE). Ver Tabla 2.5.

Tabla 2.6: Resumen del estado del arte

Referencia	Características	Rendimiento
J. Coetzer, B.M. Herbst, J.A. Du Preez. Off-line signature verification: A comparison between human and machine performance. Proc. 10th Int. Workshop on Frontiers in Handwriting Recognition, La Baule, France, pp. 481-485, 2006.	**Evaluadores:** 22 Humanos y Sistema automático (HMM). **Base de Datos:** Dolfing's **Usuarios:** 51 **Firmas:** 15 **Total firmas:** 765 firmas de prueba **Firmas genuinas:** 432 **Firmas falsificadas:** 333	**Mejor Rendimiento:** Humanos FRR= 9.0% FAR=10.5% Sistema automático EER=12.6%
M. I. Malik, M. Liwicki, A. Dengel, and B. Found. Man vs. Machine: A Comparative Analysis for Forensic Signature Verification. Proc. of the 16th International Graphonomics Society Conference, pp. 9–13, 2013	**Base de datos de firma:** La Trobe, recopilados bajo la supervisión de Bryan Found y Doug Rogers en los años 2001, 2002, 2004, 2005 y 2006 (C. Bird, 2007). Las imágenes se escanearon a una resolución de 600 ppp y se recortaron en el Instituto Forense de los Países Bajos para este estudio. En Malik M., et al, 2013,se presenta un desglose detallado de los datos utilizados.	**Mejor Precisión:** Base de Datos año 2001 Humano= 100% Sistema Automático=93.6% Base de Datos año 2004 Humano= 97% Sistema Automático=87% Base de Datos año 2005 Humano=100% Sistema Automático=68% Base de Datos año 2006 Humano=91% Sistema Automático=92%
M. I. Malik, M. Liwicki, A. Dengel. Part-based automatic system in comparison to human experts for forensic signature verification. Proc. Int. Conf. on Document Analysis and Recognition, Washington DC, USA, pp. 872–876, 2013	Las firmas son recopiladas por FHE y escaneadas a una resolución de 600 ppp. Contiene 125 firmas. Hay 25 firmas de referencia por el mismo firmante y 100 firmas cuestionadas por varios firmantes. Las 100 firmas cuestionadas comprenden 3 firmas genuinas escritas por el firmante de referencia en su estilo de firma normal y 7 firmas encubiertas escritas por el firmante de referencia.	**Precisión Promedio:** Humano=38.8% Sistema Automático=71.7% **Mejor Precisión:** Humano=91.0% Sistema Automático=92.8%
M. Liwicki, C. E. van den Heuvel, B. Found, and M. I. Malik. Forensic signature verification competition 4NSigComp2010 - detection of simulated and disguised signatures," in ICFHR, pp. 715–720, 2010	**Base de datos de firma:** La Trobe. Firmas de entrenamiento:209 firmas 9 firmas de referencia del mismo firmante y 200 firmas cuestionadas (76 firmas genuinas del mismo firmante; 104 firmas simuladas; 20 firmas encubiertas del firmante de referencia). Firmas de prueba: 125 firmas 25 firmas de referencia del mismo firmante y 100 firmas cuestionadas (3 firmas genuinas firmadas por el mismo firmante; 90 firmas simuladas; 7 firmas encubiertas del firmante de referencia.	En un total de 3100 opiniones de autoría de firma: 1254 (40.5%) fueron correctas, 224 (7.2%) fueron engañosas y 1622 (52.3%) no fueron concluyentes. Esto se traduce en una tasa de error del 15,2% en las decisiones (precisión del 84,8%).

El análisis del estado del arte de los trabajos previos, relacionados con el rendimiento del humano en el reconocimiento de firma a través del crowdsourcing, la plataforma más utilizada es Amazon Mechanical Turk (Mturk). Esta plataforma es la utilizada en el desarrollo de los experimentos de investigación de esta tesis.

Capítulo 3

3.- Estableciendo el rendimiento humano en reconocimiento de firma manuscrita

En la actualidad existe una gran variedad de aplicaciones en las que la firma es el rasgo utilizado para acreditar la dientidad de un individuo (e.j. banca, comercio, salud, educación). En la mayoría de estas aplicaciones, los humanos generalmente realizan la supervisión del proceso de firma, sin tener la responsabilidad del proceso de reconocimiento. El cotejo de firmas se suele realizar de forma no automatizada en caso de requerimento (e.j. denuncia de falsificación). Este cotejo lo realizan expertos forenses. Durante la adquisición, la persona que supervisa se encarga normalmente solo de comprobar que el sujeto ha firmado. La mayoría de los humanos supervisores no tienen la experiencia específica de los expertos forenses (FDE). En esta tesis nos referimos a estos usuarios sin experiencia profesional en reconocimiento de firma como **profanos**. Uno de los objetivos de esta tesis es evaluar el potencial rendimiento de estos profanos en el cotejo de firmas, para así poder mejorar sistemas semi-automáticos de reconocimiento.

Para recopilar datos sobre el desempeño de humanos en la tarea de reconocimiento de firma se utilizarán herramientas de crowdsourcing que pueden ser implementados en plataforma de la web. La herramienta Amazon Mechanical Turk (MTurk) está muy extendida en la literatura relacionada (Martinho-Corbishley D., et al, 2015; Panjwani S., et al, 2014; Howe J., et al, 2006; Buhrmester M., et al, 2011; Morocho D., et al, 2016; Kittur E. H., et al, 2008; Morales A., et al, 2017; Morocho D., et al, 2016). El crowdsourcing que es una herramienta que utiliza una multitud de seres humanos para resolver diferentes tipos de problemas, encuestas y otras actividades. Esta herramienta ha sido utilizada en en el campo del reconocimiento biométrico incluyendo el reconocimiento facial (Best-Rowden L., et al, 2014), el reconocimiento de la marcha (Martinho-Corbishley D., et al, 2015) y seguridad biométrica (Panjwani S., et al, 2014). Esta herramienta será la utilizada en esta tesis.

En este capítulo se presenta un análisis del rendimiento humano en reconocimiento de firma manuscrita estableciendo una línea de base (baseline), vía crowdsourcing. Se plantean escenarios diferentes, a través de Tareas de Inteligencia Humana (HIT), desarrolladas en HTML, en la plataforma de Amazon Mechanical Turk (MTurk). Cada tarea intenta explorar qué información resulta más relevante para un ser humano a la hora de realziar un cotejo de firmas. Para la ejecución de las HITs se definen las siguientes condiciones:

- **Personas sin experiencia como Examinadores de Documentación Forense (FDE)**, durante esta tesis las personas que ayudan en el proceso de verificación y etiquetación de firmas son profanos en la materia, pero dentro de la plataforma de MTurk se los conoce como trabajadores o "workers". Se quiere establecer el rendimiento de una persona media sin formación específica en la tarea.
- **Tiempo y costo de ejecución de las diferentes HITs**, es el precio de pago y el tiempo necesario para ejecutar las instrucciones o tareas encomendadas en el reconocimiento de firma. El tiempo y costo tienen impacto en el rendimiento de los workers.
- **Planificación y desarrollo de las HITs**, una HIT está compuesta por funciones, instrucciones, y/o tareas que permiten ejecutar y ser evaluadas por los workers, por ejemplo: i) instrucciones para el workers, ii) conjunto de firmas etiquetadas y sin etiquetar, iii) opciones de respuestas, y iv) justificación de respuesta. Los escenarios de los diferentes experimentos incluyen comparaciones únicas entre una muestra genuina y una muestra no etiquetada basadas en imágenes, videos o secuencias de tiempo, y comparaciones con múltiples conjuntos de entrenamiento y pruebas.

En esta tesis, las bases de datos utilizadas en los diferentes experimentos de este capítulo son: BIOSECURE-DS2 (Ortega-Garcia J., et al, 2010) y BiosecurID (Fierrez J., et al, 2010). Las pruebas realizadas a través de MTurk vía crowdsourcing son ejecutas con 10, 60 y 400 workers, dependiendo de la tarea asignada y el objetivo de las mismas.

3.1.- Plataforma Amazon Mechanical Turk (MTurk)

Las tareas masivas (Crowdsourcing) asistidas por personas aprovechan las habilidades humanas y los beneficios de un muestreo mundial de datos a través del internet (Howe J., et al, 2006; Buhrmester M., et al, 2011; Kittur E. H., et al, 2008; Buhrmester M., et al, 2011).

Amazon Mechanical Turk (MTurk) es una plataforma en la web, perteneciente a Amazon, su funcionamiento se basa en el crowdsourcing y permite el desarrollo de múltiples tareas a gran escala y a bajo costo (Kittur E. H., et al, 2008; Amazon Mechanical Turk, 2016). La Figura 3.1, muestra un esquema de las 3 variables que intervienen en el proceso a través de Mturk, que son: 1) la persona o entidad que propone la tarea (Requester), 2) el desarrollo y despliegue de las Tareas de Inteligencia Humana (HIT), y 3) los participantes (workers), encargados de la realización efectiva de las tareas predefinidas (Martinho-Corbishley D., et al, 2015; Panjwani S., et al, 2014; Coetzer J., et al, 2006).

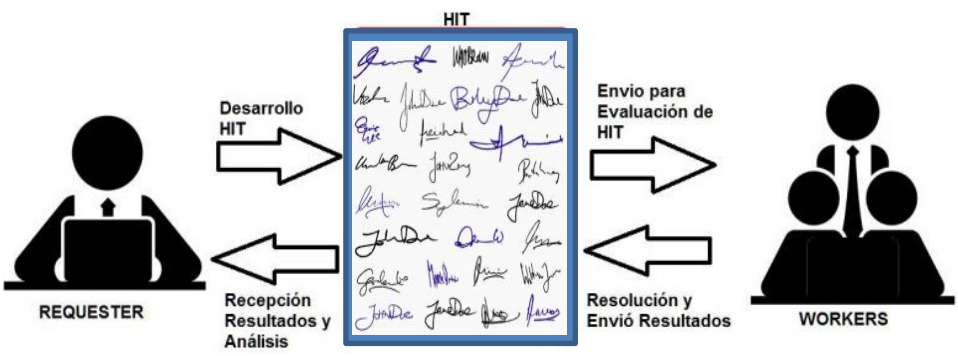


Figura 3.1: Proceso de crowdsourcing través de MTurk

Tareas de Inteligencia Humana (HIT): son la parte medular del proceso en MTurk y son desarrolladas por los Requesters (la persona o entidad que propone la tarea) y ejecutadas por los workers. La HIT, es una interface diseñada en HTML, que contiene parámetros como: las instrucciones (indicaciones), la tarea, el tiempo, y las opciones de respuestas (Ver Figura 3.2).

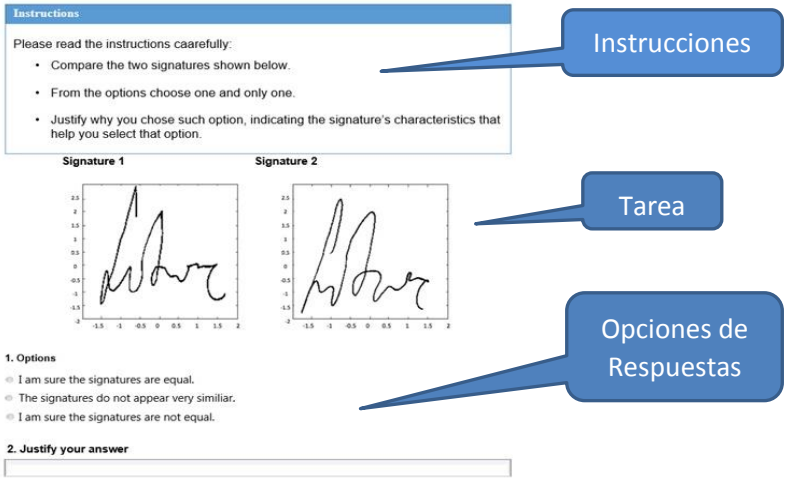


Figura 3.2: Tarea de Inteligencia Humana (Amazon Mechanical Turk, 2016)

Workers: son un grupo de personas que se encargan de la resolución de las HITs. Un worker debe hacer el mismo proceso del Requester para poder tener acceso a las HITs de MTurk y poder ejecutarlas (Ver Figura 3.3). MTurk realiza el pago por cada tarea realizada por el worker, y el costo por tarea está definido por el Requester. En la plataforma Mturk, existen 2 tipos de “workers”: el “worker” regular y el “worker” maestro. En la actualidad la plataforma MTurk cuenta con más de 500.00 personas en todo el mundo en aproximadamente 190 países.

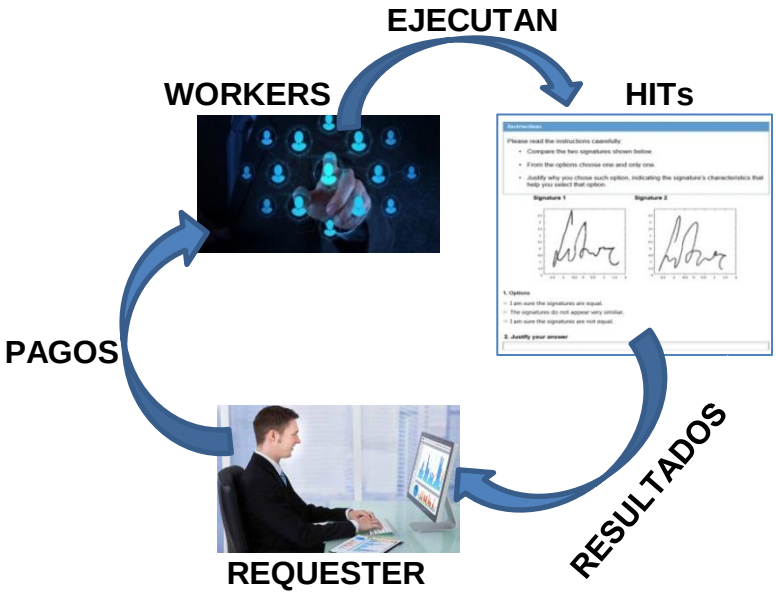


Figura 3.3: Proceso de intervención de un “worker”

3.2.- Verificación de firma a través de HITs vía crowdsourcing en MTurk

El objetivo de las HIT diseñadas en esta sección y enviadas a los “workers”, responden a una inquietud principal: **¿Qué tan buenos son los humanos (no FDE) verificando la autenticidad de una firma manuscrita?** Para responder esta inquietud se diseña 5 tareas diferentes, para ser realizadas por los workers.

La Figura 3.4, muestra el esquema del sistema crowdsourcing para establecer un “baseline” del rendimiento humano en reconocimiento de firma. El sistema se divide en Front-end que presenta las tareas a los workers y captura sus respuestas y Back-end, que comprende los procesos y algoritmos que se ejecutan en los servidores MTurk. Los “workers”, pueden ser diferentes en cada experimento. Es recomendable diseñar tareas simples para que se realicen en poco tiempo.

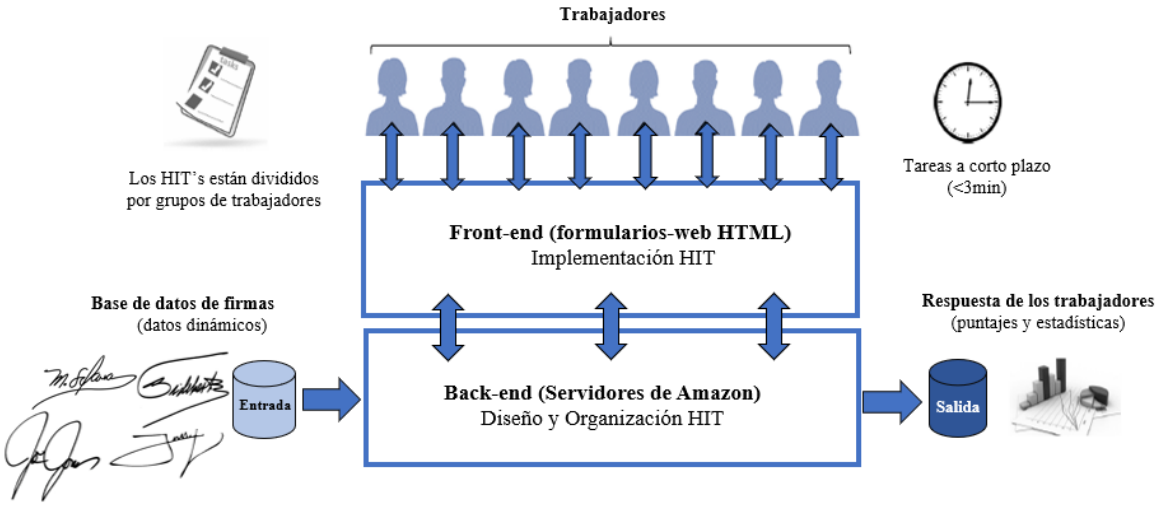


Figura 3.4: Esquema de crowdsourcing para establecer un "baseline" del rendimiento humano en reconocimiento de firma (Morocho D., et al, 2016).

Los aspectos de diseño de las HITs más relevantes son:

Calificación: representa el historial del "worker" a través de las calificaciones obtenidas en HITs anteriores. Existen 2 tipos de "workers". Para los experimentos se solicita workers de tipo regular y maestro. El "worker" regular es aquel worker que cumple con los requisitos básicos de rendimiento, y el "worker" maestro es un worker que ha demostrado un rendimiento superior en la realización de HITs en la plataforma MTurk.

Ubicación geográfica: representa la nacionalidad de los workers. Para los experimentos realizados en esta tesis, no se solicita ubicaciones geográficas específicas de los workers. Para nuestro caso se trabaja con participantes de Estados Unidos de América (47%), la India (34%) y de otras nacionalidades (19%).

Tiempo de ejecución de la HIT por worker: representa el tiempo disponible que tiene el worker para realizar la HIT. La complejidad de la HIT y la cantidad de datos tienen una relación directa con el tiempo que necesitan los workers. Se realizaron algunas pruebas iniciales con un pequeño grupo de participantes (workers), esto con el objetivo de determinar el tiempo necesario en tareas de reconocimiento de firma. Existen tareas que necesitan más tiempo que otras, debido a la complejidad de la firma. Considerando las tareas iniciales, se determina, que el tiempo necesario para realizar una HIT en reconocimiento de firma para cada worker son 15 segundos. Las diferentes interfaces gráficas implementadas para cada HIT se programan en lenguaje HTML, donde contienen los diferentes bloques de funcionalidad de la tarea inteligente humana, tales como: instrucciones, firmas, opciones, y justificación (Ver Figura 3.5):

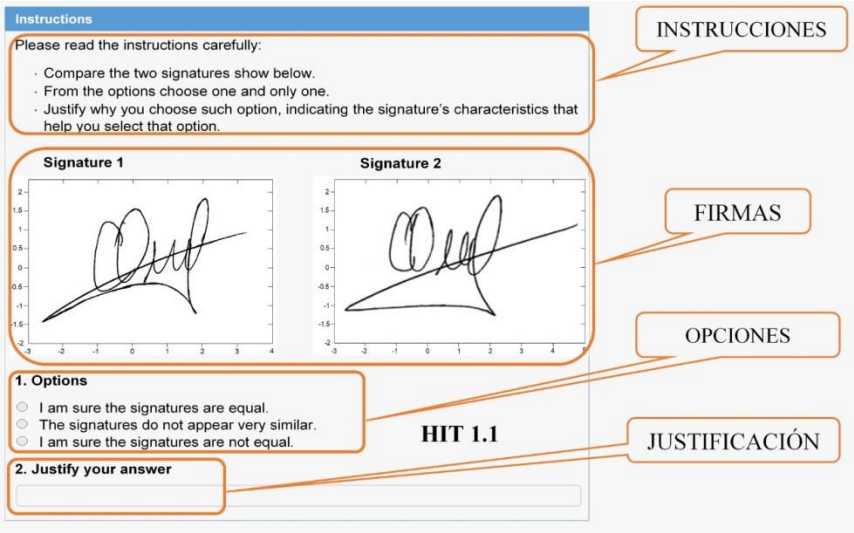


Figura 3.5: Interface desarrollado para establecer el baseline de reconocimiento de firma

La interface de usuario de la Figura 3.5, se diseña a partir de los siguientes campos:

Instrucciones: en esta área, se detallan el proceso y las pautas que debe seguir cada worker.

Firmas: en esta área se muestra un conjunto diferente de firmas etiquetadas (firmas genuinas) y sin etiquetar (muestras genuinas y falsificadas que los workers deben etiquetar). La información mostrada varía según la tarea, por ejemplo: imágenes (versiones estáticas de la firma), videos (que muestran la forma en que se realiza la firma) y secuencias de tiempo relacionadas con tres características de firma dinámica (eje x, eje y, y la presión). Las firmas mostradas a los workers se sintetizan en la pantalla según la información dinámica disponible en las bases de datos empleadas (BIOSECURE-DS2 y BiosecurID).

Opciones: esta área incluye la respuesta de los workers al HIT predefinido. En nuestras HIT, estas respuestas están relacionadas con el análisis de autenticidad (genuino o falsificado) realizado por el worker.

Justificación: en esta área, el worker debe indicar las razones por las cuales eligió las diferentes opciones. Esta información ayuda a comprender mejor los resultados obtenidos.

3.2.1.- Diseño de las HIT

Esta sección se enfoca en el análisis del rendimiento humano en reconocimiento de firma a través de pruebas de verificación de firmas genuinas versus falsificadas, con ayuda de imágenes, videos y características en tiempo de cada una de las firmas.

Pruebas Preliminares: MTurk es una plataforma que permite subir HITs, donde se necesita preparar y estructurar de forma eficiente todos los procesos de las HITs, por lo tanto es necesario realizar 4 pruebas preliminares, que permitan conocer el funcionamiento de la plataforma, interactuar con los workers y definir de forma eficiente las instrucciones y los datos de configuración de los experimentos. La realización de las pruebas preliminares permiten obtener el nivel de aceptación y comprensión de los workers y determinar el tiempo y costo de cada HIT propuesta. El porcentaje de aceptación es del 75.4%, y el tiempo aproximado es un minuto, además, esta pruebas permiten mejorar la interfaz y las instrucciones de la HIT, con el objetivo de mejorar la comprensión y facilitar su proceso de ejecución. Para determinar el análisis del rendimiento humano en reconocimiento de firma se realiza el diseño de 5 HITs, dividido en 3 experimentos:

Experimento No.1: Comparación uno a uno (Una firma de entrenamiento vs. Una firma de prueba): En este experimento se realizan 3 HITs diferentes (HIT1.1, HIT1.2, HIT1.3) (Morocho D., et al, 2016), donde los workers deben definir la autenticidad de una muestra dada utilizando como referencia solo una firma genuina. El objetivo de este experimento es analizar el rendimiento del humano en reconocimiento de firma de acuerdo a la diferente información disponible. Los datos mostrados en el experimento No.1 están diseñados de acuerdo a 3 HITs.

- HIT 1.1 se muestra solo la imagen de la firma.
- HIT1.2, se muestra la imagen y el video del proceso de firma dinámica.
- HIT1.3, se muestra la imagen y las secuencias de tiempo de la firma, que es el caso típico de los sistemas de reconocimiento dinámico en línea.

Estas tareas incluyen 12 firmantes diferentes (con 2 muestras genuinas y 1 falsificaciones por firmante) de la base de datos Biosecure-DS2 (Fierrez J., et al, 2010), donde el tiempo de ejecución de las HITs es de 2 minutos, a un costo de $0.04 por “worker” (60 workers). Las tablas 3.1a y 3.1b, muestra la descripción y la interface de cada HIT.

Tabla 3.1a: Descripción e interface de HIT 1.1 - HIT 1.2

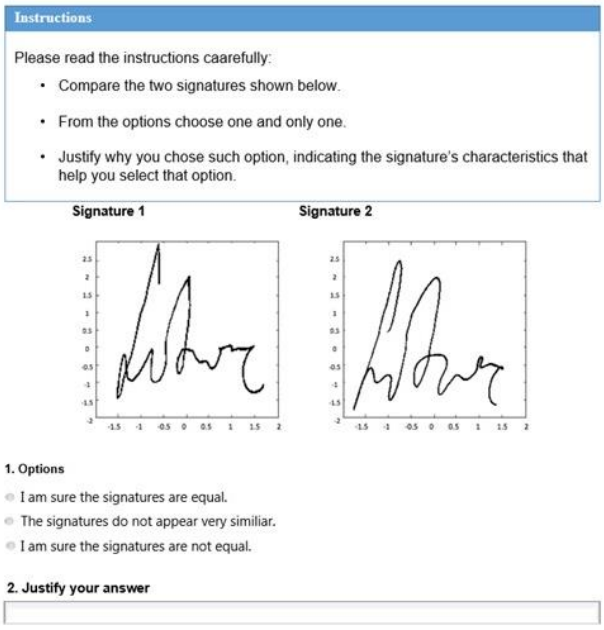
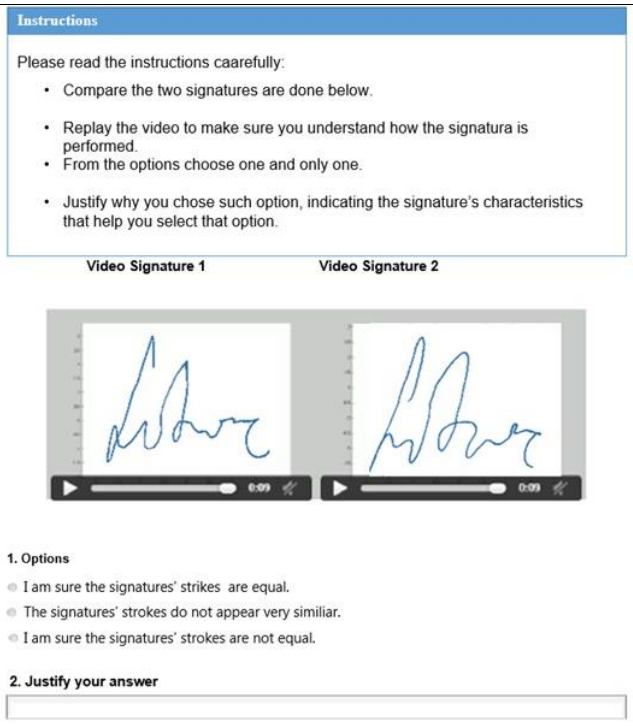
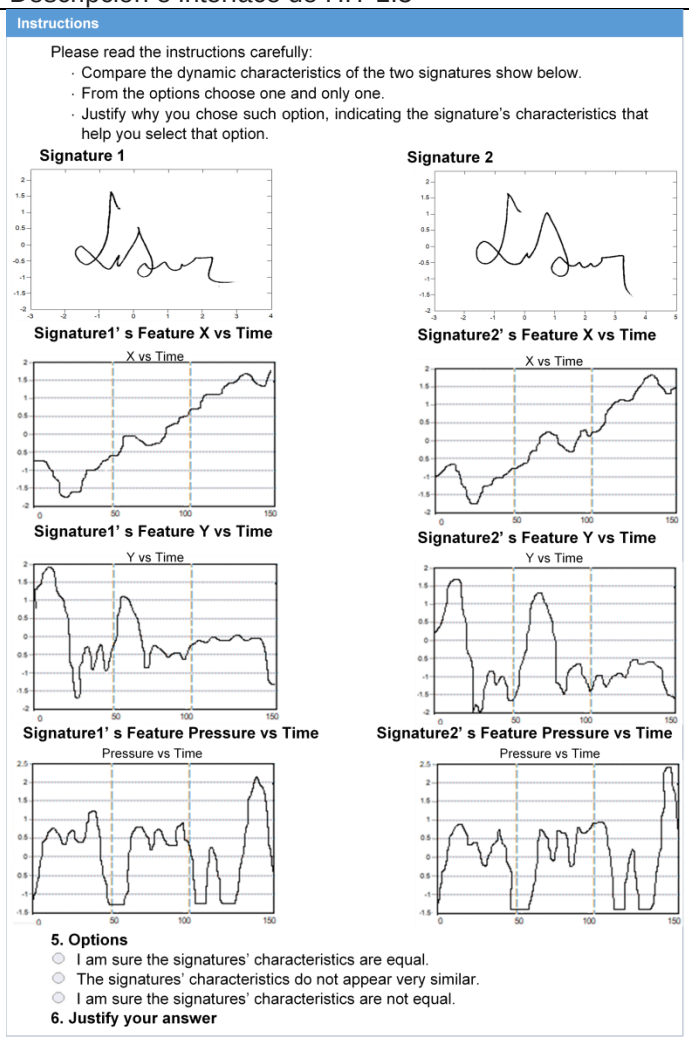
HIT 1.1: presenta la imagen de la firma de entrenamiento y la firma de prueba, con el objetivo de identificar la similitud entre firmas a través de información estática.	
HIT 1.2: presenta el video del trazo de la firma de entrenamiento y la firma de prueba, con el objetivo de identificar la similitud entre firmas a través de información dinámica.	

Tabla 3.1b: Descripción e interface de HIT 1.3

HIT 1.3: presenta la imagen de la firma de entrenamiento y la firma de prueba con sus respectivas características de posicionamiento en el eje x, e y, y la presión p, con el objetivo de identificar la similitud entre firmas a través de información estática.	

Experimento No.2: Comparación uno a muchos (Una firma de entrenamiento vs Ocho firmas de prueba): En este experimento se realiza la HIT 2 (Morocho D., et al, 2016), donde se muestran una imagen de firma genuina y ocho imágenes de firmas sin etiquetar (cinco genuinas y tres falsificadas). El objetivo de este experimento es analizar el rendimiento cuando los workers tienen información contextual. La información sobre el número de firmas genuinas y falsificadas no se facilita a los workers. Esta tarea incluye 6 firmantes diferentes (con 5 genuinos y 3 falsificaciones por firmante) de la base de datos Biosecure-DS2 (Ortega-Garcia J., et al, 2010). En este experimento el objetivo es evaluar el rendimiento de las respuestas rápidas inspiradas en escenarios operacionales reales en los que el worker debe proporcionar una respuesta en poco tiempo. El tiempo de ejecución de la HIT es de 1 minuto, a un costo de $0.02 por "worker" (30 workers), (ver Figura 3.6).

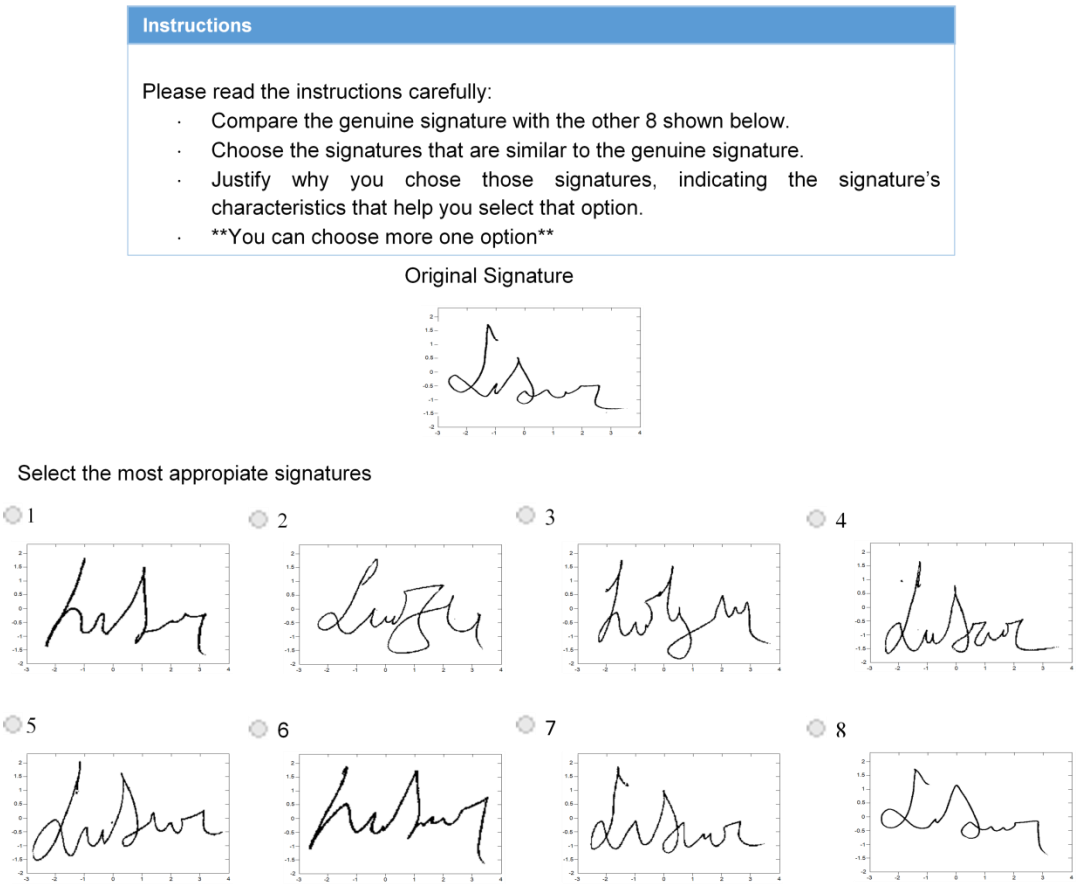


Figura 3.6: Interface de HIT 2

Experimento No.3: Comparación muchos a uno (Cuatro firmas de entrenamiento vs una firma de prueba): En este experimento se realiza la HIT 3 (Morocho D., et al, 2016), este experimento imita el protocolo de evaluación tradicional de los sistemas de reconocimiento automático de firmas en el que se compara una firma de evaluación con un conjunto de firmas de evidencia. En este caso, se muestra al worker una firma no etiquetada (genuina o falsificada) y cuatro firmas auténticas etiquetadas (Ver Figura 3.7). Para mejorar la información disponible en experimentos anteriores y adaptarlo al protocolo experimental habitual en sistemas automáticos, la respuesta del worker se proporciona como un valor de confianza entre 0 y 10, donde 0 significa "Estoy seguro de que es una firma falsificada" y 10 significa "Estoy seguro de que es una firma genuina". Esta tarea incluye 20 firmantes diferentes (5 firmas auténticas y 3 firmas falsificadas por firmante) de la base de datos de BiosecurID (Fierrez J., et al, 2010). Los 20 firmantes se dividen en 5 grupos de 4 firmantes cada uno (4 x 8 = 32 firmas) que se distribuyen por igual entre los workers (cada worker procesa 32 firmas). El tiempo de ejecución de la HIT es de 3 minutos/actividad, a un costo de $0.03 por worker/actividad (60 workers, 4 actividades).

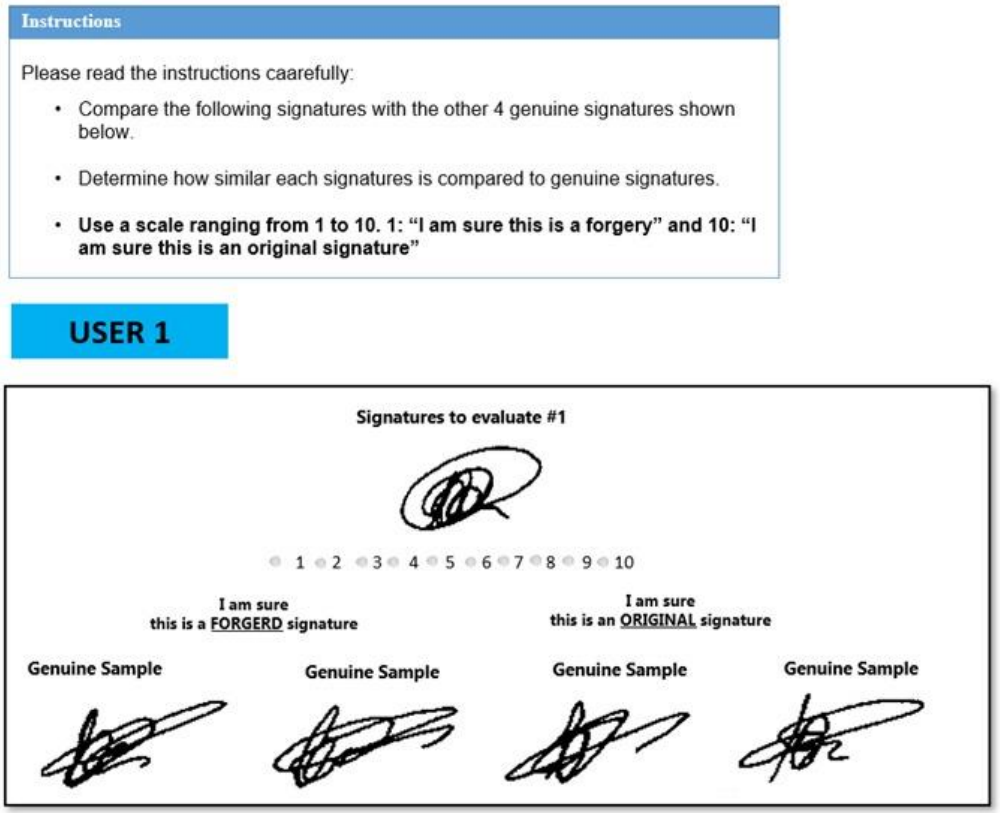


Figura 3.7: Interface de HIT 3

En resumen las 5 HITs están diseñadas con los siguientes 4 parámetros:

1. **Los datos de las firmas:** Las imágenes, videos y secuencias de tiempo, son obtenidos de dos bases de datos públicas: BIOSECURE-DS2 (experimento 1 y 2) y BiosecurID (experimento 3):

   Base de Datos BIOSECURE-DS2: La Base de Datos Multimodal BIOSECURE tiene datos de rasgos como: huella dactilar, rostro, mano, iris, voz y firma. Incluye tres conjuntos de datos DS1 y dos subcorpus de firmas correspondientes a los conjuntos de datos DS2 y DS3:

   - DS1 fue capturado remotamente a través de Internet,
   - **DS2 se adquirió en un entorno de escritorio**
   - DS3 en condiciones móviles.

   Los conjuntos de datos de firmas fueron producidos por un grupo de 667 usuarios. El conjunto de datos DS2 se capturó utilizando un digitalizador Wacom Intuos3 A6 a 100 Hz y el conjunto de datos DS3 se capturó con un PDA. Se solicitó a los usuarios que firmaran mientras estaban de pie y sostenían el PDA en una mano, emulando condiciones de operación realistas. Las señales de posición, presión, azimut y altitud de la pluma están disponibles en DS2, mientras que sólo la posición está disponible en DS3 debido a la naturaleza de la pantalla táctil del PDA. (Martínez-Diaz, et

al., 2009). Las firmas fueron capturadas en dos sesiones y en grupos de 5. Se dejó un promedio de dos meses entre cada sesión. Durante cada sesión, se pidió a los usuarios realizar 3 series de 5 firmas genuinas y 5 falsificaciones entre cada conjunto. Siguiendo este protocolo, cada usuario realizó 5 falsificaciones para los 4 usuarios anteriores en la base de datos. Así, 30 firmas genuinas y 20 falsificaciones están disponibles para cada usuario. Los experimentos 1 y 2 utilizan un subconjunto de la base de datos BIOSECURE-DS2, donde el experimento 1, contienen 12 firmantes diferentes (con 2 muestras genuinas y 1 falsificación por firmante), y el experimento 2, contiene 6 firmantes diferentes (con 5 auténticas y 3 falsificaciones por firmante).

Base de Datos BiosecurID: Esta base de datos fue recogida por 6 diferentes instituciones de investigación española. Incluye los siguientes rasgos biométricos: habla, iris, rostro, firma, escritura a mano, huellas dactilares, mano y pulsación de teclas. Los datos fueron capturados en 4 sesiones distribuidas en un período de 4 meses a un total de 400 usuarios. Las señales de posición, presión, azimut y altitud de firma se adquirieron usando un digitalizador Wacom Intuos3 A4 a 100 Hz. Durante cada sesión, dos firmas fueron capturadas al principio y dos al final, dando lugar a 16 firmas genuinas por usuario. Cada usuario realizó una falsificación por sesión de firmas de otros tres usuarios en la base de datos, en total se obtuvieron 12 firmas falsificadas. (Galbally J., et al, 2010). Para el experimento 3 se utiliza el subconjunto de datos de BisecureID que contiene 132 usuarios (firmantes), donde cada firmante o usuario está compuesto de 28 firmas/usuario, generando un total de 3696 firmas para ser evaluadas a través de la etiquetación manual, donde 2112 firmas son genuinas, y 1584 son firmas falsas. El experimento 3 utiliza un subconjunto de la base de datos BiosecurID, que contiene 20 firmantes y se dividen en 5 grupos de 4 firmantes cada uno (4 x 8 = 32 firmas).

2. Los firmantes seleccionados son los que tienen muestran peor rendimiento (en términos de EER), proporcionado a través de un sistema de verificación de firmas on-line basado en el algoritmo DTW y siete funciones de tiempo derivadas de las secuencias **x**, **y**, **p** (Martinez-Diaz, M., et al, 2014). Se busca seleccionar firmas que supongan un reto para los sistemas automáticos y comprobar así el desempeño de los humanos con ellas.

3. El número de workers varía para los diferentes HITs: 60 workers (HIT1.1-HIT1.3), 30 workers (HIT2) y 60 workers (HIT3).

3.3.- Base de datos generados

Este capítulo de la tesis, se enfoca en la recopilación de datos obtenidos de los workers en los experimentos de evaluación de firmas.

3.3.1.- Recolección y clasificación de datos de la plataforma Mturk

Los workers tienen un tiempo para la ejecución de las HITs. Después del tiempo expirado para realizar la actividad, se descarga desde MTurk los resultados generados por los workers. La clasificación de los datos se realiza en dos etapas básicas:

Aceptación o Rechazo de resultados: Los trabajos realizados por los workers en la plataforma de Mturk, pueden aceptados o rechazados, dependiendo de si las respuestas o resultados no están conforme a los esperados.

Clasificación de datos: En esta etapa la clasificación de datos se realiza de 2 formas: forma subjetiva (interpretación y clasificación) y objetiva (ponderada software matemático). La clasificación subjetiva se aplica para las preguntas de justificación, donde permite discriminar las respuestas que requieren ser interpretadas las justificaciones y poder categorizarlas. La clasificación objetiva se aplica a respuestas en preguntas de similitud y opción múltiple, a través de software matemático (Excel y Matlab).

En los experimentos 1 y 2, se aplica la clasificación subjetiva y objetiva de acuerdo a las respuestas de justificación y ponderación de los workers generando los siguientes resultados:

Clasificación Subjetiva - Experimento 1 y 2: Las justificaciones realizadas por los workers generan datos relevantes de las características más útiles en la HITs, en relación a la evaluación del reconocimiento de una firma genuina o falsificada. Los datos de las firmas se presentan de forma estática (imágenes) y dinámica (videos) como son el trazo y las características (**x**, **y**, **p**) de la firma.

a b

Figura 3.8: Clasificación de justificaciones de workers en experimento 1 y 2.

La Figura 3.8a, muestra las respuestas de percepción de los workers, obtenidas en el experimento 1, generando características discriminantes tales como: Dimensión, punto inicial y final, patrón, estilo de letra y desplazamiento en los ejes x e y, que fueron consideradas el momento de justificar su respuesta. La Figura 3.18b, muestra las respuestas de percepción de los workers, obtenidas en el experimento 2, generando características discriminantes tales como: patrón, estilo de letra y punto inicial y final iguales, que fueron consideradas el momento de justificar su respuesta.

Clasificación Objetiva - Experimento 1, 2 y 3: Las datos ponderados obtenidos por los workers en las preguntas de selección múltiple y de valoración y comparación de similitud de una firma genuina versus una falsificada, permiten establecer el rendimiento del humano en el reconocimiento de firmas estáticas (imágenes) y dinámicas (videos). La clasificación objetiva de acuerdo a las respuestas obtenidas de los workers realizadas en los experimentos 1, 2 y 3, generan los siguientes resultados:

Experimento 1: en este experimento se plantea una pregunta de opción múltiple para las HIT 1.1, HIT 1.2, y HIT 1.3. Cada pregunta de opción múltiple tiene tres respuestas posibles que son: Equal (Igual), Inconclusive (No conclusiva), y Not Equal (No igual).

Experimento 2: en este experimento se pueden generar una o más respuestas ponderadas validas por parte del worker, debido a que pueden existir una o más firmas similares a la firma de evaluación. La tabla 3.2, muestra ejemplos de respuestas dadas por los workers.

Tabla 3.2: Ejemplo de respuestas de selección del experimento 2 (8 firmas a evaluar). Las posibles respuestas son: -1 (Not Equal), 0 (Inconclusive), +1 (Equal)

Worker ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
A3Q7NPNR63WDJJ	0	+1	-1	-1	+1	-1	-1	-1
A3HDJ664DGY14O	-1	0	0	-1	+1	-1	-1	0
A784BOYEKRJCD	-1	-1	-1	-1	-1	+1	+1	-1
APA3LLSQ8OOSL	0	-1	0	-1	-1	0	+1	-1
A3QLMSLL4AYTDT	+1	-1	+1	-1	+1	-1	-1	-1
A22I19IP73HB0D	0	-1	+1	-1	-1	-1	0	-1
A3Q7NPNR63WDJJ	+1	-1	0	0	-1	+1	+1	-1
AWIO57CI1ZBIH	+1	-1	0	-1	+1	-1	-1	0
A16SO41Z0J0RNR	0	-1	+1	-1	-1	0	+1	+1

Experimento 3: en este experimento los datos son generados por la ponderación y valoración en el rango de 1 a 10, dependiendo que tan genuina o falsificada sea respecto a la firma de prueba. Esta tarea se realiza a 8 firmas de prueba de diferentes usuarios. La tabla 3.3 muestra un ejemplo de las respuestas recogidas en este experimento.

Tabla 3.3: Respuestas similitud experimento 3. En este caso, las respuestas son un valor numérico asociado a cuanto de similares les parecen las firmas evaluadas, respecto a la de referencia.

Worker ID	S1	S2	S3	S4	S5	S6	S7	S8
A1EK1N6M8VCQ66	1	2	1	1	2	3	1	1
A1GR8T8TUTJO00	3	7	8	7	8	8	4	6
A1HL0NR9KG5Z0G	1	9	10	2	9	7	1	2
A1O0BPP1U58KCW	8	6	6	4	8	4	9	7
A1OP238L5ZBHZV	4	7	7	6	6	4	3	5

3.4.- Protocolos de evaluación y análisis de Resultados

Los datos obtenidos en el capítulo 3.3 provienen de las preguntas de opción múltiple, preguntas de similitud, y ponderaciones. Estos datos se generan por la aplicación de un protocolo experimental para cada HIT:

3.4.1.- Experimento No.1: Comparación uno a uno (Una firma de entrenamiento vs. Una firma de prueba)

El experimento No.1, se divide en tres HITs, cada una de ellas presenta a los workers diferente tipo de información de las firmas manuscritas, donde su objetivo es analizar el rendimiento del humano en reconocimiento de firma a través de diferente tipo de información tales como estática y dinámica (imagen, trazo y

características de la firma), donde se aplican parámetros de experimentación (Ver Tabla 3.4).

Tabla 3.4: Parámetros del Experimento No.1

PARÁMETRO	CONFIGURACIÓN
Base De Datos	Biosecure DS2
Firmantes	6
Tiempo De Ejecución	2 minuto
Cantidad De Workers	60
Costo Por Workers	$0.04
Costo Total	$2.40

Los workers tardan un tiempo promedio de ejecución de las HITs, de 1 minuto y 12 segundos, pero existen workers que necesitan más tiempo para la ejecución de las mismas, por lo tanto se asigna 2 minutos para realizar la actividad:

HIT 1.1: Esta tarea tiene como objetivo verificar la similitud entre 2 imágenes de firmas manuscritas con información estática. Se aplican con 6 firmantes diferentes de la Base de datos BIOSECURE DS2, los cuales aportan con dos firmas genuinas y una falsificada. El nombre otorgado a cada firma tiene el siguiente formato: **us**xx_**f**yy_**s**zz, donde **us**: usuario, **f**: número de firma, **s**: sesión, xx, yy, zz: son valores numéricos de identificación.

Cada respuesta es analizada para determinar las tasas de Falso Rechazo (FR) y Falsa Aceptación (FA). Además se incluye la tasa de respuesta no definida (ND), que es un parámetro de medición de respuestas de los workers que no están seguros de si es o no una firma genuina o falsa. Los resultados mostrados en la Tabla 3.5, provienen de la categorización de las respuestas de 60 workers.

Tabla 3.5: Resultados HIT 1.1

TAREA	FRR	FAR	ND
HIT 1.1	26.7%	30.0%	33.3%

La Tabla 3.5, muestra que la información estática como es la imagen de la firma manuscrita no le permite al worker tomar una decisión específica en el reconocimiento de firma, ya que existe un 33.3% de respuestas no definidas (ND o Inconclusive). Además muestra, que existe un alto porcentaje de FRR y FAR, donde representa que en un sistema de reconocimiento de firmas, el 26.7% de usuarios que pertenecen al sistema, no pueden ingresar, y que un 30% de personas que no pertenecen al sistema están siendo admitidos.

HIT 1.2: Esta tarea tiene como objetivo identificar la similitud entre 2 firmas, a través de videos que contienen información dinámica, donde se observa el trazo de una firma manuscrita. En esta actividad se utilizan los mismos firmantes de la HIT 1.1.

La HIT 1.2, muestra al worker una pregunta de opción múltiple que tiene tres posibles respuestas de acuerdo a la similitud entre la firma de prueba y la de

entrenamiento. Cada respuesta es analizada para determinar las tasas de falso Rechazo (FR), Falsa aceptación (FA) o no es definida la respuesta (ND). Los resultados mostrados en la tabla 3.6, provienen de la categorización de las respuestas de 60 workers.

Tabla 3.6: Resultados HIT 1.2

TAREA	FRR	FAR	ND
HIT 1.2	40.0%	30.0%	25.0%

La Tabla 3.6, muestra que la información dinámica ayuda a reducir en un 8.3% la cantidad de respuestas ND, respecto de la HIT 1.1, pero aun así todavía existe un 25% de workers que tienen duda en reconocer la similitud entre firmas. Sin embargo la información dinámica no ayuda a mejorar el valor de FAR ya que no muestra cambios con respecto a la HIT 1.1. El rendimiento en términos de FRR presenta un incremento de 13.3% con respecto a la HIT 1.1. Este incremento representa que los workers tienen mayor complejidad de reconocer una firma, es decir, que con un FRR del 40%, la mitad de los usuarios que pertenecen al sistema no pueden acceder siendo parte del mismo.

HIT 1.3: Esta tarea tiene como objetivo identificar la similitud entre 2 firmas, a través de información estática como son las imágenes de las firmas de prueba y entrenamiento y sus respectivas características de posicionamiento en el eje x, y, y p (presión ejercida por el firmante). En esta actividad se utilizan los mismos firmantes de la HIT 1.1. La Figura 3.19, muestra ejemplos de las imágenes que contienen los trazos de las características de la firma utilizadas en la HIT 1.3.

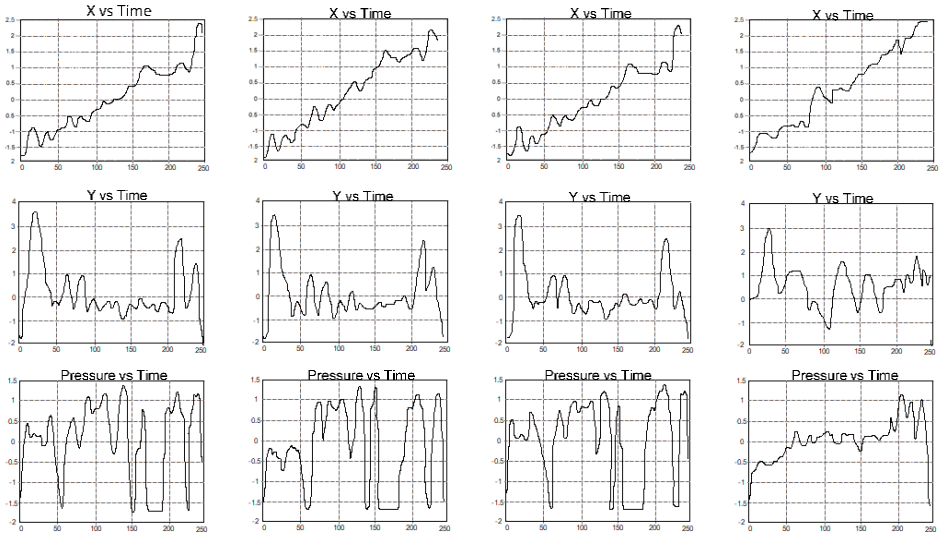


Figura 3.9: Trazos de las características en el eje x, y, y presión

La HIT 1.3, muestra al worker una pregunta de opción múltiple que tiene tres posibles respuestas de acuerdo a la similitud entre la firma de prueba y la de entrenamiento, así como las características x, y, y p. Los resultados mostrados en la tabla 3.7, provienen de la categorización de las respuestas de 60 workers.

Tabla 3.7: Resultados HIT 1.3

TAREA	FRR(%)	FAR(%)	ND(%)
HIT 1.3	43.3	33.3	21.7

La Tabla 3.7, muestra que la información facilitada a los workers en la HIT 1.3, ayudan a reducir la cantidad de respuestas no definidas (ND), en un 11.6% respecto a la HIT 1.1, y en un 3.3% respecto a la HIT 1.2, es decir el worker mejora la toma de decisión, sea correcta o incorrecta. El valor de FAR, se comporta de forma similar que la HIT 1.1, y HIT 1.2, es decir existe un 33.3% de persona que pueden acceder al sistema sin ser parte de ella. Además, el FRR se incrementa en 16.6% respecto a la HIT 1.1 y en un 3.3% respecto a la HIT 1.2, esto conlleva a que el worker a mayor información facilitada, se le hace más compleja la verificación de firma, es decir un 43.3% de personas que son parte del sistema no pueden ingresar al mismo. En un sistema que aproximadamente, la mitad de usuarios (pertenecientes al sistema) no puedan acceder al mismo, se genera en un problema muy grave.

Tabla 3.8: Resultados totales Experimento 1

TAREA	Información	FRR(%)	FAR(%)	ND(%)
HIT 1.1	Estática	26.7	30.0	33.3
HIT 1.2	Dinámica (video)	40.0	30.0	25.0
HIT 1.3	Dinámica (funciones temporales)	43.3	33.3	21.7

La Tabla 3.8, muestra los datos obtenidos de un análisis de todas respuestas generadas por los workers (personas sin experiencia en análisis de documentos forenses) en todas las tareas del experimento 1:

- El FAR prácticamente se mantiene similar en la HIT 1.1, 1,2, 1.3. (≈ 30%).
- El FRR cada vez se incrementa de 26.7% al 43.3%. (↑16.6%).
- El ND cada vez disminuye de un 33.3% a un 21.7% (↓11.6%).

Los experimentos realizados sugieren que las personas participantes tienen muchas dificultades a la hora de verificar firmas, independientemente de la información suministrada. Destacar también que el aumento de la información (estática y dinámica) no conlleva una mejora de los resultados,

3.4.2.- Experimento No.2: Comparación uno a muchos (Una firma de entrenamiento vs. Ocho firmas de prueba)

Este experimento evalúa el desempeño del humano en el reconocimiento de firmas con contexto. Para su ejecución se proporciona al worker información estática de entrenamiento y varias muestras a evaluar. Esta actividad muestra una firma de entrenamiento y un conjunto de 8 firmas de prueba. Al worker se solicita que elija las firmas de prueba que presenten mayor similitud a la firma de

entrenamiento. La Tabla 3.9, muestra los parámetros de configuración del experimento 2.

Tabla 3.9: Parámetros importantes experimento 2

PARÁMETRO	CONFIGURACIÓN
Base De Datos	Biosecure DS2
Firmantes	6
Tiempo De Jecución	1 minuto
Cantidad De Workers	30
Costo Por Workers	$0.02
Costo Total	$0.60

Los workers en el experimento 2 o HIT 2, tardan un tiempo promedio de ejecución de 50 segundos, pero existen workers que necesitan más tiempo para su ejecución, por lo tanto se asigna un minuto para realizar la actividad.

En el experimento 2, las respuestas van en relación a las firmas de prueba que mayor similitud presentan respecto de la firma de entrenamiento. Por lo tanto, las respuestas indican si las firmas elegidas son acertadas (A), son falsos rechazos (FR) o falsas aceptaciones (FA). Para el análisis y procesamiento de las respuestas obtenidas en la HIT 2 hay que tomar en cuenta la ubicación y el orden de las firmas de prueba dentro de la interfaz presentada a los workers. La Figura 3.10, muestra el orden y la ubicación de las firmas de prueba (5 firmas genuinas y 3 firmas falsificadas).

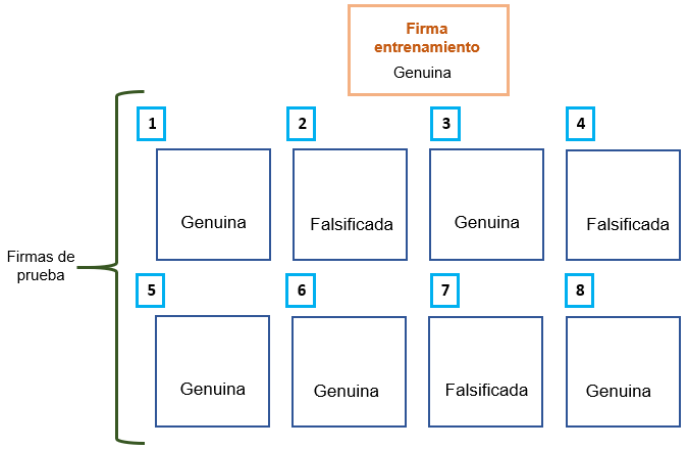


Figura 3.10: Ubicación de firmas aplicadas en experimento 2

La tabla 3.10, arroja resultados del experimento 2, donde se observa una tasa de Falso Rechazo muy elevada. Pese a la tasa baja de Falsa aceptación, los resultados son pobres. La inclusión de información de contexto vuelve al evaluador más “desconfiado” y sesga sus respuestas. En este contexto, los evaluadores tienden a etiquetar la mayoría de firmas como falsas.

Tabla 3.10: Resultados Experimento 2

Tarea	FRR	FAR
HIT 2	80.0%	7.8%

3.4.3.- Experimento No.3: Comparación muchos a uno (Cuatro firmas de entrenamiento vs una firma de prueba)

El experimento 3 o HIT 3, permite determinar el rendimiento del humano usando un protocolo similar al utilizado por los sistemas automáticos. En esta tarea se muestran cuatro firmas de entrenamiento y una firma de prueba, se solicita al worker que califique con un valor de similitud entre 1 a 10, donde: 1 se considera como una firma 100% falsificada, y 10 se considera como una firma 100% genuina, para ejecutar ésta HIT, el worker debe observar las 4 firmas de entrenamiento y debe evaluar y ponderar a la firma de prueba.

En el experimento 3, las respuestas van en relación al grado de similitud de la firma de prueba respecto de las firmas de entrenamiento. La escala de similitud va de 1 a 10, considerando como umbral el valor de 5, por lo tanto, si la respuesta está en el rango de 1-5 la firma se considera falsificada. Si la respuesta está entre 6 y 10 se considera la firma genuina, permitiendo categorizar si la respuesta es un falso rechazo o una falsa aceptación (Ver Figura 3.11).

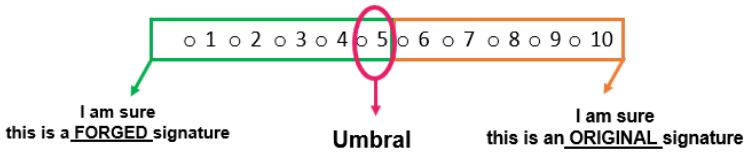


Figura 3.11: Escala de similitud de firmas

Para obtener los datos de ponderación con las dos condiciones mencionadas, se debe tomar en cuenta la ubicación y el orden de las firmas de prueba y de entrenamiento. La Figura 3.12, muestra la ubicación y el orden de la una firma de prueba y las cuatro firmas de entrenamiento.

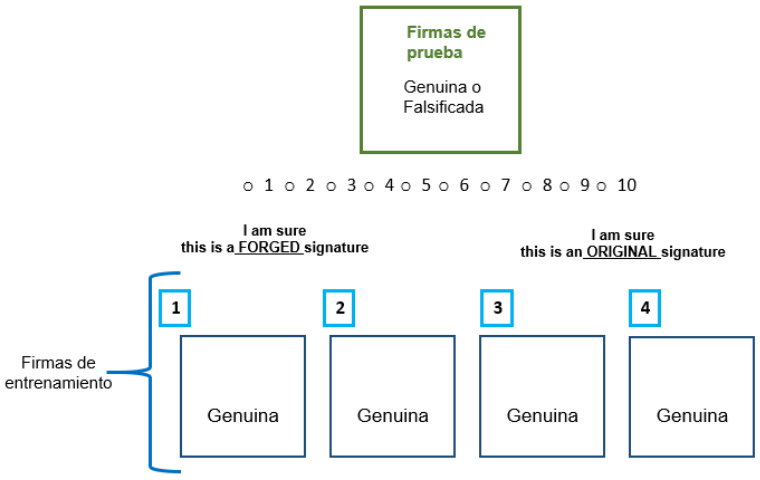


Figura 3.12: Disposición firmas experimento 3

Prueba con 500 workers: En el experimento 3 se realiza el diseño de la tarea con una variación de 60 a 500 workers. La Tabla 3.11, muestra los parámetros de configuración del experimento.

Tabla 3.11: Parámetros de configuración 500 “workers”

PARÁMETRO	CONFIGURACIÓN
Base De Datos	BiosecurID
Firmantes	20
Tiempo De Ejecución	3 minutos(2')
Cantidad De Workers	500
Costo Por Workers	$0.03
Número De Actividades	4
Costo De Actividad Por Worker	$0.12
Costo Total	$60

Los workers al ejecutar la tarea, sus respuestas permiten obtener datos estadísticos con el objetivo de calcular los valores de FAR y FRR, para observar el rendimiento de los workers en el reconocimiento de firma (Ver Tabla 3.12).

Tabla 3.12: Resultados 500 workers

TAREA	FRR	FAR
HIT 3	37.6%	31.4%

Los valores obtenidos en la Tabla 3.12 vuelven a mostrar tasas elevadas de error con valores similares de FAR y FRR.Esto refleja que firmas genuinas son rechazadas en una proporción similar a las firmas falsificadas que están siendo aceptadas. Los datos y el análisis de la tabla 3.12, muestran claramente que un worker por sí solo no es suficientemente hábil en el reconocimiento de firmas manuscritas.

Evolución FAR y FRR de la combinación de respuestas de workers: En el siguiente experimento se comprueba el poder de la multitud y si la combinación de respuestas de malos evaluadores puede dar lugar a una buena evaluación. De forma indirecta nos permitirá también evaluar la correlación entre las respuestas de los diferentes evaluadores. La evolución de FAR y FRR se obtiene de la combinación de las respuestas de los workers. En primera instancia se realiza la combinación de respuestas obteniendo la media de FAR y FRR de varios workers de forma aleatoria. Se realizan 10 repeticiones de este proceso y se calcula la media de los resultados obtenidos en las repeticiones.

Para este segundo análisis se utilizan los resultados obtenidos con 500 workers. Se combinan los resultados de 1 a 500 workers para observar el comportamiento de los valores de FRR y FAR.

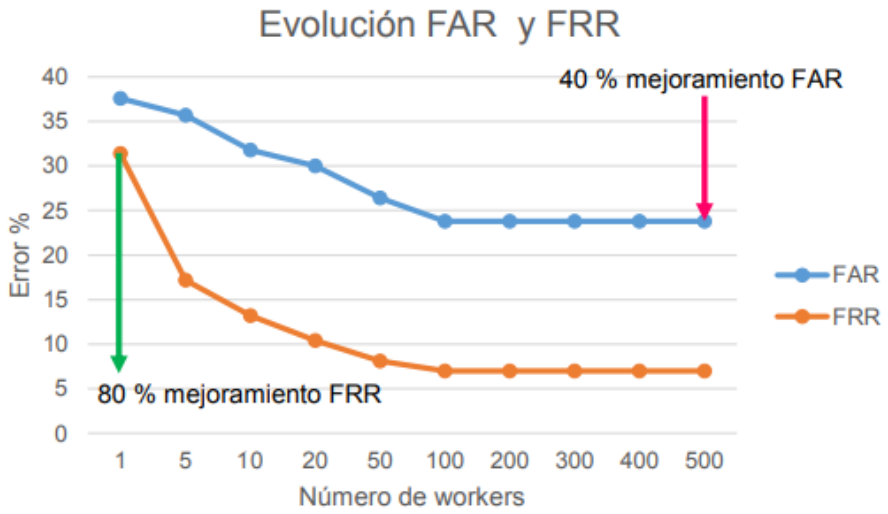


Figura 3.13: Evolución de FRR y FAR (Respuestas combinadas de hasta 500 workers)

La Figura 3.13, muestra la evolución de estos parámetros, donde se observa la tendencia de las curvas de FAR y FRR, y al combinar las respuestas de los workers se obtiene un 80% de mejora en términos de FRR, es decir, mejora el reconocimiento de firma. La Tabla 3.27, muestra la mejora (↓80%) del FRR de 31.4% a 7%, lo que significa que rechaza menos firmas genuinas como falsificadas, además el FAR mejora (↓40%) de 37.6% a 23.8%, lo que significa que acepta menos firmas falsificadas como genuinas.

Tabla 3.13: Evolución FRR y FAR de respuestas combinadas para 500 workers

CANTIDAD DE WORKERS	FAR %	FRR %
1	37,6	31,4
5	35,7	17,2
10	31,8	13,2
20	30	10,4
50	26,4	8,1
100	**23,8**	**7**
200	23,8	7
300	23,8	7
400	23,8	7
500	23,8	7

Los datos y el análisis de la Tabla 3.13, muestran como al combinar las respuestas de los workers se puede mejorar el rendimiento del humano en reconocimiento de firma, además se puede observar que al combinar las respuestas de 100 workers alcanza la más alta mejora en FRR y FAR.

La comparación del desempeño humano por calificaciones humanas versus un sistema automático, se aplica un protocolo estándar. Las respuestas de los workers pueden fusionarse para determinar el potencial de complementariedad de las capacidades humanas. La Tabla 3.14, presenta los resultados del desempeño de los 400 workers (combinando los promedios de las respuestas), con 240 firmas empleadas. Se incluye también el rendimiento obtenido por un sistema automático

de firmas basado en características off-line (Ferrer M., et al, 2012) utilizando un protocolo similar al utilizado por los humanos (4 muestras de entrenamiento y 8 muestras de prueba para cada firmante). Los sistemas de reconocimiento de firma offline se basa exclusivamente en el trazo de la firma y están pensados para aplicaciones dodne no se dispone de la información dinámica (mucho más discriminante). El rendimiento obtenido a través de la combinación de las respuestas humanas es mucho mejor en comparación con los resultados individuales y sugiere que hay un gran margen de mejora en esta área de investigación. Además, el rendimiento obtenido por el sistema automático es mejor que el rendimiento individual. Sin embargo, su desempeño está lejos del desempeño obtenido por la combinación de humanos.

Tabla 3.14: Evaluación del desempeño humano (EER) con intervención a nivel de clasificación vs. ASV off-line (Falsificación simulada)

Sistemas	**EER**
Baseline Automático - ASV basado en características off-line. (Morales A., et al, 2016)	20.27%
Rendimiento humano individual	32.2%
Combinación de respuestas humanas	13.8%.

3.5.- Conclusiones y Contribuciones

Las conclusiones y las contribuciones muestran el aporte o no del humano en los sistemas de reconocimiento de firma.

3.5.1.- Conclusiones.

1. Las personas sin experiencia en FHE y sin información de las muestras a evaluar, son malos en el reconocimiento de firma. Tabla 3.15, muestra el promedio de rendimiento en términos de ND, donde se puede observar que el 26.7% de personas dudan en reconocer una firma genuina de una falsificada.

Tabla 3.15: Resultado promedio del Experimento 1

TAREA	**FRR(%)**	**FAR(%)**	**ND(%)**
Promedio HIT 1.1, HIT 1.2, HIT 1.3	36.7	31.1	26.7

2. Los resultados en reconocimiento de firma, no mejoran a mayor cantidad de información, sea muestras estáticas o dinámicas. La Tabla 3.14, muestra el promedio de rendimiento en términos de FRR y FAR, donde se puede observar que con un FAR del 36% y un FRR del 31%, reflejan que las personas no puedan reconocer una firma genuina de una falsificada es muy compleja, brindando ayuda de señales estáticas y dinámicas.

3. Al realizar una combinación de 100 respuestas de los humanos, generan datos interesantes y alcanzan buenos resultados. La Tabla 3.16, muestra la mejora

(↓80%) del FRR de 31.4% a 7%, lo que significa que rechaza menos firmas genuinas como falsificadas, además el FAR mejora (↓40%) de 37.6% a 23.8%, lo que significa que acepta menos firmas falsificadas como genuinas.

Tabla 3.16: Mejor resultado de FRR y FAR para 100 respuesta combinadas de 500 workers

CANTIDAD DE WORKERS	FAR %	FRR %
100	23,8	7

3.5.2.- Contribuciones.

En este capítulo de la tesis, se muestra las diferentes contribuciones, a través de artículos de investigación publicadas en diferentes congresos relacionadas con el crowdsourcing:

Articulo Congreso	**D. Morocho, A. Morales, J. Fierrez and R. Tolosana. Signature recognition: establishing human baseline performance via crowdsourcing. Proc. 4th Int. Workshop on Biometrics and Forensics (IWBF), pp. 1-6, 2016.**
Contribución	El articulo explora el crowdsourcing para establecer una línea de base del desempeño humano en el reconocimiento de firmas. El potencial de plataformas como Mturk aplicadas en el reconocimiento biométrico es grande y conlleva a explorar nuevos conocimientos sobre las tareas de colaboración masivas con humanos. Los resultados obtenidos en este artículo sugieren que más características de la firma, no significa necesariamente un mejor rendimiento. La toma de decisión del worker está estrechamente relacionada con la información proporcionada y se pueden obtener mejores FAR o FRR según las muestras mostradas. Finalmente, la fusión de calificaciones humanas ha mostrado un gran potencial de mejora en términos de FRR (70% de mejora cuando se fusionan las respuestas de 10 trabajadores).

Articulo Congreso	**D. Morocho, J. Hernandez-Ortega, A. Morales, J. Fierrez, J. Ortega-Garcia. On the evaluation of human ratings for signature recognition. IEEE Int. Carnahan Conf. on Security Technology (ICCST), 2016**.
Contribución	Este artículo explora la capacidad del humano en el reconocimiento de la autenticidad de una firma manuscrita, a través de tareas inteligentes vía crowdsourcing que permite evaluar el rendimiento del humano en el reconocimiento de firmas. A través de varias evaluaciones con diferentes características de la firma se puede evaluar la similitud entre ellas. Las respuestas de los profanos se utilizan para analizar el rendimiento de los seres humanos relacionados con cada uno de los escenarios de los protocolos. Las respuestas generadas por 400 workers aplicados a 240 firmas de la base de datos pública de BiosecurID, generan resultados que evidencian la complejidad de verificación de firmas, con un FAR que van del 50% al 75%. En este escenario el análisis de los resultados sugieren que las capacidades del humano en reconocimiento de firma manuscrita, dependen de las características presentadas y la complejidad de la firma.

Articulo Revista	**A. Morales, D. Morocho, J. Fierrez, R. Vera. Signature authentication based on human intervention: performance and complementarity with automatic systems. IET Biometrics, vol. 6(4), pp. 307-315, 2017. (Journals IET Biometrics, The Institution of Engineering and Technology (IET), Special Issue: Selected Papers from the International Workshop on Biometrics and Forensics)**
Contribución	Este trabajo explora la intervención humana en los sistemas de autenticación de firma, se considera la intervención a nivel de clasificación, con un análisis de cómo los humanos se desempeñan en las tareas de autenticación de firma. Los experimentos basados en el análisis de la respuesta de 500 workers ayudan a establecer una línea de base humana. Los resultados sugieren que los profanos se desempeñan peor que los sistemas ASV y resaltan las dificultades asociadas a esta tarea. La tasa de error promedio de los profanos es de alrededor del 30%, pero las respuestas combinadas generan un potencial de las capacidades humanas.

Capítulo 4

4.- Reconocimiento de firma asistida por humanos

Según los resultados obtenidos en las evaluaciones realziadas, las personas sin formación específica (profanas) no son buenas en tareas como el cotejo de firmas. No obstante, es bien sabido que con una formación especializada, un experto forense es capaz de reconocer falsificaciones con un alto grado de precisión. Existe un paso intermedio, que es dotar a las personas sin experiencia de ciertas guías que ayuden a dirigir su evaluación. En este capítulo evaluaremos el desempeño de personas no expertas en la extracción de atributos (previamente seleccionados en base a la literatura) que sirvan para asistir a los sistemas automáticos de reconocimiento de firmas.

En biometría, los esquemas asistidos por humanos aprovechan las capacidades humanas de abstracción y las capacidades del sistema automatizado para modelar estadísticamente los datos (Kumar N., et al, 2011; Reid D., et al, 2014; Klare B. F., et al, 2014; Samangouei P., et al, 2015; Tome P., et al, 2014). El uso de anotaciones por humanos en sistemas automáticos de reconocimiento biométrico ha generado resultados alentadores (Tome P., et al, 2014). La anotación de atributos realizada por humanos ha generado alternativas para mejorar los sistemas de reconocimiento automático de cara (Kumar N., et al, 2011; Klare B. F., et al, 2014; Samangouei P., et al, 2015; Tome P., et al, 2014) o marcha (Reid D., et al, 2014). Los resultados obtenidos en (Malik M. I., et al, 2013), sugieren que los FDE pueden lograr un rendimiento similar a los sistemas automáticos con precisiones superiores al 90%. Los FDEs tienen alta capacidad para analizar la autenticidad de las firmas, debido a su alto desempeño en la clasificación de firmas genuinas y falsificadas (Coetzer J., et al, 2006). ¿Podrían alcanzar una tasa similar personas no entrenadas pero si guiadas correctamente?

El crecimiento de los sistemas automatizados está disminuyendo la intervención humana en muchas aplicaciones de reconocimiento. Sin embargo, no hay que subestimar al humano por su capacidad analítica y su percepción. Existen escenarios en los que un humano puede ayudar o contribuir a una verificación automática de firmas, por ejemplo: la banca, puntos de venta, notaría pública o la entrega de paquetes.

Qué acciones tomar y en qué medida esas acciones pueden ayudar a los sistemas automáticos de verificación de firma es una de las preguntas que motivó esta tesis. En la Figura 4.1, se puede observar la intervención del humano en los ASV, aplicado en múltiples niveles o fases de un sistema biométrico. Las posibles intervenciones del humano son: evaluación de calidad para eliminar muestras de mala calidad, anotación de atributos, clasificación o comparación de muestras, y la toma de decisiones. Sin embargo, el rendimiento de los humanos en tareas de

verificación de firmas permanece sin explorar y sus capacidades sin ser valoradas (Malik M. I., et al, 2013).

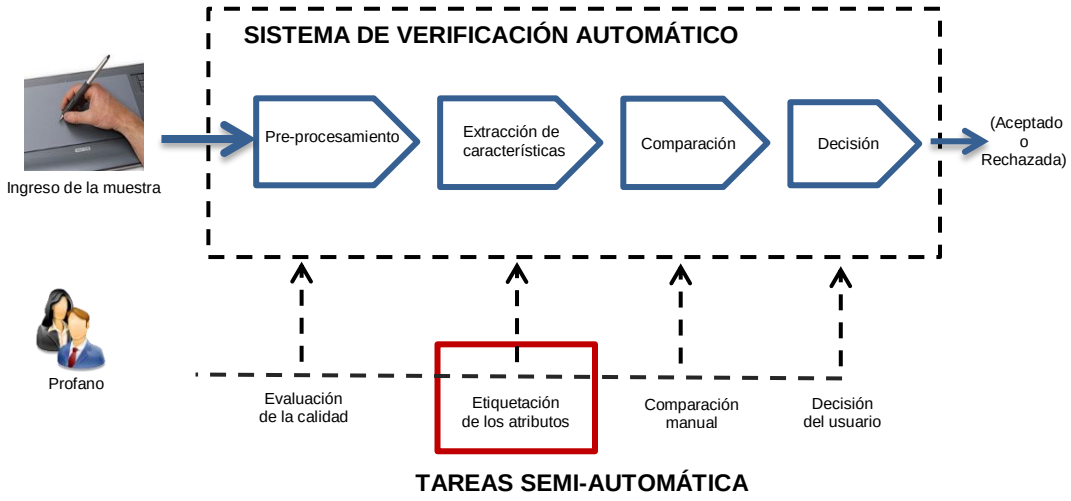


Figura 4.1: Esquema básico de reconocimiento de firma asistido por humanos. (Morocho D., et al, 2016)

En este capítulo de la tesis se analiza el potencial del humano en reconocimiento de firma basado en atributos, a través de un conjunto reducido de características discriminantes que se puedan etiquetar manualmente en un período de tiempo razonable. Por ejemplo, en el caso para una entrega de paquetes debe ser menor a 10 segundos, o para una transacción bancaria menos de 1 minuto. Este análisis se realiza a través de un nuevo sistema de reconocimiento de firmas basado en atributos, y un sistema automático de reconocimiento de firmas on-line.

4.1.- Reconocimiento de firmas basado en atributos

El rendimiento humano en reconocimiento de firma obtenido mostrado en el capítulo anterior (Morocho D., et al, 2016), sugiere que los profanos tienen dificultades para reconocer correctamente la autenticidad de las firmas. Sin embargo, es conocido que los FDEs pueden lograr desempeños competitivos basados en su entrenamiento especializado, así como su experiencia.

En este capítulo de la tesis se analiza los atributos seleccionados que son anotados de forma manual (realizados por humanos, pero inspirados en el trabajo de los FDEs). Además, se ha desarrollado una interface (aplicación) en Matlab2012 © GUI, constituida con fuentes de información, inspiradas en el trabajo realizado por FDE. La misma está diseñada para ser utilizada por un humano sin experiencia en procesos de reconocimiento de firma o análisis de FDE. Todos los atributos se anotan desde una única imagen binaria estática de la firma (los datos dinámicos no se utilizan y cada firma se anota por separado).

Atributos de la Firma Inspirados en FDEs: Existen muchas características de una firma que se pueden analizar (Olivera L., et al, 2005; Burkes, T. M., et al, 2009; E2290-07a Standard Guide for Examination of Handwritten Items, ASTM, 2007). Existen muchos atributos de la firma, pero para los experimentos realizados en esta tesis se han seleccionado un conjunto de 13 atributos (inspirados en el análisis FDE), que se basan en dos principios: i) **La eficiencia**, donde la anotación de los atributos debe ser rápida para un humano sin experiencia de FDE; y ii) **El rendimiento**, donde los atributos deben ser discriminativos y útiles para el reconocimiento de firmas.

4.2.- Atributos categóricos y escalares de la firma

La lista de atributos de una firma utilizada en la autenticación de firmas, ya sea en escenarios forenses o automáticos, es extensa (Olivera L., et al, 2005; Burkes, T. M., et al, 2009; Martinez-Diaz M., et al, 2014). Estos atributos se pueden clasificar de acuerdo con diferentes criterios como la naturaleza de la información (por ejemplo, grafología o grafometría), o las diferentes partes de la firma (por ejemplo, floritura, texto), entre otras. La Tabla 4.1 presenta una taxonomía de las características más populares analizadas en la literatura de FDE.

Tabla 4.1: Taxonomía de las características utilizadas en el análisis de FDE: (a) Morfológicas; (b) Dinámicas; (c) Habilidad del escritor; (d) Estilo de escritura. (Morales A., et al, 2016)

Morfológico	Dinámica	Capacidad de escritor	Estilo de escritura
proporcionalidad	velocidad	indecisión	forma
inclinación	presión general	ampliaciones	formateo
Alineación con la línea de base	presión local	habilidad	método de producción
bucles de texto	lentitud	temblor	adornos
características de floritura	parada	arreglo	lateralidad
tamaño	finalización repentina	retoque	trazos cruzados y puntos
espaciado entre caracteres		legibilidad	trazos de entrada/salida
longitud de trazo		libertad de ejecución	puntuación
dirección de trazos		simplificación	conexión
orden		rango de variación	énfasis
		sujetador de pluma	capitalización
			manos

En esta tesis se propone un conjunto de atributos que permite determinar, si la percepción del humano puede ayudar a mejorar los sistemas automáticos. El conjunto de características se divide en dos grupos: atributos categóricos, y atributos de medidas o escalares. La Figura 4.2, muestra un ejemplo de los 13 atributos seleccionados que permiten caracterizar el trazo de una firma.

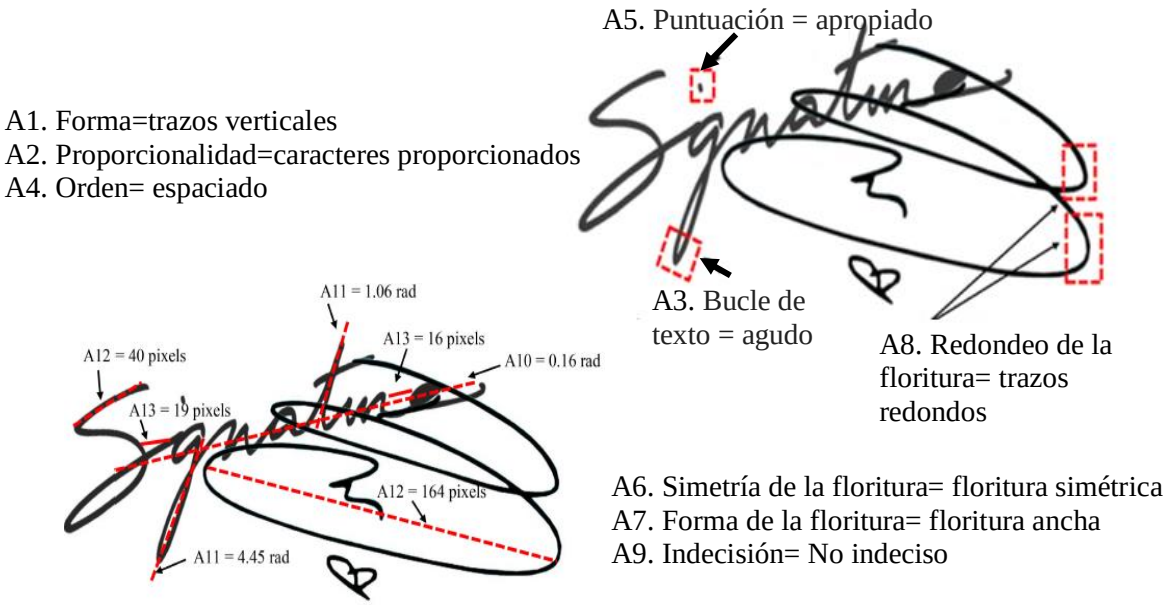


Figura 4.2: Ejemplo de una Firma con sus atributos categóricos (izq.) y atributos escalares (der.) (Morales A., et al, 2016)

4.2.1.- Atributos categóricos (A1-A9)

Los atributos categóricos son denotados por etiquetas discretas. Por ejemplo: firma espaciada/concentrada o firma proporcional/desproporcional. Está compuesto de 9 atributos, que permiten categorizar a una firma de otra dependiendo de su forma, su proporción, su orden, sus lazos, etc. A continuación se presentan los atributos categóricos seleccionados en esta tesis:

- **Forma (A1):** este atributo se asocia con el modelo gráfico utilizado para crear la firma (trazo dominante centrado en el contorno de la firma). Las etiquetas asociadas a este atributo son: los trazos redondeados, verticales, horizontales, o caligráfica.

- **Proporcionalidad (A2):** este atributo se relaciona con la simetría y el tamaño de la escritura a mano. Las etiquetas asociadas a este atributo son: proporcional o regular, desproporcional o irregular, o mixta.

- **Bucles de texto (A3):** este atributo se relaciona con el estilo predominante de los bucles, los lazos y giros de la rúbrica o de los caracteres por ejemplo en letras como "l, g, p, f, j, y", y cambios de dirección como en letras mayúsculas "A, M, N". Las posibles etiquetas son: redonda, ovalada u otra.

- **Orden (A4):** este atributo se refiere a la distribución gráfica de las partes que forman la firma: orden claro, confuso, concentrado o espaciado.

- **Puntuación (A5):** este atributo analiza cualquier signo de puntuación o trazo distintivo que pueda caracterizar la firma, como es el caso de la "i" o "j", además verifica si la firma tiene la puntuación adecuada, si tiene signos de puntuación pero está en el lugar incorrecto o no puntuación.

- **Simetría de la floritura (A6):** este atributo se refiere a los trazos de floritura y su simetría.

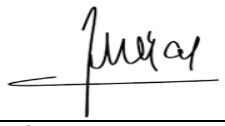
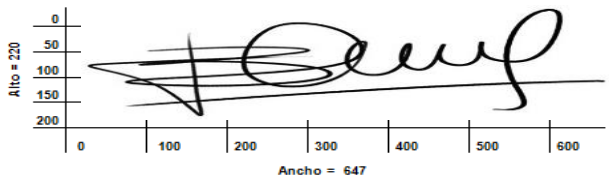
- **Forma de la floritura (A7):** este atributo está relacionado con la forma completa de la floritura que se caracteriza comúnmente por movimientos delgados o anchos realizados por movimientos rápidos y muy dependientes de la persona.

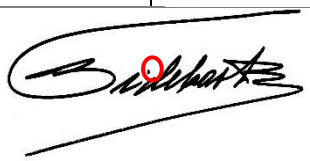
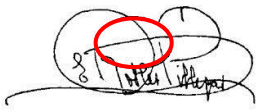
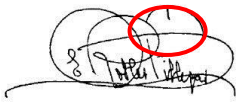
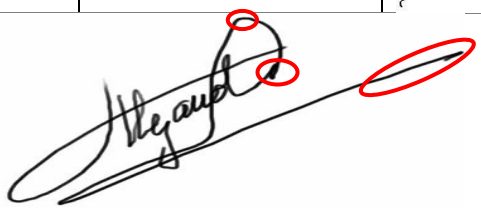
- **Redondeo de la floritura (A8):** este atributo está relacionado con el estilo de los trazos del adorno, que generalmente incluye cambios de dirección que pueden clasificarse en agudo (muy abrupto) o redondo (cambio suave de dirección).

- **Indecisión (A9):** este atributo revela el nivel de indecisión percibida en la firma. La indecisión produce la ampliación de los caracteres, la tendencia de las curvas a convertirse en ángulos, los parches y los retoques, los temblores, entre otros. Las posibles etiquetas son: el usuario no dudó al hacer la firma, y el usuario dudó mientras hizo la firma.

La Tabla 4.2, presenta un resumen de los atributos o características categóricas que se utilizan para analizar la autenticidad de una firma.

Tabla 4.2: Resumen de atributos o características categóricas.

<table>
<tr><td rowspan="2">Forma
(A1)</td><td>Vertical
  </td><td>Redonda
  </td></tr>
<tr><td>Horizontal
  </td><td>Caligráfica
  </td></tr>
<tr><td>Proporcionalidad
(A2)</td><td colspan="2"> 
</td></tr>
<tr><td>Bucles de texto
(A3)</td><td colspan="2">  </td></tr>
</table>

<table>
<tr><td rowspan="2">Orden
(A4)</td><td>Claro
  </td><td>Confuso
  </td></tr>
<tr><td>Concentrado
  </td><td>Espaciado
  </td></tr>
<tr><td>Puntuación
(A5)</td><td colspan="2">  </td></tr>
<tr><td>Atributos de floritura
(A6 – A7 – A8)</td><td>Simetría de la floritura
  </td><td>Forma de la floritura
  </td><td>Redondeo de la floritura
  </td></tr>
<tr><td>Indecisión
(A9)</td><td colspan="2">  </td></tr>
</table>

4.2.2.- Atributos escalares (A10-A13)

Los atributos escalares o de medición son calculados con ayuda de los puntos claves localizados manualmente. Por ejemplo, la distancia entre caracteres o trazos, y la selección de un punto clave. Estos atributos tratan de explotar la capacidad del humano para destacar las regiones de la firma más representativa.

- **Alineación con la línea de base (A10):** este atributo es conocido como inclinación. Se define como el ángulo de la firma con respecto al eje horizontal imaginario, podría ser un desafío en firmas de alta complejidad.

- **Inclinación de los trazos (A11):** este atributo se relaciona con los ángulos de los distintos caracteres. Mide la pendiente (ángulo respecto de la línea base). El anotador tiene que elegir cuáles son los trazos más relevantes.

- **Longitud de trazos (A12):** este atributo se relaciona con las medidas inclinadas, el anotador tiene que seleccionar hasta tres trazos representativos (puntos inicial y final) para calcular automáticamente sus longitudes (en píxeles).

- **Espaciado entre caracteres (A13):** este atributo mide la separación (en píxeles) entre hasta cuatro caracteres relevantes en la firma (caracteres a elegir por el etiquetador).

Los atributos A12 y A13 se miden en píxeles. Para mejorar la interoperabilidad entre diferentes escáneres, los valores se pueden convertir al Sistema Internacional de Unidades utilizando el parámetro de resolución (por ejemplo, 600 ppp de la base de datos de BiosecurID).

La Tabla 4.3, presenta un resumen de los atributos o características escalares que se utilizan para analizar la autenticidad de una firma. Estos atributos son aquellos que para obtener un dato se requiere que se grafique un ángulo o una distancia.

Tabla 4.3: Resumen de características o atributos de medición o escalares.

Alineación con la línea de base (A10)	
Inclinación de los trazos (A11)	
Longitud de trazos (A12)	
Espacio entre caracteres (A13)	

4.2.3.- Interface de etiquetación manual de atributos de una firma

La Figura 4.3, muestra la interface de etiquetación manual de atributos categóricos y escalar, donde el profano (etiquetador) realiza el proceso de etiquetación de los distintos atributos de la firma de la base de datos BiosecurID, generando una nueva base de datos de los atributos etiquetados llamada DB_Labeling_ESPE.

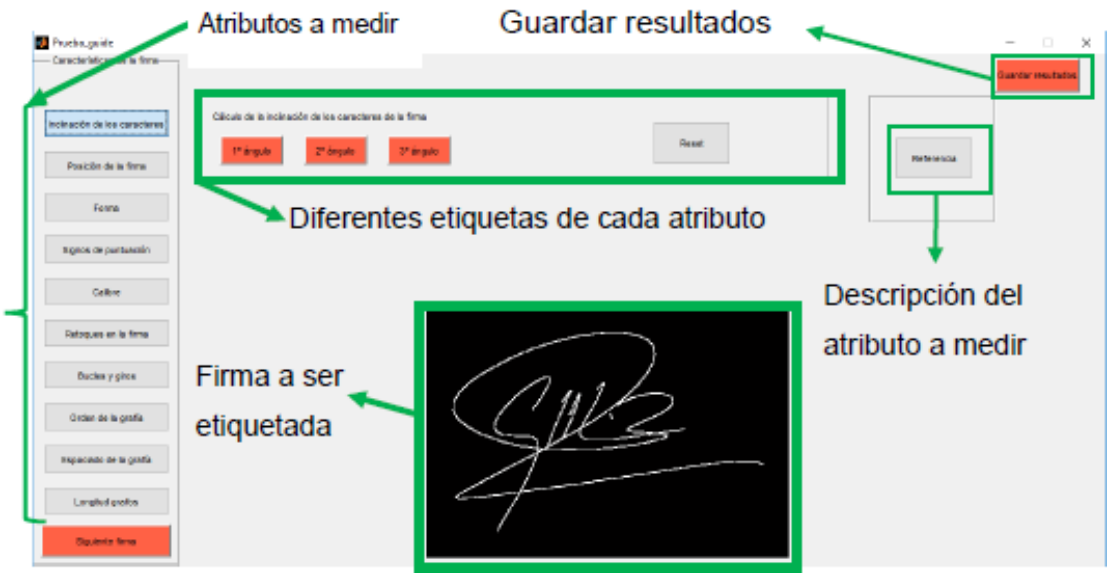


Figura 4.3: Diseño de la Interfaz de etiquetación de atributos categóricos y escalar.

4.2.4.- Comparación basada en atributos categóricos y escalares

Los atributos se combinan en un vector único para cada firma. Se asigna un valor discreto entre 0 y 6 (dependiendo del número de etiquetas del atributo) a cada característica. La distancia entre dos vectores de atributos se calcula utilizando la distancia de Manhattan normalizada por la desviación absoluta media de cada atributo. Las medidas escalares son valores que dependen del atributo (longitud en píxeles para A12-A13 o ángulos en radianes para A10-A11). Se asume $\mathbf{f} = [f_1, f_2, \dots, f_M]$ como el vector de características (con M características) de una firma dubitada y $g^k]\ k \in 1, \dots T$ como un conjunto de inscripción con T firmas indubitadas. La distancia entre el vector de características $\mathbf{f}$ de la firma dubitada y el conjunto de inscripción $\{\boldsymbol{g}^k\}_{k=1}^{T}$ es calculada como:

$$d = \sum_{i=1}^{M} \left(\frac{|f_i - \overline{g_i}|}{\sigma_i} \right) \quad (1)$$

Donde, $\overline{\boldsymbol{g_i}}$ es el promedio del conjunto de entrenamiento y $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \ldots, \sigma_M]$ es la desviación estándar de las características de entrenamiento. En los experimentos, T es igual a 4 y M = 20 (obsérvese que los atributos A11-13 comprenden 10 medidas). En el caso de los atributos categóricos la distancia se limita a dos valores 1 o 0 (por cada atributo) en caso de coincidencia o discrepancia.

4.2.5.- Protocolos de experimentación y Base de datos.

La base de datos utilizada en los experimentos es la base de datos multimodal BiosecurID (Fierrez J., et al, 2010). El subcorpus empleado consta de 132 firmantes de la UAM, corpus adquiridos en cuatro sesiones diferentes, con 16 firmas genuinas (cuatro por sesión) y 12 falsificaciones simuladas (tres por sesión) para cada firmante (132 × 28 = 3696 firmas). Las firmas simuladas son creadas por escritores diferentes al firmante que intenta imitar el estilo natural de una firma genuina. Las firmas se realizaron en un área marcada sobre plantillas de papel (25 mm × 120 mm) con un lápiz de tinta que también capturó las trayectorias x e y, y la presión del lápiz p durante el proceso de firma, con una frecuencia de muestreo de 100 Hz. La base de datos incluye tanto la secuencia dinámica [**x**, **y**, **p**] como la imagen estática escaneada (600 ppp) de las hojas.

4.2.5.1.- Base de datos DB_Labeling_ESPE

Para la adquisición de la base de datos DB_Labeling_ESPE, se cuenta con el subconjunto de 132 usuarios (firmantes) de la base de datos BiosecureID, teniendo un total de 3696 firmas para ser evaluadas a través de la etiquetación manual, donde 2112 firmas son genuinas, y 1584 son firmas falsas. La base de datos de atributos generada comprende más de 800,000 datos [La base de datos de atributos completa está disponible en: http://www.atvs.ii.uam.es/databases.jsp].

Corpus de la base de datos: Para la adquisición de la base de datos se contó con la colaboración de 11 personas sin experiencia en FDEs, no se proporcionó información sobre la autenticidad (genuina o falsificada) de las muestras al anotador y todas las firmas se analizaron por separado. Las características de este grupo de trabajo son:

- Género de los participantes: 1 mujer y 10 hombres
- Edad: de 24 a 27 años

- Nivel académico: estudiantes de maestría de la Universidad de las Fuerzas Armadas ESPE.
- Experiencia con el tema: sin experiencia en FDE.

Cada etiquetador entrega en un archivo de 20 datos extraídos del proceso de etiquetación de cada firma etiquetada, donde la base de datos DB_Labeling_ESPE, se compone de 813.120 datos (20 datos extraídos × 28 firmas × 132 usuarios × 11 etiquetadores).

4.2.6.- Experimentos y análisis de resultados de evaluación.

Los experimentos están enfocados a determinar la utilidad de etiquetas humanas para mejorar procesos de reconocimiento automático de firma manuscrita.

4.2.6.1.- Discriminabilidad de atributos

Este experimento tiene como objetivo evaluar el poder discriminativo de los atributos anotados manualmente. En primera instancia, las etiquetas categóricas se codifican en valores numéricos de 1 al número de etiquetas posibles para cada atributo (por ejemplo, seis para A1 o cuatro para A2).

Sea A_{ij} una matriz con los valores del atributo i ∈ {1, 2,..., 20} anotados por el anotador j ∈ {1, 2,..., 11} para toda la base de datos (considerando que $A11$, $A12$ y $A13$ tienen más de una anotación):

$$\hat{A}_i^j(n,p) = \frac{1}{2}\left(tanh\left(0.01\left(\frac{A_i^j(n,p)-\mu_i}{\sigma_i}\right)\right)+1\right) \quad (2)$$

Inicialmente todos los valores $A_{ij}(n,p)$ se normalizan, debido a que μ_i (media) y σ_i (desviación estándar) del atributo i de todas las firmas genuinas de todos los anotadores j. Se utiliza la función de normalización de $tanh$ para reducir el impacto de los valores no deseados en los modelos generados a partir de las características etiquetadas (Jain, A.K, et al, 2005). El índice n ∈ {1,..., N = 132} es el firmante y p ∈ {1,..., P = 16} es el número de muestra. Se define dos índices de discriminabilidad D_R (Comparaciones aleatorias) y D_F (Comparaciones falsificadas). En D_R, la firma n se evalúa contra muestras de firmas de diferentes firmantes.

$$D_R(i) = \frac{1}{(N-1)N}\sum_{n=1,n\neq m}^{N}\sum_{m=1}^{N}\frac{|\mu_i(n)-\mu_i(m)|}{\sigma_i(n)-\sigma_i(m)} \quad (3)$$

donde, $\mu_i(n)$ es el atributo para el firmante n calculado a través de las 16 firmas genuinas disponibles para ese firmante (y todos los anotadores). De manera similar, $\sigma_i(n)$ es la desviación estándar del atributo i para el firmante n. El índice de discriminabilidad de las falsificaciones simuladas D_F para el atributo i se calcula como:

$$D_F(i) = \frac{1}{N}\sum_{n=1}^{N}\frac{|\mu_i(n) - \widetilde{\mu_i}(m)|}{\sigma_i(n) - \widetilde{\sigma_i}(m)} \tag{4}$$

donde, $\widetilde{\mu_i}(m)$ y $\widetilde{\sigma_i}(m)$ son la media y la desviación estándar de las falsificaciones simuladas del firmante n calculadas en las 12 falsificaciones disponibles para ese firmante (y todos los anotadores). En el caso de los atributos con más de una anotación ($A11$, $A12$ y $A13$), las anotaciones se procesan por separado y luego se combinan en un valor por promedio. Los resultados desprenden que la discriminabilidad de los atributos es mayor en las falsificaciones aleatorias (ver Figura 4.4). Sin embargo, los resultados dependiendo del escenario (falsificaciones aleatorias o simuladas), donde algunos atributos pueden ser más discriminatorios que otros. Por ejemplo, la característica indecisión $(A9)$ es más discriminante para las falsificaciones simuladas que para las aleatorias. Esto se debe a las indecisiones del falsificador que no están presentes en las firmas genuinas (utilizadas para las comparaciones aleatorias). Para la característica forma $(A1)$ es altamente discriminativa para comparaciones aleatorias pero no para falsificaciones.

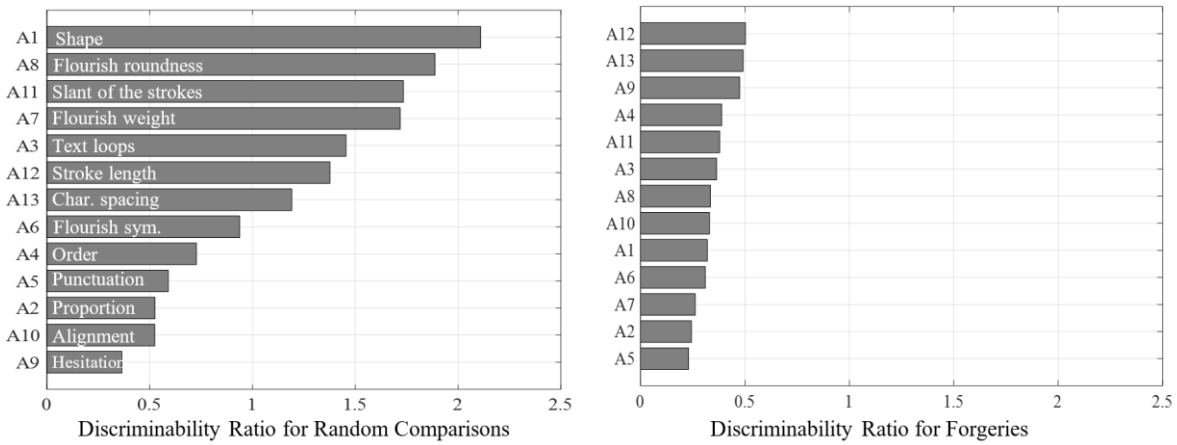


Figura 4.4: Índice de discriminación de los diferentes atributos para comparaciones de falsificaciones aleatorias (izq.) y simuladas (der.) (Morales A., et al, 2016)

La anotación de los atributos depende de la percepción de cada anotador y puede variar entre ellos. Para medir la inestabilidad de un atributo i, se calcula el índice $\bar{S}(i)$, como

$$\bar{S}(i) = \frac{1}{NPT}\sum_{n=1}^{N}\sum_{p=1}^{P}\frac{1}{11}\sum_{j=1}^{11}\left|A_i^j(n,p) - mode\left(\boldsymbol{A}_i(n,:)\right)\right| \tag{5}$$

donde, N (firmantes), P (muestras), y T (etiquetas) son el número de cada atributo, (N = 132, P = 16; T varía para cada característica). $A_{ij}(n,p)$, es el valor del atributo i por el anotador j para la muestra p del firmante n. $A_{i:}(n,:)$, es una matriz (dimensión 176 × 1) con todos los valores del atributo i de todas las muestras del firmante n y todos los anotadores (11 anotadores × 16 muestras genuinas por firmante = 176).

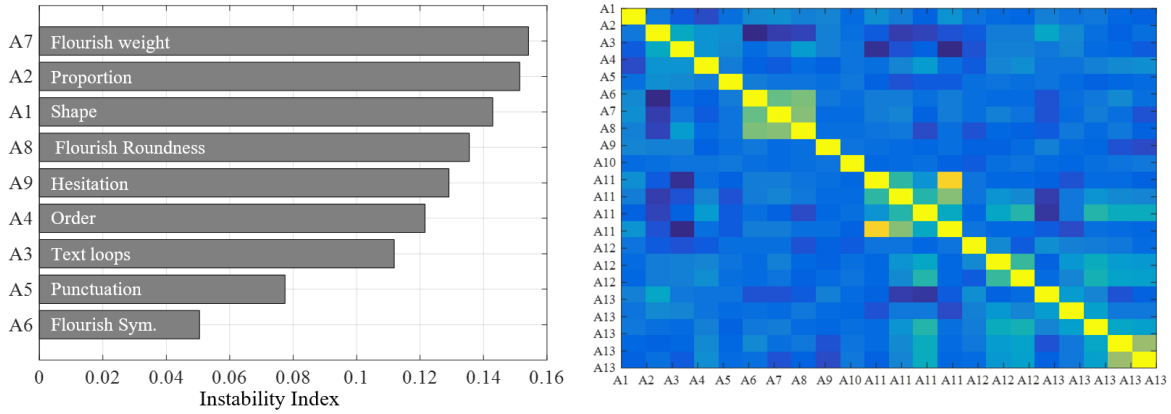


Figura 4.5: Índice de inestabilidad de los atributos categóricos para firmas genuinas (izq.) y matriz de correlación de los atributos (der.) (Morales A., et al, 2016)

La Figura 4.5 (izq.), muestra el índice de inestabilidad para las nueve características categóricas. Las características escalares dependen de los puntos seleccionados por los anotadores. Por lo tanto, los índices de inestabilidad de las características escalares son mayores que las características categóricas. Los resultados muestran cómo algunos atributos como la forma de la floritura (A7), proporción (A2), Forma (A1) y redondeo de la floritura (A8) son menos estables que otros, como la simetría de la floritura (A6), Puntuación (A5) o Bucles de Texto (A3). La Figura 4.5 (der.), muestra la matriz de correlación de todos los atributos, donde, presenta una baja correlación entre las características, a excepción de los tres atributos relacionados con las características de floritura (A6, A7 y A8) y las cuatro medidas de inclinación de los trazos (A11).

4.2.6.2.- Rendimiento del humano en anotación de atributos de la firma

Los experimentos están diseñado para responder las siguientes preguntas: ¿Cuál es el rendimiento de los atributos de firma anotados manualmente? ¿Cuál es la complementariedad del humano (en términos de rendimiento) entre el reconocimiento basado en atributos y el reconocimiento automático tradicional de firmas en línea?. El experimento se divide en dos categorías:

- **Escenario 1:** falsificaciones aleatorias: el modelo del usuario se evalúa utilizando muestras genuinas de otros usuarios (diferentes al dueño de la firma) como ataques impostores (simulación de usuarios que intentan falsificar la identidad del usuario con su propia firma).

- **Escenario 2:** falsificaciones simuladas: también conocidas como falsificaciones especializadas, el modelo del usuario se evalúa utilizando imitaciones hechas por otros usuarios de la base de datos BiosecurID (Fierrez J., et al, 2010), que contiene diferentes niveles de habilidad.

El conjunto de firmas de entrenamiento está compuesto por cuatro firmas genuinas de la primera sesión de cada usuario. Se obtienen puntuaciones genuinas comparando el modelo de entrenamiento con las 12 muestras originales restantes de cada usuario (sesiones 2-4) para un número total de puntuaciones genuinas igual a 1584 (132 × 12). Las puntuaciones del impostor para las falsificaciones aleatorias se obtienen comparando el modelo de entrenamiento con las primeras muestras genuinas de todos los usuarios (diferentes al dueño de la firma) para un número total de puntuaciones aleatorias de impostor igual a 17,292 (132 × 129 × 1). Las 1584 puntuaciones de impostor para el escenario de falsificación simulada se obtienen comparando las muestras de entrenamiento con las 12 falsificaciones simuladas para cada usuario (132 × 12).

Sistema on-line (Martinez-Diaz M., et al, 2014; Martinez-Diaz M., et al, 2015): una función basada en el algoritmo Dynamic Time Warping (DTW) (clasificado entre los tres algoritmos principales en las evaluaciones internacionales de tecnología (Malik, M.I., et al, 2013; Houmani, N., et al, 2012)). El algoritmo DTW (Martinez-Diaz M., et al, 2015) se aplica a funciones de secuencias de tiempo extraídas de cada firma. Un conjunto de siete funciones de tiempo se derivan de las secuencias [x, y, p]. Las secuencias se escogen después de la selección de características (según el rendimiento del conjunto de características) de un conjunto más amplio de secuencias definidas en (Martinez-Diaz M., et al, 2014). El algoritmo DTW coincide con dos conjuntos diferentes de secuencias en función de la distancia euclidiana entre las funciones de tiempo. El puntaje de clasificación se obtiene como la distancia promedio entre una firma de prueba y el conjunto de enrolamiento.

Sistema off-line (Galbally, J.,et al, 2015; Ferrer, M., et al, 2012): los patrones binarios locales (LBP) y los patrones direccionales locales (LDP) se utilizan para caracterizar las regiones de firma (12 bloques superpuestos para cada firma). La transformada de coseno discreta se aplica para reducir la dimensionalidad de los vectores de características y se entrenan dos clasificadores de máquina de vectores de soporte de mínimos cuadrados diferentes (LSSVM), utilizando cada uno de los conjuntos de características (características LBP y LDP). El puntaje final se calcula como la suma de los dos puntajes LSSVM, provenientes de cada uno de los clasificadores. El sistema off-line se aplica a las versiones de las muestras de firma off-line (la base de datos de BiosecurID (Fierrez J., et al, 2010) incluye las versiones on-line y off-line de las mismas muestras de firma).

En los experimentos realizados, T es igual a 4 (4 firmas para modelar al usuario) y I = 11 (los atributos A11–A13 comprenden diez mediciones). En el caso de los atributos categóricos (atributos A1-A9), se considera una distancia fija igual a 1, cuando la etiqueta del vector de características y el modo de los vectores de la galería (el valor más frecuente del atributo para este firmante) no son iguales. Por lo tanto, la distancia entre los atributos categóricos varía de 1 a 9 (número de atributos entre la muestra de la prueba y el conjunto de la galería con diferentes etiquetas). Ambas distancias (categórica y escalar) se normalizan de manera similar a la ecuación (2). El puntaje final se obtiene de la suma de ambas distancias.

Comparación del rendimiento basado en atributos: Los experimentos restantes tratan de evaluar el rendimiento de los atributos de firma anotados manualmente y la mejora obtenida cuando se combinan con los dos sistemas de autenticación de firmas detallado en (Morales A., et al, 2016). La base de datos de BiosecurID incluye información en línea (adquirida con una tableta digital), y la imagen estática de las firmas (escaneada a 150 ppp). Las imágenes estáticas se utilizan como entrada del sistema off-line, mientras que las secuencias en línea se utilizan como entrada del sistema on-line y la aplicación para la anotación de atributos (el anotador trabaja con una versión de la imagen sintética derivada de la secuencias [x, y, p]), utilizadas en los experimentos de anotación de atributos. La versión sintética es generada a partir de secuencias dinámicas, obtenidas de dispositivos digitales, que son muy comunes en aplicaciones reales como: puntos de ventas, bancos, entrega de paquetes, etc.

La Figura 4.6, y en la Tabla 4.4, presenta los resultados del promedio del rendimiento en términos de EER y FRR cuando FAR es igual al 10% en todos los anotadores, y el mejor rendimiento en término de EER y FRR cuando FAR es igual al 10% para el mejor anotador.

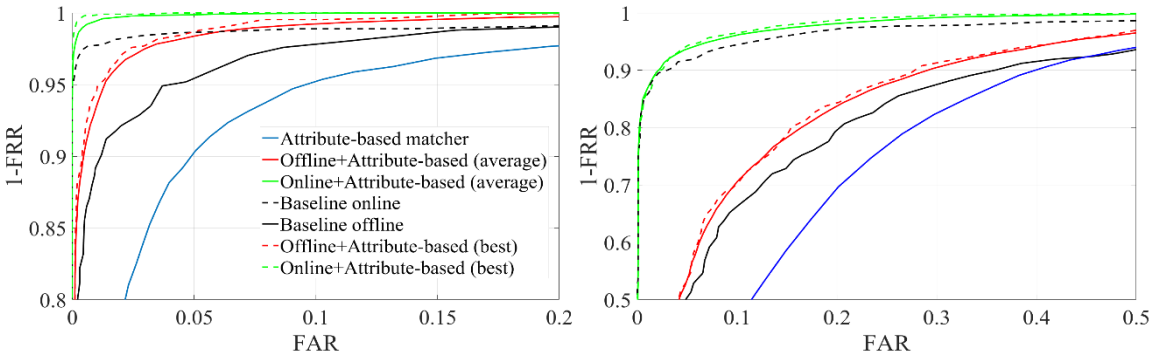


Fig. 4.6: Curvas ROC para diferentes falsificaciones aleatorias (izquierda) y simuladas (derecha) obtenidas por los diferentes sistemas evaluados (Morales A., et al, 2016).

El rendimiento obtenido en este experimento por el comparador propuesto basado en atributos, es similar al rendimiento obtenido por el sistema ASV off-line (baseline). El mejor rendimiento del comparador on-line se debe a la información

más discriminante disponible en las secuencias dinámicas en comparación con las características obtenidas de imágenes estáticas únicas (tanto los sistemas basados en atributos como los sistemas off-line que se basan en información estática).

Tabla 4.4: Rendimiento para los diferentes sistemas en la base de datos de BiosecurID (mejora con respecto a los sistemas de base agregados como subíndices) (Morales A., et al, 2016)

Sistema	EER, % Falsificaciones aleatorias		FRR (FAR = 10%) Falsificaciones aleatorias	
Sistema off-line (baseline)	4.72	20.27	2.13	34.31
sistema on-line (base)	1.85	6.85	1.21	6.12
Basado en atributos (promedio)	6.89	24.22	4.64	54.23
Basado en atributos (mejor anotador)	4.25	22.31	2.04	46.32
Off-line + atributos(promedio)	$2.63\downarrow_{44\%}$	$16.80\downarrow_{17\%}$	$0.84\downarrow_{60\%}$	$30.21\downarrow_{12\%}$
Off-line + atributos (mejor anotador)	$1.66\downarrow_{65\%}$	$15.55\downarrow_{23\%}$	$0.46\downarrow_{78\%}$	$29.80\downarrow_{13\%}$
On-line + atributos(promedio)	$0.72\downarrow_{61\%}$	$5.98\downarrow_{13\%}$	$0.10\downarrow_{92\%}$	$4.65\downarrow_{24\%}$
On-line + atributos (mejor anotador)	$0.20\downarrow_{89\%}$	$5.55\downarrow_{19\%}$	$0.01\downarrow_{99\%}$	$4.24\downarrow_{31\%}$

El siguiente análisis es explorar la complementariedad entre los sistemas de línea de base y el comparador propuesto basado en atributos. Las puntuaciones se normalizan de manera similar a la ecuación (2) y se combinan utilizando una suma ponderada. Las ponderaciones se seleccionan heurísticamente en función del rendimiento alcanzado en el experimento anterior: 0,8 × puntuación on-line + 0,2 × puntuación basada en atributos, y 0,8 × puntuación off-line + 0,2 × puntuación basada en atributos. La tabla 4.4, muestra los resultados y sugieren que el comparador propuesto basado en atributos se puede utilizar para mejorar significativamente el rendimiento de los sistemas de referencia en escenarios de falsificación aleatoria y simulada. En el escenario de comparación aleatoria, es posible observar mejoras desde el 44% (anotadores promedio, off-line + comparador basado en atributos) hasta el 90% (el mejor anotador, online + comparador basado en atributos). En el caso de falsificaciones simuladas, las mejoras van desde el 16% (anotadores promedio, on-line + basado en atributos) hasta el 23% (mejor anotador, off-line + basado en atributos).

Rendimiento de calificaciones: La comparación del desempeño humano por evaluaciones humanas es un protocolo estándar para la evaluación de esquemas asistidos por humanos (Jain, A.K., et al, 2005; Phillips, P.J., et al, 2015). Se propone un experimento de combinación de respuestas de profanos basadas en la regla de la suma a nivel de puntuación (se combinan las puntuaciones obtenidas por diferentes profanos). El análisis y los experimentos con 2 y 5 anotadores se repiten 50 veces (con selección aleatoria de anotadores) y el experimento con 10 anotadores se repite 11 veces. La Tabla 4.5, muestra los resultados promediados y la mejora obtenida por la combinación de profanos.

Tabla 4.5: Rendimiento que combina puntuaciones de diferentes números de anotadores (mejora con respecto a los sistemas de referencia agregados como subíndices) (Morales A., et al, 2016).

Sistema	# Anotadores	EER, %		FRR (FAR=10%)	
		Aleatorio	Simulado	Aleatorio	Simulado
Basado en atributos	1	6.89	24.22	4.64	54.23
Basado en atributos	2	$3.96\downarrow_{42\%}$	$21.14\downarrow_{13\%}$	$2.11\downarrow_{54\%}$	$42.96\downarrow_{21\%}$
Basado en atributos	5	$2.38\downarrow_{65\%}$	$18.66\downarrow_{23\%}$	$1.53\downarrow_{67\%}$	$33.43\downarrow_{38\%}$
Basado en atributos	10	$1.68\downarrow_{75\%}$	$18.21\downarrow_{24\%}$	$1.02\downarrow_{78\%}$	$31.78\downarrow_{41\%}$
Off-line + atributos	1	2.63	16.80	0.84	30.21
Off-line + atributos	2	$1.92\downarrow_{27\%}$	$15.75\downarrow_{6\%}$	$0.34\downarrow_{59\%}$	$22.48\downarrow_{25\%}$
Off-line + atributos	5	$1.65\downarrow_{37\%}$	$14.79\downarrow_{12\%}$	$0.27\downarrow_{68\%}$	$21.31\downarrow_{29\%}$
Off-line + atributos	10	$1.33\downarrow_{49\%}$	$13.98\downarrow_{17\%}$	$0.18\downarrow_{74\%}$	$20.09\downarrow_{33\%}$
On-line + atributos	1	0.72	5.98	0.10	4.65
On-line + atributos	2	$0.47\downarrow_{34\%}$	$5.88\downarrow_{2\%}$	$0.01\downarrow_{99\%}$	$3.81\downarrow_{18\%}$
On-line + atributos	5	$0.33\downarrow_{54\%}$	$5.61\downarrow_{6\%}$	$0.01\downarrow_{99\%}$	$3.23\downarrow_{30\%}$
On-line + atributos	10	$0.28\downarrow_{61\%}$	$5.57\downarrow_{6\%}$	$0.01\downarrow_{99\%}$	$2.88\downarrow_{38\%}$

Los resultados muestran que la complementariedad de las anotaciones hechas por diferentes humanos (profanos) tienen unas mejoras en términos de EER que varían de 27% a 75% para escenarios aleatorios y del 2% a 42% para escenarios de falsificación. Estas mejoras son aún mayores para la FRR con valores que van del 54% al 99% para escenarios aleatorios y del 18 al 41% para escenarios de falsificación. Como en experimentos anteriores, la mejora es mayor en aplicaciones off-line que en on-line. Las mayores tasas de error obtenidas en los sistemas off-line permitirán hacer una mejora del rendimiento.

4.3.- Atributos comparativos de la firma

El desempeño humano obtenido en (Morocho D., et al, 2016; Morocho D. et al, 2016; Morocho D. et al, 2016; Morocho D., et al, 2016; Reid D., et al, 2013) sugiere que a los humanos (profanos), les resulta difícil verificar correctamente la autenticidad de las firmas pero que su desempeño mejora si reciben una guía que marca los atributos en los que deben fijarse.

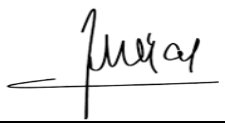
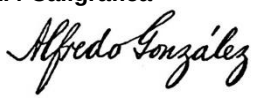
Los atributos absolutos constriñen la capacidad de abstracción humana y no siempre son fácilmente identificables a ojos de un no experto. Esto hace que la subjetividad de la medida sea alta. Así por ejemplo, cuando se debe definir el trazo de una firma como vertical u horizontal, las firmas que presenten trazos en ambas direcciones son difícilmente de etiquetar. Los atributos comparativos son aquellos, donde es necesario otorgar una ponderación numérica determinada dentro de un rango de valoración a través de la etiquetación manual por humanos. En lugar de un valor absoluto (SI/NO), se da una medida de la preponderancia de ese atributo en la firma. Además ofrecen respuestas más enriquecedoras al dar un valor de ponderación en un trazo de la firma. La idea de los atributos comparativos es reemplazar las etiquetas absolutas por rangos de ponderación (Ver Figura 4.7). Por ejemplo, utilizando atributos comparativos, una firma no está

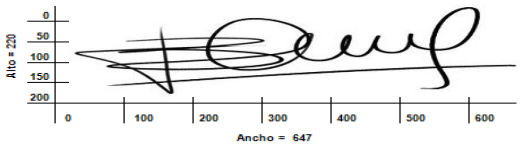
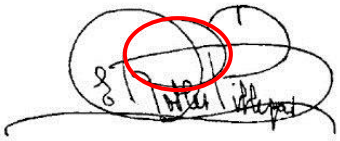
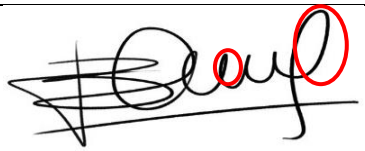
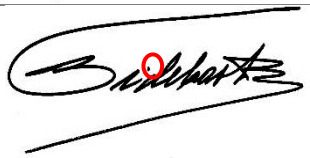
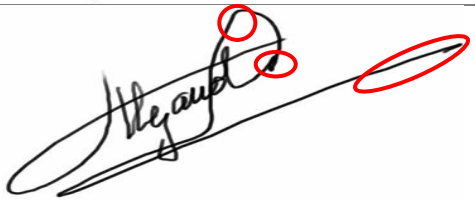
etiquetada como vertical, sino que está etiquetado, ¿cuánto de vertical es la firma?

Figura 4.7: Atributos absolutos en función de atributos comparativos (Morocho D., et al, 2017).

Se establece un rango de ponderación de 1 a 5, donde se incluyen características relacionadas con la forma, la puntuación, los retoques y los bucles, tanto para el texto como para el los adornos de la floritura de la firma La comparación basada en atributos se enfoca en la evaluación del rendimiento de 17 atributos de firma divididos en: 11 atributos comparativos (ver Tabla 4.2) provenientes de 5 características categóricas (C1: Forma, C2: Orden, C3: Proporcionalidad, C4: Lazos de Floritura, C5: Lazos de texto), que son etiquetados y ponderados con valores que van de 1 a 5, y 2 características categóricas con etiquetas binarias (A1: Puntuación e A2: Indecisión) y 4 características escalares (D1: Orientación, D2: Inclinación de caracteres, D3: Espaciado entre caracteres, D4: Longitud de trazos).

Tabla 4.6: Resumen de atributos comparativos.

Forma (C1)	**C1.1 Vertical** 	**C1.2 Redonda** 
	C1.3 Horizontal 	**C1.4 Caligráfica** 
Orden (C2)	**C2.1 Claro** 	**C2.2 Confuso** 
	C2.3 Concentrado 	**C2.4 Espaciado** 

Proporcionalidad (C3)	
Lazos de floritura (C4)	
Lazos de texto (C5)	
Puntuación (A1)	
Indecisión (A2)	
Orientación (D1)	
Inclinación de caracteres (D2)	
Espacio entre caracteres (D3)	
Longitud de trazos (D4)	

La figura 4.8, muestra ejemplos de etiquetación de atributos comparativos, donde se puede evidenciar la ponderación de evaluación, como es el caso: que tan redonda, o que tan horizontal es la firma. Para la ponderación hay que considerar que la intervención humana tiene una gran subjetividad causada por la percepción personal y la experiencia de cada anotador.

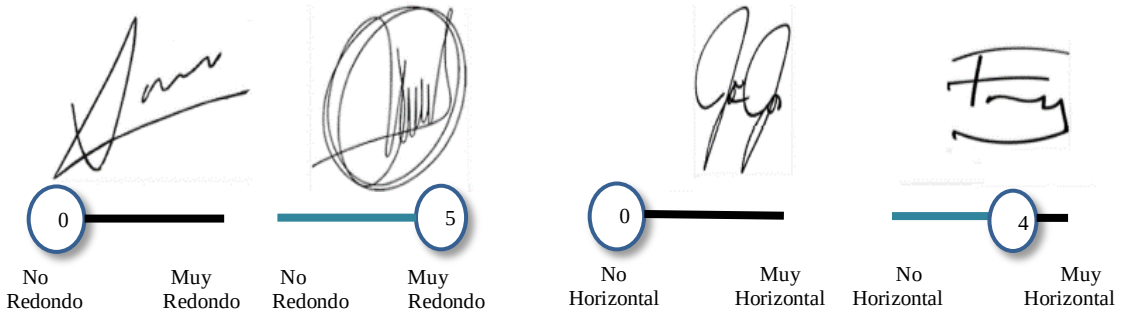


Figura 4.8: Ejemplos de etiquetación de atributos comparativos: que tan redonda es la firma (izq.), y que tan horizontal es la firma (der.) (Morocho D., et al, 2017).

4.3.1.- Interface de etiquetación manual de atributos de una firma

La Figura 4.9, muestra la interface de etiquetación manual de atributos comparativos, llamado Sys-HSL (System Handwritten Signatures Labeling), donde el profano (etiquetador) realiza el proceso de etiquetación de los distintos atributos de la firma de la base de datos BiosecurID, generando una nueva base de datos de los atributos etiquetados llamada Bio-HSL (Biometric Handwritten Signatures Labeling).

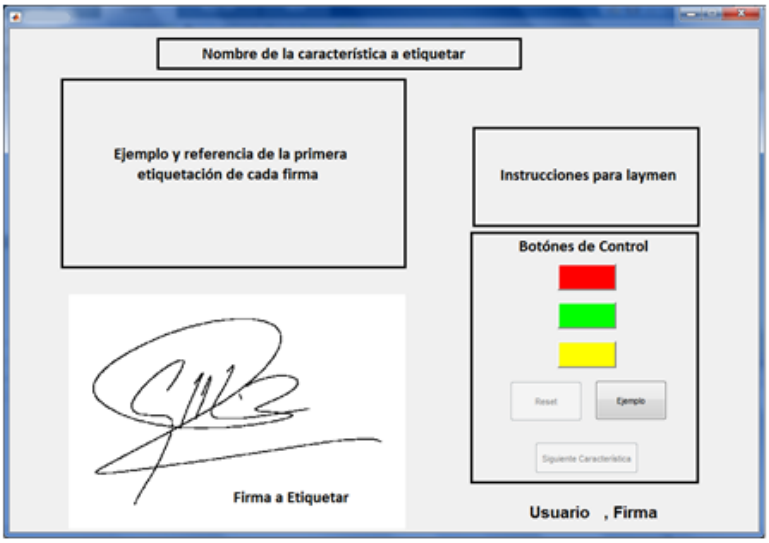


Figura 4.9: Aplicación Sys-HSL (Software de etiquetación de atributos comparativos).

4.3.2.- Comparación basada en atributos comparativos

La aplicación de etiquetación ofrece una breve introducción a cada uno de los atributos como ayuda para cada uno de los anotadores (humanos-profanos), por lo cual por su intervención hay que tener en cuenta la gran subjetividad causada por la percepción y la experiencia de cada anotador.

La comparación de los vectores de datos de las características se realiza aplicando la distancia Manhattan modificada (ver sección 4.2.4 ecuación (1)), y para la normalización de los scores obtenidos se aplica la normalización por tangente hiperbólica (ver sección 4.2.6, ecuación (2)) .

4.3.3.- Protocolos de experimentación y Base de datos Bio-HSL

El protocolo utilizado en este experimento son el mismo utilizado en el capítulo 4.2.5. La adquisición de la base de datos Bio-HSL (Biometric Handwritten Signatures Labeling), se cuenta con el subconjunto de 130 usuarios (firmantes, se excluyeron los dos últimos) de la base de datos BiosecureID-SONOF DB. La base de datos de atributos comprende más de 4.968.600 datos (21 anotadores × 130 firmantes × 28 muestras × 65 datos).

Corpus de la base de datos: Para la adquisición de la base de datos se contó con la colaboración de 21 personas sin experiencia en FDEs, no se proporcionó información sobre la autenticidad (genuina o falsificada) de las muestras al anotador y todas las firmas se analizaron por separado. Las características de este grupo de trabajo son:

- Género de los participantes: 8 mujeres y 13 hombres
- Edad: de 24 a 27 años
- Nivel académico: estudiantes de maestría de la Universidad de las Fuerzas Armadas ESPE.
- Experiencia con el tema: sin experiencia en FDE.

Cada etiquetador entrega en un archivo de 65 datos extraídos del proceso de etiquetación de cada firma etiquetada, donde la base de datos Bio-HSL, se compone de 4.968.600 datos (65 datos extraídos × 28 firmas × 130 usuarios × 21 etiquetadores).

4.3.4.- Experimentos y análisis de resultados de evaluación.

Para la ejecución de las evaluaciones de las firmas a través de la etiquetación se utiliza la base de datos BiosecurID-SONOF DB y el software de etiquetación manual Sys-HSL (System Handwritten Signatures Labeling). De la base de datos se tomó un subconjunto que está compuesto por 130 usuarios, donde cada usuario se compone de 28 firmas, de las cuales 16 son genuinas (G) y 12 falsificadas (F). El profano etiqueta un total de 3640 firmas presentadas a través de la interface en el orden que se indica en la tabla 4.7.

El desarrollo de la aplicación tiene por objetivo presentar al etiquetador o profano características basadas en el análisis extrínseco de la firma manuscrita, que son muy utilizados por los peritos forenses en el reconocimiento de firmas.

Tabla 4.7: Orden de presentación de las firmas al etiquetador (profano)

#Firma	Sesión 1	#Firma	Sesión 2	#Firma	Sesión 3	#Firma	Sesión 4
1	G	8	G	15	G	22	G
2	G	9	G	16	G	23	G
3	F	10	F	17	F	24	F
4	F	11	F	18	F	25	F
5	F	12	F	19	F	26	F
6	G	13	G	20	G	27	G
7	G	14	G	21	G	28	G

4.3.4.1.- Funcionalidad del Sistema de Etiquetación Manual

La aplicación Sys-HSL, está compuesta de etapas que cumplen con una funcionalidad como:

- **Lectura de Firmas:** Los datos de las firmas son cargados de la Base de Datos BiosecureID SONOF DB.

- **Pre-procesamiento de firmas:** Se realiza el pre-procesado de los datos y carga el último ertiuetado más uno, además se realiza el análisis de las coordenadas cartesianas tanto en x,y obtenidas de la lectura de la Base de datos de firmas manuscritas.

- **Etiquetación de características:** para el proceso de etiquetación por parte del profano se deben tomar en cuenta ciertas consideraciones:
  - Se presentan 11 características (categóricas y escalares) de la firma en un total de 15 pantallas de interface.
  - En cada pantalla de la interface, aparece un botón de EJEMPLO, que muestra más información de la característica.

- En la interface siempre se visualiza el número de usuario y el número de firma que se está etiquetando.
- Las características categóricas tendrán una ponderación valorada de [1, 2, 3, 4, 5].
- Para las características escalares, se visualiza la primera etiquetación como referencia para las posteriores etiquetaciones del mismo usuario, con el fin de tener resultados óptimos.
- La primera etiquetación de las características de medición se guarda en una imagen tipo jpg.
- La interface no cambia a una nueva pantalla, si el profano no etiqueta los atributos de la firma de forma completa

➢ **Almacenamiento de Datos Etiquetados:** Este proceso permite almacenar los atributos etiquetados por el profano y generar la base de datos Bio-HSL:

- Finalizada la etiquetación de las 11 características, estos datos se guardan.
- Si el profano cierra el programa y posteriormente lo vuelve a ejecutar, este inicia desde la última firma etiquetada completamente.

➢ **Diagrama de flujo del programa Sys-HSL:** El funcionamiento de la etiquetación a través del programa Sys-HSL.

4.3.4.2.- Experimentos de evaluación

A través de la aplicación de etiquetado se genera una nueva base de datos Bio-HSL (Biometric-Handwritten Signature Labels), que contiene datos etiquetados por 21 profanos. Cada profano etiqueta 13 atributos (A1- A7) más 10 atributos escalares (D1 más 3 medidas para cada D2, D3 y D4) para cada una de las 3,640 firmas en la base de datos (130 × 28). Por lo tanto, la base de datos comprende 1.758.120 atributos: 23 atributos × 28 firmas por firmante × 130 firmantes × 21 profanos. Los experimentos de evaluación se dividen en los dos escenarios presentados con anterioridad (Falsificaciones Aleatorias y Simuladas).

Para el protocolo experimental, se emplea el protocolo propuesto para la comparación de atributos absolutos (Morales A., et al, 2016). Utilizamos nuevamente las 4 muestras genuinas de la primera sesión como conjunto de entrenamiento. La distancia entre la matriz de entrenamiento (matriz con 4 × 23 atributos) y una firma dada (vector con 1 × 23 atributos) se calcula utilizando la distancia de Manhattan propuesta con anterioridad.

El rendimiento en términos de EER aleatoria y simulada, permite determinar el potencial de los atributos comparativos. En esta base de datos se cuenta con un mayor número de etiquetadores, lo cual permite un mejor análisis particularizado de los mismos. La Figura. 4.10, muestra el rendimiento humano (profano), donde los resultados muestran el desempeño de cada uno de los profanos en términos

de EER promedio para firmas aleatorias y simuladas. Los resultados muestran que el 38% de los profanos presentan un EER aleatorio por debajo del 5%, donde el mejor EER es 3.90% y el peor EER es 10.32%. Para las falsificaciones simuladas, tenemos que el 52% de los profanos tienen un EER simulado de menos del 21% con el mejor EER del 18.83% y el peor EER del 26.22%.

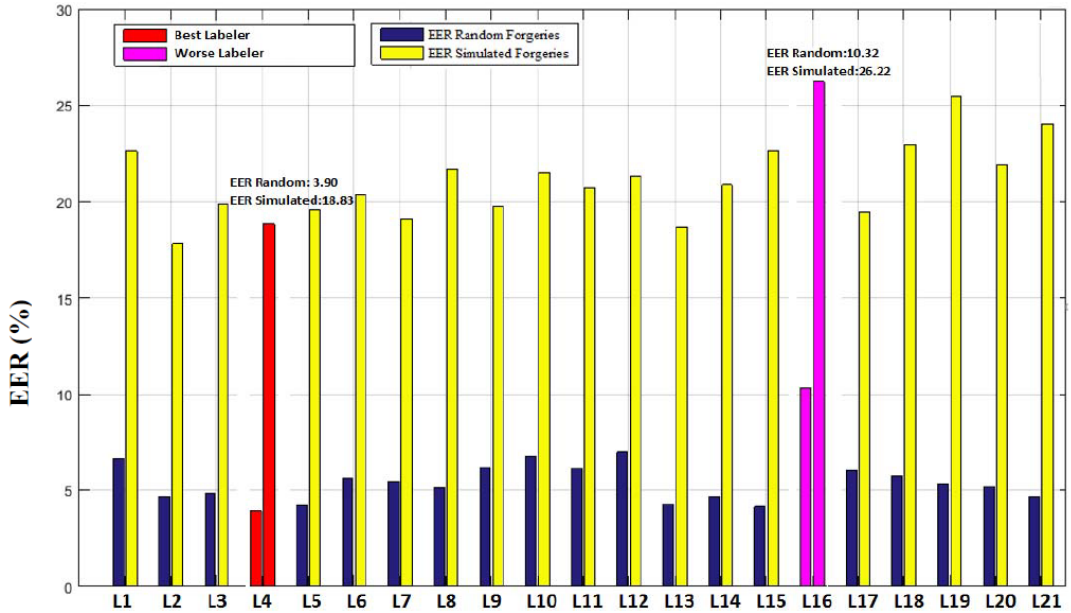


Figura 4.10: Rendimiento de Profano: el mejor y el peor anotador se resaltan con un color diferente (Morocho D., et al, 2017).

El análisis de los resultados sugiere que las características escalares son más discriminatorias para firmas aleatorias que para firmas simuladas, como se puede apreciar en la Figura 4.11, es menos compleja diferencias una firma aleatoria de una firma genuina, que una firma simulada de una firma genuina. Si se considera el atributo de distancia entre caracteres (D3), una firma simulada es muy similar a una genuina, evidenciando los mismos caracteres, y se pueden obtener medidas de distancia entre caracteres muy similares entre ellas. En cambio una firma aleatoria está compuesta de otros caracteres diferentes a la firma genuina, evidenciando distancias entre caracteres diferentes entre ellas.

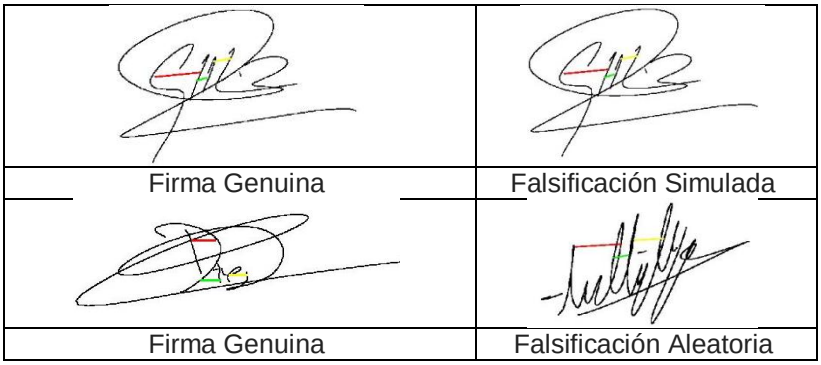


Figura 4.11 Análisis del atributo D3 para firmas genuinas versus falsificación simulada y aleatoria.

La Tabla 4.8, muestra los resultados obtenidos promediando todos los rendimientos del anotador y los resultados obtenidos en trabajos anteriores utilizando la misma base de datos y el protocolo experimental. Los resultados sugieren el rendimiento superior de los atributos comparativos frente a los atributos absolutos. Sin embargo, existe una gran brecha con los sistemas automáticos, especialmente con los ASV on-line. Es importante destacar que el rendimiento de los métodos basados en atributos está altamente relacionado con los profanos. Los resultados mostrados en la Tabla 4.8, se promediaron a partir de 21 profanos en el caso de atributos comparativos y 11 profanos en el caso de atributos categóricos (profanos diferentes en cada caso). Las diferencias en los rendimientos (1% en falsificaciones aleatoria y 3% para falsificaciones simuladas) son estadísticamente significativas para las 384,930 puntuaciones calculadas para las comparaciones aleatorias (12 × 130 × 21 + 129 × 130 × 21) y las 65,520 puntuaciones calculadas para el falsificaciones simuladas (12 × 130 × 21 × 2).

Tabla 4.8: Evaluación de desempeño humano (EER%) a nivel de Atributos Absolutos y Comparativos vs. Sistemas Automáticos

Sistema	EER [%]	
	Aleatorio	Simulado
ASV basado en características On-line (Morales A., et al, 2016)	1.9	6.9
ASV basado en características Off-line (Morales A., et al, 2016)	4.72	20.27
Semi-Automático: Atributos Absolutos (Morales A., et al, 2016)	6.89	24.22
Semi-Automático: Atributos Comparativos (Morocho D., et al, 2017b)	**5.57**	**21.20**

4.4.- Conclusiones y Contribuciones

En este capítulo de la tesis, se muestra las conclusiones y las diferentes contribuciones, a través de artículos de investigación publicadas en diferentes congresos relacionadas con atributos etiquetados por profanos.

4.4.1.- Conclusiones

- El rendimiento humano se mejora claramente cuando se le da una guía que permita enfocar su análisis de las firmas.
- Los atributos absolutos ofrecen un rendimiento similar a los sistemas automáticos offline y ayudan a mejorar los sistemas automáticos.
- Los atributos comparativos ofrecen mejor rendimiento que los absolutos.

4.4.2.- Contribuciones

En este capítulo de la tesis, se muestra las diferentes contribuciones, a través de artículos de investigación publicadas en diferentes congresos relacionadas con atributos etiquetados por profanos.

Articulo Revista	**A. Morales, D. Morocho, J. Fierrez, R. Vera. Signature authentication based on human intervention: performance and complementarity with automatic systems. IET Biometrics, vol. 6(4), pp. 307-315, 2017. (Journals IET Biometrics, The Institution of Engineering and Technology (IET), Special Issue: Selected Papers from the International Workshop on Biometrics and Forensics)**
Contribución	Esta investigación evalúa la intervención humana a nivel de función basada en atributos inspirados en el trabajo de los FDE. Los experimentos incluyen 11 anotadores diferentes, 3696 firmas y más de 800,000 atributos etiquetados. Los resultados sugieren el potencial de las capacidades humanas para mejorar los sistemas de autenticación automática tanto en aplicaciones sin conexión como en línea. La combinación de intervención basada en atributos y sistemas ASV a nivel de puntuación y las mejoras van del 16 al 90% según el escenario. Los resultados revelan nuevas perspectivas sobre cómo los seres humanos se desempeñan en la autenticación de firmas, y algunas formas en que los sistemas ASV pueden mejorarse con la intervención humana. Para aplicaciones prácticas se debe tener un conjunto reducido de las características más discriminatorias, ya sea que se etiqueten de forma automática o manual en un corto período de tiempo. El poder discriminativo de la información dinámica de la firma podría utilizarse para aumentar las diferencias entre muestras genuinas y falsificadas.

Articulo Congreso	**D. Morocho, A. Morales, J. Fierrez and R. Vera-Rodriguez. Human-Assisted Signature Recognition based on Comparative Attributes. International Conference on Document Analysis and Recognition (ICDAR) e International Workshop on Human-Document Interaction, (HDI), vol. 8, pp 5-9, 2017b**
Contribución	Este artículo analiza el rendimiento de los atributos comparativos etiquetados manualmente por personas para el reconocimiento de firmas. Los atributos comparativos intentan explotar las capacidades de los humanos para extraer información discriminante de las firmas. En lugar de atributos absolutos (por ejemplo, ¿este trazo es vertical?), Los atributos comparativos ofrecen respuestas más completas (por ejemplo, ¿qué tan vertical es este trazo?). Quince atributos comparativos son etiquetados por 21 profanos, donde cada uno lleva a cabo el etiquetado de 28 firmas de 130 usuarios de la base de datos de BiosecurID, generando una nueva base de datos Bio-HSL, que contiene 4,968,600 atributos. Los resultados muestran que los atributos comparativos superan a los atributos absolutos en el reconocimiento de firma semiautomático con tasas de error que van desde 5.5% para comparaciones aleatorias hasta 21.2% para falsificaciones.

Articulo Congreso	**D. Morocho, A. Morales, J. Fierrez, and J. Ortega-Garcia. Humans in the loop: Study of semi-automatic signature recognition based on attributes". In Proceedings of the 51st International Carnahan Conference on Security Technology (ICCST'2017), IEEE, pp 1–5, Madrid, Spain, 2017a**
Contribución	Este trabajo explora el desempeño de los seres humanos en el reconocimiento de firmas basado en la intervención en diferentes niveles. La literatura muestra que los FDE pueden lograr rendimientos similares a los mejores sistemas de verificación automática de firmas de última generación. El desempeño de los profanos está lejos del desempeño obtenido por los sistemas FDE y ASV. Sin embargo, las intervenciones de colaboración basadas en el crowdsourcing, así como los atributos comparativos, han mostrado resultados alentadores. Los resultados muestran el gran potencial de estos esquemas y su potencial en aplicaciones que involucran intervenciones humanas.

Capítulo 5

5.- Conclusiones y Trabajo futuro

Esta Tesis se basa en la investigación y el análisis de como la intervención humana puede ayudar a los sistemas de autenticación de firma.

5.1.- Conclusiones

1. La intervención humana a nivel de clasificación, considera un análisis de cómo los humanos se desempeñan en las tareas de autenticación de firma. Los experimentos basados en el análisis de la respuesta de 500 personas (workers), a través del crowdsourcing, permiten ayudar a establecer un "baseline" por humanos. Los resultados muestran que los workers se desempeñan peor que los sistemas ASV y resaltan las dificultades asociadas a esta tarea. La tasa de error promedio de los workers es de alrededor del 30%, pero las opiniones agregadas muestran el potencial de las capacidades humanas cuando se combinan las respuestas de diferentes workers.

2. El potencial de plataformas como Mturk en la investigación de reconocimiento biométrico es grande y se pueden desarrollar nuevos conocimientos sobre la base de tareas de colaboración masivas de seres humanos.

3. La intervención humana a nivel de función basada en atributos inspirados en trabajos de FDEs. Se realizó experimentos con 2 grupos de anotadores utilizando la misma base de datos y el protocolo experimental: en primera instancia con 11 anotadores diferentes, etiquetando 3696 firmas y obteniendo más de 800,000 atributos etiquetados. En segunda instancia con 21 anotadores diferentes, etiquetando 3640 firmas y obteniendo más de 1.7 millones de atributos etiquetados. Cabe destacar que el rendimiento de los métodos basados en atributos esta altamente relacionado con los profanos. Las diferencias en los rendimientos del $EER_{Aleatoria}$ del 1% y del $EER_{Simulada}$ del 3%, son estadísticamente significativas para las 384,930 puntuaciones calculadas para las falsificaciones aleatorias, y las 65,520 puntuaciones calculadas para las falsificaciones simuladas. Los resultados evidencian que el rendimiento de los atributos comparativos es mejor que los atributos absolutos.

4. Los experimentos realizados en esta tesis han generado la base de datos Bio-HSL, que contiene datos etiquetados por 21 profanos, donde cada profano etiqueta 23 atributos: 11 atributos comparativos (C1.1-C1.4, C2.1-C2.4, C3, C4, C5), 2 atributos categoricos (A1, A2), y 10 atributos escalares (D1, y 3 medidas por cada atributo D2, D3 y D4) para cada una de las 3,640 firmas, generando una base de datos de 1.758.120 atributos.

5. Los resultados sugieren que las etiquetas comparativas ofrecen información más discriminante que los atributos absolutos. Los resultados reportados en este trabajo revelan nuevas perspectivas sobre cómo los seres humanos se desempeñan en la autenticación de firmas, y algunas formas en que los sistemas ASV pueden mejorarse con la intervención humana.

5.1.- Trabajo Futuro

La información generada en esta Tesis, presenta futuras investigaciones en esta área, consideradas de interés por el autor.

Esta investigación genera inquietudes de trabajo futuro para seguir con esta línea de investigación, que permitan ayudar a los sistemas automáticos de verificación de fitma:

1. ¿Cuál es la consistencia de los atributos anotados aplicados a otra base de datos?

2. ¿Pueden los atributos comparativos mejorar el reconocimiento automático en esquemas combinados?

3. ¿Cuál es la estabilidad de los anotadores para diferentes números de firmas?

Esta inquietudes están alineados a la continuidad de esta investigación, donde intenta mejorar los sistemas de verificación de firma a través de la intervención del humano.

Referencias

L. C. Araujo, L. H. Sucupira, M. G. Lizárraga, L. L. Ling, and J. B. T. Yabu-Uti. User authentication through typing biometrics features. IEEE transactions on signal processing, vol. 53, no. 2, pp. 851-855, 2005.

L. Best-Rowden, S. Bisht, J. C. Klontz and A. K. Jain. Unconstrained face recognition: Establishing baseline human performance via crowdsourcing. IEEE International Joint Conference on, pp. 1-8, 2014a.

L. Best-Rowden, H. Han, C. Otto, B. F. Klare and A. K. Jain. Unconstrained face recognition: Identifying a person of interest from a media collection. IEEE, vol. 9, no. 12, pp. 2144-2157, 2014b.

C. Bird, B. Found, and D. K. Rogers. Forensic Handwriting Examiners' Skill in Detecting Disguise Behavior from Handwritten Text Samples. vol. 22, 2012.

V. Blankers, C. van den Heuvel, K. Franke, and L. Vuurpijl. The ICDAR 2009 signature verification competition. In Proc. Int. Conf. Document Analysis and Recognition, vol. 3, pp 1403–1407, 2009.

L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. Apparel classification with style. in Proc. 11th ACCV, pp. 321–335, 2013.

M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's Mechanical Turk: A new source of inexpensive, yet highquality, data. Perspectives on psychological science, vol. 6, no. 1, pp. 3-5, 2011.

T. M. Burkes, D. P. Seiger and D. Harrison. Handwriting examination: Meeting the challenges of science and the law. Forensic Science Communications, vol. 11, no. 4, 2009.

D. M. Burt and D. I. Perrett. Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and color information. Proc. Biol Sci., vol. 259, no. 1355, pp. 137–143, 1995.

H. Chen, A. Gallagher, and B. Girod. Describing clothing by semantic attributes. in Proc. 12th ECCV, pp. 609–623, 2012.

S.E. Choi, Y.J. Lee, S.J. Lee, K.R. Park, and J. Kim. Age Estimation Using a Hierarchical Classifier Based on Global and Local Facial Features. Pattern Recogn., vol. 44, no. 6, pp.1262-1281, 2011.

J. Coetzer. Off-line signature verification. Ph.D. Thesis, University of Stellenbosch , 2005.

J. Coetzer, B.M. Herbst, and J.A. Du Preez. Off-line signature verification: A comparison between human and machine performance. Proc. 10th Int. Workshop on Frontiers in Handwriting Recognition, La Baule, France, pp. 481-485, 2006.

J. Coetzer, B.M. Herbst, and J.A. du Preez. Offline Signature Verification Using the Discrete Radon Transform and a Hidden Markov Model. Eurasip Journal on Applied Signal Processing - Special Issue on Biometric Signal Processing, H. Bourland, I. Pitas, K.K. Lam, and Y. Wang, editors, vol. 2004, no. 4, pp. 559-571, 2004.

H. Coetzer and R. Sabourin. A human-centric off-line signature verification system. Proc. Int. Conf. on Document Analysis and Recognition, Curitiba, Brazil, pp. 153-157, 2007.

A. Dantcheva, C. Velardo, A. D'Angelo, and J. L. Dugelay. Bag of soft biometrics for person identification. Multimedia Tools Appl., vol. 51, no. 2, pp. 739–777, 2011.

Y. Deng, P. Luo, C. C. Loy, and X. Tang. Pedestrian attribute recognition at far distance. in ACMMM. ACM, 2014.

M. Ferrer, J. Vargas, A. Morales, A. Ordonez. Robustness of offline signature verification based on gray level features. IEEE, Trans. Information, Forensics and Security, 7(3):966-977, 2012.

J. Fierrez, J. Galbally, J. Ortega-Garcia, M. R. Freire, F. Alonso-Fernandez, D. Ramos and E. Anguiano. BiosecurID: A Multimodal Biometric Database. Pattern Analysis and Applications. Springer, vol. 13, no. 2, pp. 235-246, 2010.

J. Fierrez and J. Ortega-Garcia. Handbook of Biometrics, chapter On-line signature verification. Eds. A. K. Jain, A. Ross and P. Flynn, Springer, pp. 189-209, 2008

K. Franke, L. Schomaker, C. Veenhuis, L. Vuurpijl, I. Erp, and M. van Guyon. WANDA: A common ground for forensic handwriting examination and writer identification. ENFHEX News, pp. 23-47, 2004.

A. Gilperez, F. Alonso-Fernandez, S. Pecharroman, J. Fierrez, and J. Ortega-Garcia, "Off-line Signature Verification using Contour Features. in: ICFHR, 2008.

H. Han, C. Otto, X. Liu and A. K. Jain. Demographic Estimation from Face Images: Human vs. Machine Performance. in IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 37, no. 6, pp. 1148-1161, 2015.

A. Hassaine, S. Al-Maadeed, J. Alja'am, A. Jaoua, and A. Bouridane. The ICDAR2011 Arabic Writer Identification Contest. In Document Analysis and Recognition (ICDAR), pp. 1470 –1474, 2011.

N. Houmani, A. Mayoue, S. Garcia-Salicetti, B. Dorizzi, M. I. Khalil, M. N. Moustafa, H. Abbas, D. Muramatsu, B. Yanikoglu, A. Kholmatov, M. Martinez-Diaz, J. Fierrez, J. Ortega-Garcia, J. Roure Alcobe, J. Fabregas, M. Faundez-Zanuy, J. M. Pascual-Gaspar, V. Cardenoso, Payo, and C. Vivaracho-Pascual. BioSecure signature evaluation campaign (BSEC'2009): Evaluating online signature algorithms depending on the quality of signatures. Pattern Recognition, vol. 45, no.3, pp. 993-1003, 2012.

J. Howe. The Rise of Crowdsourcing. Wired, vol. 14, no.6, 2006.

D. Impedovo and G. Pirlo. Automatic signature verification: The state of the art. IEEE Trans. on Systems, Man, and Cybernetics (Part C), vol. 38, no. 5, pp. 609-635, 2008.

E. Jaha and M. Nixon. Soft Biometrics for Subject Identification using Clothing Attributes. In IEEE International Joint Conference on Biometrics, pp. 1-6, 2014a.

E. S. Jaha and M. S. Nixon. Analysing soft clothing biometrics for retrieval. in Proc. BIOMET, pp. 234–245, 2014b.

E. S. Jaha and M. S. Nixon. Viewpoint invariant subject retrieval via soft clothing biometrics. in Proc. ICB, pp. 73–78, 2015.

E. S. Jaha and M. S. Nixon. From Clothing to Identity: Manual and Automatic Soft Biometrics. IEEE Trans. Inf. Forensics Secur., vol. 11, no. 10, pp. 2377–2390, 2016.

K. Jain, S. C. Dass and K. Nandakumar. Soft biometric traits for personal recognition systems. Conf. Biometric Authentication. Hong Kong, 2004a.

K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. Pattern Recognit, vol. 38, no. 12, pp. 2270–2285, 2005.

K. Jain, K. Nandakumar and A. Ross. 50 years of biometric research: Accomplishments, challenges, and opportunities. El Sevier, vol. 79, pp. 1-26, 2016.

K. Jain, A. A. Ross and K. Nandakumar. Introducción a la biometría. Boston, 2011.

K. Jain , A. Ross and S. Prabhakar. Una introducción al reconocimiento biométrico. IEEE, vol.1, no. 14, pp. 4-20, 2004b.

A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with Mechanical Turk. Proc. of the SIGCHI conference on human factors in computing systems, pp. 453-456, 2008.

F. Klare et al. Suspect Identification Based on Descriptive Facial Attributes. Proc. of International Joint Conference on Biometrics, Clearwater, Florida, USA, pp. 1-8, 2014.

N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Describable visual attributes for face verification and image search. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 33, no. 10, pp.1962–1977, 2011.

F. Leclerc and R. Plamondon. Automatic signature verification: The state of the art 1989–1993. Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 8, no. 3, 643–660, 1994.

S. Liu, J. Feng, Z. Song, T. Zhang, H. Lu, and C. Xu, et al. Hi magic closet, tell me what to wear!. In Proc. ACM MM12, pp. 619–628, 2012a.

S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In Proc. CVPR, pp. 3330–3337, 2012b.

M. Liwicki, M. I. Malik, L. Alewijnse, C. E. van den Heuvel, and B. Found. ICFHR2012 Competition on Automatic Forensic Signature Verification (4NsigComp 2012). in Frontiers in Handwriting Recognition (ICFHR), pp. 819-824, 2012.

M. Liwicki, M. I. Malik, C. E. Van den Heuvel, X. Chen, C. Berger, R. Stoel, M. Blumenstein, and B. Found. SigComp11: Signature verification competition for on- and offline skilled forgeries. in Document Analysis and Recognition (ICDAR), pp. 1480-1484, 2011.

M. Liwicki, C. E. Van den Heuvel, B. Found, and M. I. Malik. Forensic signature verification competition 4NSigComp2010 - detection of simulated and disguised signatures. in Frontiers in Handwriting Recognition (ICFHR), pp. 715–720, 2010.

M. D. MacLeod, J. N. Frowley, and J. W. Shepherd. Whole body information: Its relevance to eyewitnesses. in Adult Eyewitness Testimony. Cambridge, U.K.: Cambridge Univ. Press, 1994.

M. I. Malik, M. Liwicki, and A. Dengel. Part-based automatic system in comparison to human experts for forensic signature verification. Proc. Int. Conf. on Document Analysis and Recognition, Washington DC, USA, pp. 872–876, 2013a.

M. I. Malik, M. Liwicki, A. Dengel, and B. Found. Man vs. Machine: A Comparative Analysis for Forensic Signature Verification. Proc. of the 16th International Graphonomics Society Conference, pp. 9–13, 2013b.

M. I. Malik, and M. Liwicki. From Terminology to Evaluation: Performance Assessment of Automatic Signature Verication Systems. in: ICFHR, pp. 609-614, 2012.

M. Martinez-Diaz and J. Fierrez. Signature databases and evaluation. Springer, pp. 1367-1375, 2015a.

M. Martinez-Diaz, J. Fierrez and S. & Hangai. Signature matching. Springer, pp. 1192-1196, 2009.

M. Martinez-Diaz, J. Fierrez and S. Hangai. Signature matching. Springer, pp. 1382-1387, 2015b.

M. Martinez-Diaz, J. Fierrez, and R.P. Krish, et al. Mobile signature verification: feature robustness and performance comparison. IET Biometrics, vol. 3, no.4, pp. 267–277, 2014.

D Martinho-Corbishley, M Nixon. Super-Fine "Attributes with Crowd Prototyping. IEEE Transactions, 2018.

D. Martinho-Corbishley, M. Nixon, and J. Carter. Soft Biometric Recognition from Comparative Crowdsourced Annotations. Accepted Proc. Int. Conf. on Imaging for Crime Detection and Prevention, ICDP, pp. 1–6, 2015.

C. A. Meissner and J. C. Brigham. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. Psychology, Public Policy, and Law. vol. 7, no. 1, pp. 3–35, Mar. 2001.

M. Minear and D. Park. A Lifespan Database of Adult FacialStimuli. Behavior Research Methods, Instruments, and Computers, vol. 36, no. 4, pp. 630-633, 2004.

A. Morales, D. Morocho, J. Fierrez and R. Vera-Rodriguez. Signature authentication based on human intervention: performance and complementarity with automatic systems. IET Biometrics, IEEE, vol. 6, no. 4, pp. 307-315, 2017.

D. Morocho, J. Hernandez-Ortega, A. Morales, J. Fierrez and J. Ortega-Garcia. On the evaluation of human ratings for signature recognition. In Security Technology (ICCST), IEEE, pp. 1-5, 2016a.

D. Morocho, A. Morales, J. Fierrez and R. Tolosana. Signature recognition: establishing human baseline performance via crowdsourcing. In Biometrics and Forensics (IWBF), IEEE, pp. 1-6, 2016b.

D. Morocho, A. Morales, J. Fierrez and J. Ortega-Garcia. Humans in the loop: Study of semi-automatic signature recognition based on attributes. IEEE, pp. 1-5, 2017a.

D. Morocho, A. Morales, J. Fierrez and R. Vera-Rodriguez. Towards human-assisted signature recognition: improving biometric systems through attribute-

based recognition. IEEE International Conference on Identity, Security and Behavior Analysis, pp. 1-6, 2016c.

D. Morocho, A. Morales, J. Fierrez and R. Vera-Rodriguez. Human-Assisted Signature Recognition Based on Comparative Attributes. In Document Analysis and Recognition (ICDAR), IEEE, vol.8, pp. 5-9, 2017b.

D. Morocho, M. Proaño, D. Alulema, A. Morales and J. Fierrez. Signature Recognition: Human performance analysis vs. automatic system and feature extraction via crowdsourcing. Proc. Mexican Conf. on Pattern Recognition, Springer International Publishing, pp. 324-334, 2016d.

C. Ng, Y. Tay, and B. Goi. Recognizing human gender in computer vision: a survey. in PRICAI. Springer, pp. 335–346, 2012.

V. Nguyen, and M. Blumenstein. An application of the 2d Gaussian filter for enhancing feature extraction in off-line signature verification. in: ICDAR, pp. 339-343, 2011.

L. Oliveira, E. Justino, C. Freitas, and R. Sabourin. The graphology applied to signature verification. Proc. 12th Conf. of the Int. Graphonomics Society, Salerno, Italy pp. 286-290, 2005.

J. Ortega-Garcia, et al. The Multi-Scenario Multi-Environment BioSecure Multimodal Database (BMDB). IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 32, no. 6, pp. 1097-1111, 2010.

A. J. O'Toole, P. J. Phillips, F. Jiang, J. Ayyad, N. P´enard, and H. Abdi. Face recognition algorithms surpass humans matching faces over changes in illumination. IEEE Trans. PAMI, vol. 29, no. 9, pp. 1642–1646, 2007.

A. J. O'Toole, D. A. Roark, and H. Abdi. Recognizing moving faces: A psychological and neural synthesis. TRENDS in Cognitive Sciences, vol. 6, no. 6, pp.261–266, 2002.

S. Panjwani, and A. Prakash. Crowdsourcing attacks on biometric systems. In: Symposium On Usable Privacy and Security (SOUPS 2014), Menlo Park, California, pp. 257–269, 2014.

M. Philipp. Fakten zu FISH. das forensische informations-system handschriften des bundeskriminalamtes eine analyse nach ber 5 jahren wirkbetrieb, Tech. rep., Bundeskriminalamt, Germany, in German, 1996.

P.J. Phillips, M.Q. Hill, and J.A. Swindle, et al. Human and algorithm performance on the PaSC face recognition challenge. Proc. Int. Conf. On Biometrics: Theory, Applications and Systems, Arlington, USA, pp.1–8, 2015.

R. Plamondon, and G. Lorette. Automatic signature verication and writer identification the state of the art. in: Pattern Recognition, vol. 22, pp. 107-131, 1989.

R. Plamondon and S. N. Srihari. On-line and off-line handwriting recognition: A comprehensive survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, pp. 63-84, 2000.

D. Reid and M. Nixon. Human Identification using Facial Comparative Descriptions. Proc. Int. Conf. on Biometrics, 2013.

D. Reid, M. Nixon and S. V. Stevenage. Soft Biometrics; Human Identification using Comparative Descriptions. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 36, no. 6, pp. 1216-1228, 2014.

M. Rhodes. Age estimation of faces: A review. Appl. Cognit. Psychol, vol. 23, no. 10, pp. 38–59, 2009.

K. Ricanek and T. Tesafaye. MORPH: A longitudinal image database of normal adult age-progression. in Proc. Int. Conf. Automat. Face Gesture Recognit, pp. 341–345, 2006.

P. Samangouei, V. M. Patel and R. Chellappa. Attribute-based Continuous User Authentication on Mobile Devices. Proc. Int. Conf. on Biometrics: Theory, Applications and Systems, Washington DC, USA, pp. 1-6, 2015.

P. Samangouei, V. M. Patel, R. Chellappa, D. Chandra and B. Barbello. Continuous user authentication on mobile devices: Recent progress and remaining challenges. IEEE Signal Processing Magazine, vol.33, no.4, pp. 49-61, 2016.

Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a semantic representation for person re-identification and search. in Proc. CVPR, pp. 4184–4193, 2015.

J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. in RASC, 2002.

P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. Proc. IEEE, vol. 94, no. 11, pp. 1948–1962, 2006.

J. Sita, B. Found, and D. Rogers. Forensic handwriting examiners expertise for signature comparison. Journal of Forensic Sciences, vol. 47, pp.1117-1124, 2002.

S. N. Srihari, B. Zhang, C. Tomai, S. Lee, Z. Shi, and Y. C. Shin. A system for handwriting matching and recognition. in: Symp. on Document Image Understanding Technology, pp. 67-75, 2003.

Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In Proc. CVPR, 2014.

Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In Proc. CVPR, 2014.

M. Toews and T. Arbel. Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 9, pp. 1567–1581, 2009.

R. Tolosana, R. Vera-Rodriguez, R. Guest, J. Fierrez and J. Ortega-Garcia. Complexity-based Biometric Signature Verification. in Proc. 14th IAPR Int. Conference on Document Analysis and Recognition, ICDAR, Kyoto, Japan, 2017.

P. F. J. Tome, R. Vera-Rodriguez and M. S. Nixon. Soft biometrics and their application in person recognition at a distance. IEEE Transactions on Information Forensics and Security, vol. 9, no. 3, pp. 464-475, 2014.

D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk. Attribute-based people search in surveillance environments. in WACV, 2009.

K. Yamaguchi, T. Okatani, K. Sudo, K. Murasaki, and Y. Taniguchi. Mix and match: Joint model for clothing and attribute recognition. in Proc. BMVC, pp. 51.1–51.12, 2015.

D.Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll. SVC2004: First International Signature Verification Competition. ICBA vol. 2004, LNCS 3072, Springer-Verlag Berlin Heidelberg, D. Zhang and A. K. Jain, Eds., pp. 16-22, 2004.

M. B. Yilmaz, B. Yanikoglu, C. Tirkaz, and A. Kholmatov. Offline signature verification using classifier combination of HOG and LBP features. In Biometrics (IJCB), pp. 1-7, 2011.

L. Zheng, Y. Yang, and A. G. Hauptmann. Person re-identification: Past, present and future. arXiv preprintarXiv: 1610.02984, 2016.

J. Zhu, S. Liao, Z. Lei, D. Yi, and S. Li. Pedestrian attribute classification in surveillance: Database and evaluation. In ICCV workshop on Large-Scale Video Search and Mining, 2013