

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-27

Identificación de Firmas Dinámicas por Métodos no Lineales usando Máquinas de Vectores de Soporte

Fredy Andrés Torres Muñoz



INSTITUTO TECNOLÓGICO METROPOLITANO

FACULTAD DE INGENIERÍAS

DEPARTAMENTO DE ELECTRÓNICA Y TELECOMUNICACIONES

MEDELLÍN, COLOMBIA

2015

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Identificación de Firmas Dinámicas por Métodos no Lineales usando Máquinas de Vectores de Soporte

Fredy Andrés Torres Muñoz

Proyecto de grado para optar al título de

Ingeniería Electrónica

Director

Prof., Jorge Alberto Jaramillo Garzón, PhD.



INSTITUTO TECNOLÓGICO METROPOLITANO

FACULTAD DE INGENIERÍAS

DEPARTAMENTO DE ELECTRÓNICA Y TELECOMUNICACIONES

MEDELLÍN, COLOMBIA

2015

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Dynamic Signature Identification by Non-linear Methods using Support Vector Machines.

Fredy Andrés Torres Muñoz

Graduation Project for the degree of
Ingeniería Electrónica

Advisor

Prof., Jorge Alberto Jaramillo Garzón, PhD.



INSTITUTO TECNOLÓGICO METROPOLITANO
FACULTY OF ENGINEERING
DEPARTMENT OF TELECOMMUNICATIONS AND ELECTRONICS
MEDELLÍN, COLOMBIA

2015

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

“Siempre se es moderno, en toda época, desde el momento en que uno piensa poco más o menos como sus contemporáneos y de forma un poco distinta que sus maestros”

A. Koyré

“Es de suponer que quién descubrió tantas verdades admirables en el mundo sensible no habría estado muy descarriado en el mundo intelectual”

Voltaire

*Mide lo que puedas medir,
y lo que no se pueda medir, hazlo medible.*

Galileo Galilei

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

AGRADECIMIENTOS

Gracias a Dios, por darme la sabiduría e inteligencia para la realización de este proyecto.

Agradezco enormemente a todas las personas que de una u otra manera hicieron posible la terminación de este proyecto.

Primero, quiero agradecer a mi madre por su apoyo incondicional en todos los retos que he enfrentado en la vida. Con entrega, esfuerzo y humildad, uno a uno los he ido superando. Y sin duda alguna, por su crianza y afecto me ha hecho el hombre que soy hoy en día.

No menos importante extendiendo mi agradecimiento a mi familia, y amigos que me rodean. También quiero agradecer al doctor, magister, ingeniero pero ante todo, profesor Jorge Jaramillo Garzón quien es director de este trabajo y que por sus infinitos consejos, durante este tiempo me han ayudado a crecer como persona y como futuro profesional. También darles las gracias por enseñarme e introducirme en el área de reconocimiento de patrones. Su conocimiento e ideas han hecho que para muchas personas sea un ejemplo a seguir. Por todo esto y por su amistad, mi agradecimiento y gratitud.

Del mismo modo quiero agradecer a la línea de investigación MIRP por todo su apoyo y colaboración a lo largo del desarrollo de este trabajo, especialmente a Leonardo Duque Muñoz, Delio Augusto Aristizábal Martínez, Norma Patricia Guarnizo Cutiva, Hermes Alexander Fandiño Toro, Germán David Góez Sánchez, María Constanza Torres, Luis Roberto Mercado, Julián Arango.

¡Muchas gracias, Dios los bendiga a todos!

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

TABLA DE CONTENIDO

TABLA DE CONTENIDO	10
LISTAS DE FIGURAS.....	12
LISTAS DE TABLAS.....	13
RESUMEN.....	15
1. INTRODUCCIÓN	17
Justificación	20
Planteamiento del Problema.....	22
Hipótesis.....	26
Objetivos del trabajo de grado.....	27
2. MARCO TEÓRICO	28
2.1 Biometría	28
2.1.1 Firmas Dinámicas (On-line).....	29
2.1.2 Tipos de sistemas biométricos.....	32
2.1.3 Evaluación de los sistemas biométricos	33
2.2 Análisis de dinámica no lineal.....	34
2.2.1 Dimensión de Correlación ($D2$)	36
2.2.2 Entropía de Kolmogorov ($K2$).....	37
2.2.3 Entropía de Shannon (Hx).....	38
2.2.3 Exponente de Hurst (H).....	38
2.2.4 Complejidad de Lempel–Ziv (LZ).....	39
2.3 Máquinas de Vectores de Soporte (SVM).....	39
2.3.1 Caso Linealmente Separable.....	40
2.3.2 Caso Linealmente no Separable	42
2.3.3 Espacio de alta dimensionalidad – Kernels.....	43
2.3.4 Optimización por Enjambres de Partículas.....	45
3. MARCO EXPERIMENTAL	47
3.1. Base de Datos	49
3.1.1. Base de Datos (ITM-MIRP-SIGN-01 – ITMMS01)	49

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3.2.	Estimación de Características	52
3.3.	Selección de los parámetros para SVM	53
3.3.1.	Selección y Sintonización del Kernel Gaussiano (γ).....	54
3.3.2.	Factor de penalización de error.....	55
3.4.	Técnica de Evaluación.....	55
3.4.1	Validación Cruzada con K – Particiones.	56
3	EXPERIMENTOS Y RESULTADOS.....	59
4	CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO	65
	REFERENCIAS.....	67

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

LISTAS DE FIGURAS

<i>Figura 1.1. Análisis Biométrico.....</i>	<i>28</i>
<i>Figura 1.2: Ejemplos de aplicaciones biométricas. (Tomada de Jain et al., 2004).....</i>	<i>29</i>
<i>Figura 1.3: Series temporales extraídas de una firma dinámica. (Tomada de Garcia-Salicetti et al., 2003).....</i>	<i>31</i>
<i>Figura 1.4. Ajuste para una SVM lineal.</i>	<i>39</i>
<i>Figura. 1.5. Aparición del parámetro de error ξ_i en el error de clasificación. (Tomada de Betancourt A., 2005).</i>	<i>43</i>
<i>Figura. 1.6. Uso de un tipo de Kernel para la transformación del espacio de representación de los datos. (Tomada de Betancourt A., 2005).</i>	<i>44</i>
<i>Figura.3.1. Tablet gráfica utilizada para la adquisición de las firmas. (Tomada de Gaspar Pascual, 2010).</i>	<i>48</i>
<i>Figura. 3.2. Diagrama de bloques de la Metodología utilizada en este trabajo.....</i>	<i>49</i>
<i>Figura. 3.3. Series de tiempo extraídas de la firma dinámicas. (Tomada de Ahmedt-Aristizabal et al., 2011).....</i>	<i>51</i>
<i>Figura 3.4. Ventana de la interfaz desarrollada para la visualización de las firmas dinámicas y de las series temporales.</i>	<i>52</i>
<i>Figura. 3.5. Metodología de Clasificación.</i>	<i>58</i>
<i>Figura. 3.6. Esquema general con el enfoque propuesto para la clasificación de las firmas Dinámicas.....</i>	<i>58</i>
<i>Figura. 4.1. Resultados de la Clasificación en la identificación de firmantes.....</i>	<i>62</i>
<i>Figura. 4.2. Desempeño obtenido por el sistema propuesto, haciendo una organización con las características no-lineales más relevantes para la identificación de firmantes.....</i>	<i>64</i>
<i>Figura. 4.3. Desempeño obtenido por el sistema propuesto, haciendo una organización con los atributos más relevantes para la identificación de firmantes.....</i>	<i>65</i>

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

LISTAS DE TABLAS

<i>Tabla 3.1. Resumen de la composición de la base de datos utilizada.</i>	<i>51</i>
<i>Tabla.4.1. Clasificación Resultados de la firma, en términos de medidas de rendimiento del clasificador considerado (a_{gm}, a_{sen}, a_{esp}).</i>	<i>60</i>
<i>Tabla.4.2. Matriz de Confusión. (Kohavi and Provost, 1998).</i>	<i>61</i>

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

RESUMEN

Los sistemas de identificación de firmas dinámicas, se utilizan para la medición y análisis de las características del ser humano (biometría), y están basados en el reconocimiento de características comportamentales, en este caso, características que tienen que ver con el movimiento intrínseco de la mano al momento de ejercer o realizar una firma manuscrita. Estos sistemas son muy utilizados y pueden desempeñarse en entornos de seguridad (Dessimoz et al., 2007), entre otros. Por consiguiente, se puede relacionar la teoría del caos con la dinámica desarrollada en un individuo para procesar las series de tiempo extraídas de la dinámica de la firma, series como; la presión, la velocidad en los ejes (x, y), la inclinación del trazo y la altitud.

Este proyecto de investigación se centra en el diseño de una metodología para la identificación cerrada de firmas dinámicas usando Máquinas de Vectores de Soporte (SVM). La extracción de características se basa en técnicas de dinámica no lineal o en medidas de complejidad como: el Exponente de Hurst, la Entropía de Shannon, la Entropía de Kolmogorov, la complejidad de Lempel-Ziv y Lyapunov. Para la evaluación de la metodología se utiliza una base de datos construida por la línea de investigación MIRP, la cual está constituida por 800 firmas que representan 40 firmantes o individuos. La validación de los resultados se realiza usando técnicas de validación cruzada (cross validation). Finalmente se realiza un análisis estadístico con los resultados obtenidos de la metodología propuesta, para determinar si la utilización de las características fractales permite obtener una mejor generalización de las firmas dinámicas en estudio.

Palabras clave: Análisis fractal, firmas dinámicas, identificación, series de tiempo, SVM.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

1. INTRODUCCIÓN

Los sistemas de seguridad son un importante campo de investigación para las organizaciones e instituciones que necesitan reconocimiento personal. La Biometría permite la identificación personal basada en características fisiológicas o de comportamiento y constituye una de las soluciones más eficaces en términos de seguridad (Dessimoz et al., 2007). Entre la biometría, la firma presenta algunas ventajas: amplia aceptación, de uso común en las transacciones legales y comerciales y de uso de los dispositivos relativamente baratos (Faundez-Zanuy, 2007), (Jain et al., 1999).

La creciente demanda de aparatos electrónicos en reconocimiento de personas está basado en rasgos de tipo fisiológicos (Huella, iris, geometría de la mano, cara entre otros) por mostrar algunos ejemplos (acceso a ordenadores portátiles por reconocimiento de cara y huella dactilar).

Por otro lado, los sistemas de reconocimiento conductuales tales como la voz, la firma e inclusive el modo de caminar, todavía siguen siendo temas de investigación que desafortunadamente no se refleja el esfuerzo en aplicaciones prácticas para los ciudadanos.

La identificación biométrica de firmas dinámicas se basa en la suposición de que la firma de cada individuo es única y puede estar asociada con patrones específicos. Las variaciones biométricas conductuales dependen de la persona, tal como una firma puede dar a una amplia serie de factores que impiden una identificación fiable (Pascual-Gaspar et al., 2011). Por lo tanto, surge la necesidad de utilizar técnicas de caracterización no-lineales o análisis Fractal para la interpretación fiable de la dinámica fisiológica del firmante. Sin embargo estos espacios de representación no-lineal son distribuidos por estructuras complejas derivadas del comportamiento psicomotor. Se ha encontrado que las habilidades psicomotoras pueden ser controladas por los sistemas dinámicos absolutamente complejos cuyas propiedades y estructuras aún no se han descrito y

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

comprendido completamente (Longstaff y Heath, 1999). Por lo tanto, es posible suponer que el desempeño del análisis teórico de los sistemas dinámicos no lineales puede proporcionar conceptos nuevos y más potentes para analizar el comportamiento real de los sistemas complejos, como los biológicos (Solé et al., 1996).

Ha habido numerosas investigaciones para la autenticación de la firma (Plamondon y Lorette, 1989). En particular, matching; utilizando Deformación Dinámica de Tiempo (DTW) (Plamondon y Lorette, 1989) es la técnica más ampliamente estudiada de autenticación de firma on-line, mientras que los Hidden Markov Models (HMMs) (Nanni and Lumini, 2006) se han convertido en los modelos estadísticos con mejores resultados para la verificación de firmas online (Dessimoz et al., 2007).

Sin embargo, DTW es computacionalmente costoso y el proceso de remuestreo por lo general resulta en la pérdida de importantes detalles locales, de modo que las firmas falsificadas coincidan estrechamente con las firmas auténticas. A su vez, debido a la naturaleza estadística de HMM, el número de datos de entrenamiento es generalmente bastante alto para aplicaciones prácticas.

Propuestas recientes muestran algunas mejoras mediante la combinación de técnicas (Faundez-Zanuy, 2007), fusionando información local y global (Fierrez-Aguilar et al., 2005) o mediante la combinación de diferentes salidas del sistema (Nanni y Lumini, 2006). Sin embargo, excepto en algunos casos aislados, el elevado número de características usadas conduce a una mayor demanda computacional, así como el aumento de las necesidades de almacenamiento. Además, cuanto mayor es la complejidad de las características, mayor será el coste de adquisición del dispositivo.

Como la información biométrica de la firma contiene operadores no lineales embebidos, el uso de técnicas de análisis dinámicos no lineales proporciona un rendimiento más fiable de identificación biométrica. Sin embargo, muy pocos estudios se han publicado en esta materia (Ahmedt-Aristizábal et al., 2011) (Ahmedt et al., 2011) y se centró en la caracterización, sin tener en cuenta la fase de clasificación. Por lo tanto, no proporcionan un sistema de identificación completamente funcional.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

En el presente trabajo, se propone una metodología para la identificación biométrica de firmas dinámicas representadas por características fractales, tales como el exponente de Hurst, dimensión de correlación, la complejidad de Lempel-Ziv, la entropía de Shannon y la entropía Kolmogorov, y evaluado por una máquina de vectores de soporte multi-clase (SVM), que fueron estimadas a partir de la series de tiempo correspondiente a la presión, el acimut, altitud, y las posiciones horizontales y verticales de cada firma (Ahmedt-Aristizábal et al., 2011). La evaluación muestra resultados satisfactorios de sensibilidad y especificidad con respecto a las medidas utilizadas en una base de datos llamada ITM-MIRP-SIGN-01, que se compone por 800 firmas recogidas de 40 voluntarios diferentes y recolectadas con una tableta digitalizadora Wacom Intuos 4 (Ahmedt-Aristizabal et al., 2011). Por lo tanto, la integración de las representaciones fractales con la capacidad de las máquinas de vectores de soporte para representar los datos en espacios de alta dimensionalidad a través de las funciones kernel, permitirá descifrar los operadores fractales embebidos en las dinámicas fisiológicas de los firmantes y obtener herramientas de identificación con alta precisión.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Justificación

Como parte de nuestra vida cotidiana, los seres humanos deseamos vivir en paz y tranquilidad. Por tal razón, el hombre ha buscado la manera de innovar y mejorar instrumentos que garanticen su seguridad en su entorno y así de alguna manera sentirse más seguro (Jain et al, 2004).

En la identificación automática de personas entra a jugar un papel muy importante la biometría, por utilizar métodos confiables (Plamondon & Lorette, 1989) para identificación en entornos que tienen que ver con seguridad. La biometría es un conjunto de métodos que se basan en las características biológicas para identificar o comprobar la analogía de un ser humano.

Hoy en día con la constante mejora de los sistemas, se busca siempre la posibilidad de garantizar el mayor grado la seguridad en todos nosotros. La implementación de sistemas que garanticen la seguridad en la sociedad, conlleva a utilizar las metodologías de identificación de personas. Siendo una de estas la firma (Plamondon & Lorette, 1989). Donde un individuo es identificado por características de comportamiento (comportamentales) que varían del estado de ánimo de la persona y por el movimiento psicomotor de la muñeca en la medida que se ejerce la firma.

La firma de una persona es muy usual para la identificación y verificación de su identificación, esto con el fin de establecer la autenticidad de la persona en un documento o para establecer la certificación o autorización del mismo; que implican transacción, desistimiento y, en general, disposición de derechos, que requiera de la constancia de conformidad ante una situación legal o de compromiso.

Se ha considerado a la firma por sus características conductuales al momento de realizarla por el firmante, como una característica biométrica de fiabilidad en la identificación de un individuo o para la verificación de su personalidad; sin embargo, otras personas ajenas a

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

ella, deseen suplantarla para obtener algún beneficio, realizando una falsificación de la misma con rasgos más o menos parecidos para su cometido. Ante esta situación se requiere de una forma de autenticar que la firma obtenida de un usuario, sea de la persona que dice ser, no una copia o falsificación y para ello, se desea crear una metodología que acorde a las señales conductuales al momento de realizarla, se pueda realizar dicha tarea con un buen grado de confianza y certeza de la autenticidad o falsificación o copia de la misma, favoreciendo en muy buena cantidad el sector comercial y gubernamental que utiliza en gran medida documentos en medio escrito y magnético como acción legal.

El gobierno colombiano en la Ley 962 de 2005 más conocida como la Ley Anti trámites, en el artículo 6. Medios magnéticos PARÁGRAFO 3. Dice: “Cuando la sustanciación de las actuaciones y actos administrativos se realice por medios electrónicos, las firmas autógrafas que los mismos requieran, podrán ser sustituidas por un certificado digital que asegure la identidad del suscriptor, de conformidad con lo que para el efecto establezca el Gobierno Nacional.” Y en el artículo 24. Presunción de validez de firmas. <Modificado por el Art 36 Decreto 0019 de 10-01-2012> dice: “Las firmas de particulares impuestas en documentos privados, que deban obrar en trámites ante autoridades públicas no requerirán de autenticación. Dichas firmas se presumirán que son de la persona respecto de la cual se afirma corresponden. Tal presunción se desestimará si la persona de la cual se dice pertenece la firma, la tacha de falsa, o si mediante métodos tecnológicos debidamente probados se determina la falsedad de la misma.”

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Planteamiento del Problema

Un sistema biométrico se orienta hacia el reconocimiento de formas y rasgos fisiológicos que son inherentes de cada ser humano, y que no cambian con el paso del tiempo, como por ejemplo, el iris, las huellas dactilares, y la simetría de las manos. Otros tipos de características que se pueden considerar en un sistema biométrico son los rasgos conductuales, los cuales se relacionan con el comportamiento de las personas, y están directamente relacionadas con el habla, la escritura manuscrita y la firma (Ahmedt-Aristizabal & et al., 2011). La verificación automática de la firma es un área de investigación bien establecida y activa, con numerosas aplicaciones, por medio del cual se logra el reconocimiento de personas a través de sus rasgos intrínsecos comportamentales. Este tipo de sistemas son importantes en entornos donde la seguridad es vital, porque son sinónimo de confianza y de privacidad, siendo su principal función la verificación e identificación de personas en un sistema que puede ser abierto o cerrado.

Las técnicas usadas para el reconocimiento biométrico se basan principalmente en mediciones geométricas, matemáticas y en métodos basados en inteligencia artificial, por ejemplo (J. Fierrez, Ortega-García & et al., 2007). En efecto, la mayoría de los sistemas corresponden a sistemas cerrados, que no pueden tratar con muestras que no hayan estado representadas en el conjunto de datos de entrenamiento (Ahmedt-Aristizabal & et al., 2011) (Schmidt, Riffo, Mery, 2011) y otros.

En los sistemas cerrados, por su lado, es común utilizar como medida de desempeño el porcentaje de aciertos del sistema respecto al total de identificaciones realizadas, es decir, el porcentaje de firmas que fueron correctamente asignadas a cada individuo o clase (Pascual-Gaspar, 2011), con el fin de construir fronteras de decisión sobre las clases que se encuentren en un espacio de características representativas en un conjunto de N clases, y así obtener un sistema óptimo y confiable en su funcionamiento, teniendo en cuenta que, una nueva clase se encuentre previamente registrada en la base de datos.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Diversos métodos han sido descritos para la adquisición de datos dinámicos (firmas on-line). (Nelson & Kishon, 1991; Hamilton et al., 1995) utilizan la sensibilidad de la presión adquirida de la tableta digitalizadora mientras que (Sakamoto et al., 2001; Tham et al., 2001) utilizan una pluma equipada con un sensor de presión y un sistema de condicionamiento de señal que puede extraer las características de la distribución de presión de la firma manuscrita.

Se puede apreciar variedad de propuestas en el tipo de características dinámicas que se pueden estimar. (Kato & Mawashima, 1993) utilizan posición de la pluma y presión, mientras que (Taguchi et al., 1988) utiliza la inclinación de la pluma y (Yoshimura, 1991) utiliza para la verificación la dirección del movimiento de la pluma. El algoritmo propuesto por (Jin et al., 1992), calcula las distancias entre la entrada y la base de datos para cada trazo, sin embargo se presentan inconvenientes cuando el número de trazos varía. Además (Sakamoto et al., 2001) utiliza las trayectorias de posición de la pluma, la presión y la inclinación de manera combinada, no obstante los resultados fueron satisfactorios para los caracteres Japoneses y no fueron probados para el alfabeto occidental.

Algunas metodologías combinan características dinámicas y estáticas (Jin et al., 2005). , e inclusive otras propuestas utilizan la presión, velocidad y altura del lápiz, (Hangai et al., 2000), velocidad del baricentro (Thumwarin & Matsuura, 2004), modelos de puntos de distribución (Tsai, 2005), y programación dinámica continua (Kameya & Oka, 2003) para la estimación de características.

Se han implementado diversos algoritmos aplicados al pre procesamiento de los datos dinámicos adquiridos. Algunos de ellos utilizan normalización (normalization) (Ma et al., 2000), codificación de predicción lineal (linear prediction coding) (Wu et al. 1997), dinámica en tiempo de deformación (dynamic time warping) (Martens & Claesen, 1996, 1997) árbol de coincidencias (tree matching) y reducción de ruido (noise reduction) (Dullink et al., 1995), suavizado de datos (smoothing of data) (Hastie et al., 1992), y tipos de segmentación (segmentation) (Clark et al., 1990; Shafiei, & Rabiee, 2003). Es necesario combinar más de uno de los algoritmos mencionados anteriormente para el pre-

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

procesamiento de una señal de la firma. Por ejemplo en (Hastie et al., 1992) se aplicó normalization, dynamic time warping, smoothing of data y segmentation para procesar la información adquirida.

Entre las investigaciones que se han abordado en identificación de firmas en sistemas dinámicos, se encuentra:

En el trabajo presentado por (Nanni, 2005), se realiza una comparación experimental de sistemas basados en clasificadores de una sola clase. A partir del trabajo de (D.M.J. Tax, 2001), se propone un nuevo enfoque que amplía el trabajo previo con respecto a la fusión de máquinas expertas basado en características globales. Las máquinas expertas que se utilizaron fueron los siguientes: Modelos de Descripción Gaussiana, Mezcla de Descripción Gaussianas, Análisis de Descripción de Componentes Principales -PCAD-, Support vector data description SVD, Descripción de programación lineal (LPD), Clasificador Ventana de Parzen (PWC). Los resultados obtenidos con respecto a la base de datos MCYT, que ha sido ampliamente utilizada en esta área (Ortega-Garcia, Fierrez-Aguilar & et al., 2003), presentan una fusión entre los clasificadores (PWC) y (PCAD). Con esta fusión, obtienen el error más bajo superando drásticamente los trabajos existentes en el estado del arte.

La investigación desarrollada por (Kumar & Unikrishnan, 2013) propone una caracterización de atributos extraídos de las firmas dinámicas, basada en el método de coeficiente de varianza o coeficiente de variación, donde las características con varianzas pequeñas son seleccionadas y las características con varianza grandes son rechazadas. La metodología se denomina Técnica de Selección de Características de Clase Específica reduciendo al Mínimo el Coeficiente de Varianza (CSFMCOV), y constituye una medida normalizada de dispersión de un conjunto de datos probabilísticos, propuesta como soporte para verificar la consistencia de la distribución.

(Zhang & et al., 2007) abordan una metodología en firmas off-line, basadas en agrupación de características de dirección, texturas, dinámica y complejidad. Estas características

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

fueron extraídas de imágenes y en sus totalidades clasificadas o verificadas por una red neuronal feedforward totalmente conectada con el algoritmo de aprendizaje de retropropagación clásico, conocido comúnmente como la Red de retropropagación (BPN). Entre los resultados obtenidos muestran que el sistema es eficaz con una tasa de error promedio del 1,8%.

(Nanni & Lumini, 2005) diseñan un método basado en un trabajo previo (Fierrez-Aguilar & et al, 2005), donde se detalla el conjunto completo de 100 funciones globales. Por medio de la cual se utiliza una simple función de clasificación que estima las características que maximizan la distancia de los centroides entre los diferentes firmantes introducido en el trabajo previo. Para cada característica se calcula la distancia de Mahalanobis entre la media de los patrones de entrenamiento de cada firmante para un total de N clases y el conjunto de todas las firmas de entrenamiento. La distancia de Mahalanobis entre un patrón x ($\mu_{(c,k)}$) en su caso y un conjunto S de patrones con media μ_s y matrix de covarianza Cov_S es dado por:

$$dM(x, s) = (x - \mu_s)^T Cov_S^{-1} (x - \mu_s) \quad (1)$$

Las características se clasifican de acuerdo a la siguiente clase de separabilidad entre usuarios:

$$DistF(k) = \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} (dM_{(i,k)} + dM_{(j,k)}) \quad (2)$$

Después, dado el conjunto de entrenamiento de uno de los firmantes, compuesto por vectores de características de sus firmas, una estimación no-paramétrica de la función de densidad de probabilidad es obtenida mediante el uso del Clasificador de la Ventana de Parzen.

Los resultados experimentales se presentan con respecto a la base de datos (J. Ortega-Garcia, J. Fierrez-Aguilar, et al., 2003), y permite obtener una mejora relativa en el

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

desempeño de verificación de hasta 21% (Para falsificadores Calificados) y 37% (Para Falsificadores Aleatorios) en comparación con los trabajos del estado de arte. La base de datos consistió en 100 firmantes con 25 firmas genuinas y 25 falsificaciones de calidad, al firmante-falsificadores se proporcionan las imágenes de la firma de los clientes para ser falsificados y, después de entrenar con ellos varias veces, se les pide imitar la forma con la dinámica natural, es decir, sin descansos o desaceleraciones.

En las investigaciones mencionadas es claro que los métodos desarrollados emplean un criterio de caracterización con características globales y vectores de distancias. Sin embargo, se evidencia la no utilización de técnicas de dinámica no lineal a la cual este proyecto se enfoca.

Las investigaciones desarrolladas anteriormente están enfocadas a proyectos que involucran diferentes clasificadores de una sola clase en identificación de firmas dinámicas, pero la aplicación a las que están orientadas, no resultan útil para la realización de este proyecto de clasificación de firmas dinámicas con análisis fractal utilizando Máquinas de vectores de soporte en sistemas cerrados. Por esta razón se ha encontrado poca implementación en metodologías en donde se empleen análisis fractal, en el área de la identificación de firmas dinámicas. Por consiguiente este proyecto de investigación busca expandir el estado del arte con respecto a la metodología planteada.

Hipótesis

Teniendo en cuenta que las series de tiempo contienen información del proceso psicomotriz ejercido por el sujeto, por ésta razón es de suma importancia utilizar la trayectoria contenida en el sistema para inferir sobre la dinámica desconocida de la serie de tiempo. Es posible encontrar un desempeño más eficiente en identificar firmas dinámicas por análisis fractal o técnicas de dinámica no lineal.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Objetivos del trabajo de grado

Objetivo General

Desarrollar una metodología para la identificación de firmas dinámicas, utilizando características fractales; relacionando la representación dinámica de la firma con análisis no lineal en sistemas caóticos. Usando una Máquina de Vectores de Soporte multi-clase para la evaluación e identificación de la metodología propuesta.

Objetivos Específicos

- Caracterizar los atributos comportamentales de cada firmante, mediante técnicas de dinámica no lineal o características fractales con el fin de realizar la identificación de las firmas.
- Estimar la estrategia de clasificación basada en el clasificador SVM, e interpretar los resultados obtenidos por clase pertenecientes a los firmantes.
- Evaluar la estrategia y la metodología de clasificación por medio de SVM, para validar la propuesta de investigación.

La mayoría de las simulaciones se llevaron a cabo en el entorno de R estadística (R Core Team, 2012). Todos los paquetes son distribuidos libremente bajo la Licencia Pública General GNU. Y MATLAB (MATLAB, 1984). Todos los toolbox son distribuidos bajo la Licencia MathWorks, Inc.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

2. MARCO TEÓRICO

Se presenta en este capítulo un análisis teórico de la metodología basada en características fractales para la identificación de firmas dinámicas. Con miras a facilitar el proceso que conlleva en este trabajo. Se describen conceptos fundamentales de la biometría, análisis fractal y el aprendizaje de máquina; que para este caso, es la máquina de vectores de soporte (SVM). Como primer paso se presenta la descripción de las señales que forman parte de una firma dinámica.

Se presentan para el análisis de las señales de las firmas diversas técnicas, orientadas al reconocimiento del firmante, el cual exige que las características seleccionadas entreguen información suficiente para su identificación. La extracción de características es una de las etapas de la cual depende en gran medida el buen desempeño del sistema de reconocimiento, por tal razón, se requiere abarcar en detalle temáticas como son la reconstrucción de las medidas de complejidad, y pruebas de estructuras no lineales deterministas. En segundo paso, explicar analíticamente el método utilizado para la clasificación y la estrategia utilizada para la validación del clasificador SVM.

2.1 Biometría

Un sistema biométrico es un sistema que puede ser automático, y su finalidad es la verificación e identificación de personas o individuos utilizando características tanto físicas como comportamentales, que la mayoría de las veces es usado como un sistema de monitoreo o vigilancia para entornos de seguridad. Para el análisis que realiza los sistemas biométricos se utilizan técnicas basadas en inteligencia artificial, estadísticas y matemáticas, entre otras.

La definición de un sistema biométrico podría decirse como un sistema de reconocimiento de patrones que opera mediante la adquisición de datos biométricos de un individuo, la extracción de un conjunto de características a partir de los datos adquiridos, y la

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

comparación de este conjunto de características contra la plantilla situada en la base de datos (Jain et al., 2004). Dentro de un sistema biométrico existen dos tipos de rasgos pertenecientes a un individuo para su autenticación biométrica. Entre ellos, se encuentra los rasgos fisiológicos. Estos rasgos son los que poseemos los seres humanos por nacimiento y que al pasar de los años no cambian o sufren alguna modificación física, y si ocurre; es debido a una enfermedad desarrollada a lo largo de la vida de la persona, e.g., la huella dactilar, el iris, la retina, la palma de la mano, entre otros. Por otro lado tenemos los rasgos conductuales, los cuales están más relacionados con el comportamiento de las personas y que dependen del estado de ánimo del individuo. Dentro de esta clase de rasgos conductuales se encuentran la firma, la voz, la forma de caminar son ejemplos de este tipo de rasgos, como se aprecia en la figura 1.1.

Un sistema de reconocimiento biométrico se divide en dos grandes campos: los métodos estáticos (off-line) y los métodos dinámicos (on-line). Los métodos estáticos verifican características de la firma que no varían con el tiempo y se refiere a imágenes escaneadas de documentos manuscritos, y los métodos dinámicos contienen información temporal de la firma (como la dirección, la presión, velocidad, altitud, entre otros). De esta manera, el método dinámico tiene más información respecto al método estático, por lo tanto, puede proporcionar resultados más precisos y confiables.

2.1.1 Firmas Dinámicas (On-line)

Tradicionalmente, en muchas de las aplicaciones comerciales se han utilizado sistemas basados en el conocimiento (por ejemplo, PINs y contraseñas), aplicaciones gubernamentales han utilizado sistemas basados en tokens (por ejemplo, tarjetas de identificación), y aplicaciones forenses han dependido de personas expertas para que coincida con las características biométricas. Los sistemas biométricos se están empleando cada vez más en aplicaciones civiles a gran escala (por ejemplo, acceso a ordenadores por reconocimiento de rostro, sistema de identificación de personas en cajeros automáticos,

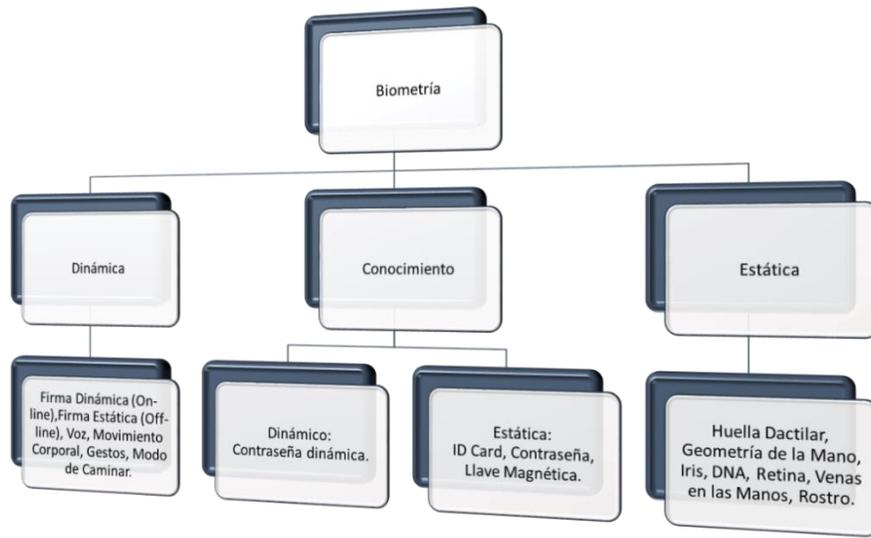


Figura 1.1. Análisis Biométrico.

Sistema de paso de fronteras mediante el reconocimiento del iris en aeropuertos, cerradura de la puerta a base de huellas dactilares utilizado para restringir el acceso a instalaciones, entre otros). Como se observa en la figura 1.2



Figura 1.2: Ejemplos de aplicaciones biométricas. (Tomada de Jain et al., 2004)

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

No obstante, la biometría basada en los rasgos conductuales abarca gran parte de la investigación que se realiza hoy en día para la identificación biométrica conductual (Cappelli et al., 2006; Przybocki et al., 2006; P. J. Phillips & Rizvi, 2006; Phillips, 2006; Alariki et al., 2014). Por consiguiente, en este trabajo se presenta una metodología con respecto a la identificación biométrica de personas por medio de la dinámica de la firma manuscrita.

La firma dinámica representa toda su información en el momento que el usuario inscribe toda su acción psicomotriz realizada por la mano sobre un dispositivo de adquisición de datos; en otras palabras, obtener las características anatómicas cuando un individuo exhibe cuando firma. Las especificaciones que debe tener un sistema de reconocimiento biométrico en el mercado deben tener en cuenta algunos aspectos (Pascual-Gaspar, 2009) como sigue:

1. Uso de un número reducido de muestra de referencia.
2. Reducir el tamaño de almacenamiento de las firmas.
3. La eficiencia computacional.
4. La seguridad, buscando sobre todo robustez ante ataques con imitaciones
5. Capacidad de adaptación a los distintos tipos de firmantes y a sus variaciones a la hora de firmar, a corto y largo plazo.
6. Evitar emplear imitaciones en la creación del modelo.
7. No basar el éxito en características muy específicas del hardware de adquisición para ganar en universalidad.

Dependiendo del dispositivo de adquisición para la captura de la firma, se pueden obtener diferentes tipos de series temporales que posteriormente se le puede hacer el análisis dinámico. La captura de la firma manuscrita del individuo se realiza por medio de Tablet digitalizadoras especiales, que adquieren la dinámica intrínseca del firmante y registran la evolución temporal de las señales generadas por el lápiz al momento de firmar; asimismo registrando las coordenadas geométricas. El número de variables que representa la

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

dinámica de la firma infieren que puede llegar hacer un sistema de identificación confiable para establecer la identidad de un individuo.

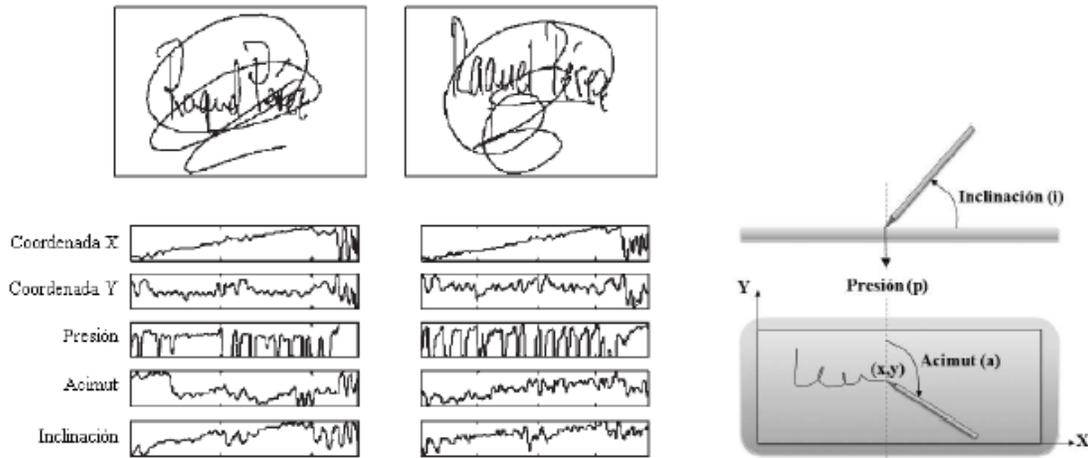


Figura 1.3: Series temporales extraídas de una firma dinámica. (Tomada de Garcia-Salicetti et al., 2003)

2.1.2 Tipos de sistemas biométricos

En el reconocimiento biométrico se encuentran dos tipos de sistemas que dependiendo del contexto de la aplicación pueden funcionar como un sistema de Identificación o como un sistema de verificación Biométrica.

- Un sistema de verificación podría definirse como la comparación de los datos biométricos de un individuo con los datos propios guardados con anticipación en una base de datos establecida. En este sistema, un individuo declara una identidad para ser reconocida y el sistema realiza una comparación (también conocida como uno a uno (1:1)) para determinar si el individuo es falso o verdadero.
- En un sistema de identificación, los datos biométricos se comparan con todos los conjuntos de características biométricas de los usuarios previamente guardadas en una base de datos. Este sistema también es conocido como uno-contra-todos o (1:N). el sistema realiza la comparación de los datos biométricos del individuo con los demás individuos que se encuentran en la base de datos y establece una búsqueda

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

de coincidencias para determinar la identidad del individuo, entregando como resultado la identificación o la no identificación de la persona.

Dentro de un sistema de biométrico, en la mayoría de las veces se necesita tener un sistema capaz de entregar resultados rápidos cuando se requiera verificar o identificar a un individuo, es por eso que es muy común utilizar sistemas de verificación, ya que utiliza menos procesamiento y comparaciones. Por lo tanto el proceso de verificación es muy reducido. Por esta razón, es habitual usar verificación cuando se quiere validar la identidad de un individuo desde un sistema con capacidad de procesamiento limitada o se quiere un proceso muy rápido (Faundez-Zanuy, 2007).

2.1.3 Evaluación de los sistemas biométricos

Tanto para la verificación y la identificación en un sistema biométrico se necesita una medida la cual determine el rendimiento del sistema. En el caso de la identificación biométrica es común utilizar como medida de desempeño el porcentaje de aciertos del sistema respecto al total de identificaciones realizadas, es decir, el porcentaje de firmas que fueron correctamente asignadas a cada individuo o clase (Pascual-Gaspar, 2011). Para realizar esta labor o tarea, hay que tener en cuenta dos posibilidades como: suponer que el individuo a identificar se encuentra registrado en una base de datos (identificación cerrada), o caso tal realizar una verificación encubierta posterior a la identificación (identificación abierta).

Para establecer una identificación cerrada de una persona, se verifica las características biométricas de la persona con las que se encuentren almacenadas en la base de datos. Tomando como premisa su previo almacenamiento de sus características. El sistema como resultado dará un porcentaje de acierto con la persona que más se asemeja las características introducidas. Y la identificación abierta es un proceso híbrido entre la verificación y la identificación cerrada, donde la persona no reclama una identidad específica, entonces se compara contra toda la base de datos para verificar si existe en la

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

base de datos, una vez se verifica que posiblemente existe, dentro de las coincidencias más probables, determina quién es el usuario (Phillips et al., 2003).

Teniendo el individuo con sus características biométricas y la base de datos con las características guardadas se define típicamente el tipo de error para evaluar el rendimiento del sistema biométrico en modo identificación. Se utiliza las curvas CMC (Cumulative Match Characteristic) y para el caso de la verificación se generan las curvas ROC (Receiver Operating Characteristic) (Bolle et al., 2005).

2.2 Análisis de dinámica no lineal

El análisis de Dinámica No Lineal (DNL) es una herramienta alternativa matemática para analizar señales de tiempo discreto, y las señales fisiológicas del cuerpo humano (Goldberger et al., 1990; Delgado et al., 2009).

El mundo está lleno de fenómeno que parecen caóticos aunque, en realidad, se ciñen a reglas estrictas pero difíciles de esclarecer por la gran cantidad de variables implicadas o de combinaciones generadas. La computación y la representación gráfica de la evolución de los valores de las variables es lo que ha permitido entrar en un nuevo campo de análisis llamado más comúnmente hoy "la frontera del caos" (Ahmedt-Aristizabal & et al., 2011).

La teoría del caos ha de ser entendida como una nueva herramienta de análisis que permite afrontar problemas hasta ahora inabordables o difícilmente analizables por la estadística (Subias, 1991). Las series de tiempo contienen información del proceso que generó dicha serie, por lo que es de suma importancia utilizar la dinámica contenida en el sistema para inferir sobre la dinámica desconocida de la señal. Con la señal se realiza la reconstrucción de las trayectorias que surgen de los sistemas analizados que tienden a concentrarse en una región finita o acotada del espacio (atractor). En diversas aplicaciones, el análisis no lineal trata de conservar las propiedades de la dinámica subyacente. Esto permite una herramienta para para el diagnóstico y la clasificación señales, demostrando que en muchas situaciones una correcta aplicación de la dinámica no lineal brinda nuevas perspectivas de rendimiento. Las técnicas de dinámica no lineal

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

son aquellas que permiten extraer información oculta de las señales y deducir el orden subyacente que ocultan fenómenos aparentemente aleatorios (Delgado et al., 2009). Los sistemas dinámicos no lineales son aquellos cuya respuesta ante un estímulo dado en algún instante de tiempo no se produce de forma proporcional a su respectiva excitación de entrada. Es por esto, que aunque exista un modelo determinístico del sistema no lineal, este al entrar en el régimen caótico genera un comportamiento impredecible a largo plazo si no se tienen de manera precisa las condiciones iniciales (Delgado et al., 2009). Los sistemas estables tienden a converger en un punto a lo largo del tiempo, conforme a su dimensión (atractor). En cambio, los sistemas inestables se salen de trayectoria o se escapan de los atractores. Entonces, para formar un sistema caótico, este exterioriza los dos comportamientos. En el sistema caótico por un lado, existe un atractor por el que el sistema se ve atraído, pero a la vez, existe una “potencia” que lo aleja de éste. De este modo, el sistema se queda confinado en una zona de su espacio de estados, pero el atractor será totalmente irregular o extraño. A diferencia del análisis lineal, el cual atribuye el comportamiento irregular de un sistema a la naturaleza aleatoria de la señal de entrada, la teoría del caos afirma que las entradas aleatorias no son las únicas fuentes posibles de irregularidad en la salida de un sistema (Castellanos, 2005). No se debe confundir un proceso caótico con un proceso aleatorio. Cuando el proceso es caótico, si se conoce con una precisión infinita la condición inicial del sistema se puede saber el estado en cualquier instante. Es decir, un proceso caótico es determinista aunque poco predecible. De un sistema del que se conocen sus ecuaciones características, y con unas condiciones iniciales fijas, se puede conocer exactamente su evolución en el tiempo. Pero en el caso de los sistemas caóticos, una mínima diferencia en esas condiciones hace que el sistema evolucione de manera totalmente distinta (Solé et al., 1996).

Entre la infinidad de medidas invariantes que pueden encontrarse, tanto estáticas como dinámicas, la caracterización utilizando dinámica no lineal típicamente suele utilizar al Máximo Exponente de Lyapunov, y la dimensión de Correlación. En el presente trabajo sin embargo, se estudiará el poder caracterizante de las anteriores, junto con el Exponente de

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Hurst, Complejidad de Lempel-Ziv, y una medida de la regularidad de series de tiempo, todas aproximaciones de la Entropía.

2.2.1 Dimensión de Correlación (D_2)

La dimensión de correlación es una medida de la complejidad de un sistema determinístico que proporciona el número de variables independientes necesarias para describir su comportamiento. Estima la dimensión fractal del atractor que se forma en el espacio de fase. La dimensión de correlación permite distinguir entre una serie de datos aleatoria y un ruido determinista. Esta medida se obtiene a partir de la suma de correlación $C(r)$, determinando la zona donde las curvas generadas para cada valor de m se comportan de forma lineal o secuencial. A partir de esta región, se calcula la pendiente de cada recta, el cual representa el valor de D_2 para cada valor de m (Grassberger & Procaccia, 1983). La dimensión de correlación se basa en la correlación integral definido por:

$$C(r) = \lim_{N_m \rightarrow \infty} \frac{1}{N_m^2} \sum_{i,j=1}^{N_m} H(r - \|x_i^m - x_j^m\|) \quad (3)$$

El límite de esta expresión debe ser interpretado de la siguiente forma: Debe considerarse el mínimo radio (r) de esfera tal que la cantidad de puntos dentro de ésta sea el apropiado, logrando capturar la información acerca de la dependencia entre estos dentro del atractor (Grassberger & Procaccia, 1983).

Donde N_m es el número de puntos del espacio de fases reconstruido y H es la función de Heaviside definida por:

$$H(x) = \begin{cases} 1 & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases} \quad (4)$$

La dimensión de Correlación se expresa de la forma:

$$D_2 = \lim_{r \rightarrow 0} \frac{\log(C(r))}{\log(r)} \quad (5)$$

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

El límite de esta expresión debe ser interpretado de la siguiente forma: Debe considerarse el mínimo radio de esfera tal que la cantidad de puntos dentro de ésta sea el apropiado, logrando capturar la información acerca de la dependencia entre estos dentro del atractor (Grassberger & Procaccia, 1983).

2.2.2 Entropía de Kolmogorov (K_2)

La entropía de Kolmogorov se utiliza para diferenciar los comportamientos aleatorios y periódicos o regulares que poseen o que describan caos. Ésta medida de complejidad se le da el nombre de entropía porque representa la velocidad media de pérdida de información en una serie temporal (montero & morán, 1994).

El cálculo de la entropía de Kolmogorov permite medir la impredecibilidad de un sistema dinámico. Es un parámetro que permite determinar el contenido y ganancia de información en un sistema determinado (Walters, 1982). Por lo tanto, la entropía de Kolmogorov se define como principio que mide la pérdida de información a lo largo de la evolución del sistema. Tiene una gran importancia, en su principal aplicación a sistemas de los cuales no se dispone más que de series temporales de valores de determinada variable, la cual se considera poseedora de significado.

En un sistema dinámico sumido en un espacio de fase con dimensión m , y que se encuentra dividido en pequeños bloques de tamaño r^m , siendo r la longitud del lado de cada una. Asumiendo entonces la existencia de un atractor dentro del espacio de fase, constituido por la evolución temporal de una trayectoria $S[n]$, siendo los puntos la representación de los estados dinámicos del sistema medidos en pequeños intervalos de tiempo τ .

Sea $p(k_1, k_2, \dots, k_m)$ la probabilidad conjunta de la trayectoria $S[\tau]$ se encuentre en la caja k_1 , $S[2\tau]$ en k_2 , hasta $S[m\tau]$ en k_m , entonces, la entropía de Kolmogorov se define como:

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

$$K = - \lim_{\substack{r \rightarrow 0 \\ m \rightarrow \infty}} \frac{1}{m\tau} \sum_{p(k_1, k_2, \dots, k_m)}^N p(k_1, k_2, \dots, k_m) \ln p(k_1, k_2, \dots, k_m) \quad (6)$$

Con la ecuación anterior, se define la estimación K_2 de la entropía de Kolmogorov de la siguiente manera:

$$K_{2,m}(r) = \frac{1}{\tau} \ln \frac{C_m(r)}{C_{m+1}(r)} \quad 0 \leq K_2 \leq K \quad (7)$$

Donde m es la dimensión de sumidero, τ es el retardo, y $C_m(r)$ es la suma de correlación definida anteriormente del cálculo de la dimensión correlación $K_2 \rightarrow \infty$ para sistemas aleatorios, mientras $0 \leq K_2 \leq \infty$ para sistemas caóticos.

2.2.3 Entropía de Shannon (H_x)

Shannon en 1948 definió la entropía desde el punto de vista de la teoría de la información como una medida de la incertidumbre media de la cantidad de información enviada en un mensaje. De esta manera la entropía es una medida que cuantifica la incertidumbre presente en un conjunto de datos debido a su cantidad de información.

Dada una variable aleatoria discreta X , con un rango de posibles valores, $X = (X_1, \dots, X_n)$ con un rango de valores $\theta_1, \dots, \theta_n$, función de probabilidad $p(x_i) = P_r\{X = x_i\}$, y el operador E , de la Entropía de Shannon indicado por H , la entropía de Shannon es indicada (Al- Mayyan et al, 2011).

$$H(X) = - \sum_{x_i \in \theta} p(x_i) \log p(x_i) = -E[\log p(x_i)] \quad (8)$$

2.2.3 Exponente de Hurst (H)

El exponente de Hurst es una medida de un sistema de memoria a largo plazo, es decir, cuantifica la influencia de los valores pasados en el valor presente o actual de la serie temporal. También puede ser vista como una medida de la persistencia del proceso a

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

seguir las tendencias, o la correlación entre los diferentes puntos de las series de tiempo (Sprott and Sprott, 2003). El exponente de Hurst tiene un rango entre 0-1, donde un valor inferior a 0,5 presenta una correlación negativa, que implica un proceso anti-persistente que tenderá a volver al punto donde comenzó. Así que, si la tendencia de la serie es positiva, puede haber un cambio de tendencia negativa en la siguiente iteración. Un valor cercano a 0,5 indica que no hay correlación entre los valores pasados y los valores presentes (Movimiento Browniano). Un valor mayor a 0,5 indica correlación positiva, y es un indicador de un proceso continuo, en el cual es probable que se mantenga constante la dirección de la trayectoria. El cálculo del exponente de Hurst del espacio de estados $S \rightarrow [n]$ se obtiene de la regresión empírica como la pendiente de la relación,

$$\frac{R}{\sigma} = \left(\frac{\tau}{2}\right) \quad (9)$$

Donde R es la variación span (diferencia entre valor máximo y valor mínimo en las series), σ es la derivación estándar, y τ es el retardo utilizado en la reconstrucción del atractor.

2.2.4 Complejidad de Lempel–Ziv (LZ)

Lempel y Ziv proponen un algoritmo de una medida de complejidad útil, la cual puede caracterizar el grado de orden y desorden y el desarrollo de patrones espacio-temporales (Lempel y Ziv, 1976). El cálculo de la complejidad de Lempel-Ziv estima la complejidad o la irregularidad de una serie de tiempo y generalmente se calcula utilizando el algoritmo descrito en (Lempel y Ziv, 1976). Los valores van cerca de 0 a 1. En particular, $LZ = 1$ significa máxima complejidad (patrón totalmente aleatoria ruido blanco), y $LZ = 0$ significa series perfectamente predecible (ecuación determinista).

2.3 Máquinas de Vectores de Soporte (SVM).

La teoría de las Maquinas de Soporte Vectorial (SVM por su nombre en inglés), es una técnica de clasificación de aprendizaje supervisado, en la literatura goza de una gran popularidad, tal vez porque se cuenta con un fuerte respaldo teórico desarrollado por Vapnik (Vapnik, 1998), basado en la idea de minimización del riesgo estructural.

En muchas aplicaciones, las SVM han mostrado tener gran desempeño, más que las máquinas de aprendizaje tradicional como las redes neuronales y han sido introducidas como herramientas con un gran potencial para resolver problemas de clasificación.

La ventaja de las SVM se debe a que primero mapea los puntos a un espacio de una dimensión mayor y encuentra un hiperplano donde separe y maximice el margen entre las clases en el nuevo espacio, como se puede ver en la figura 1.4.

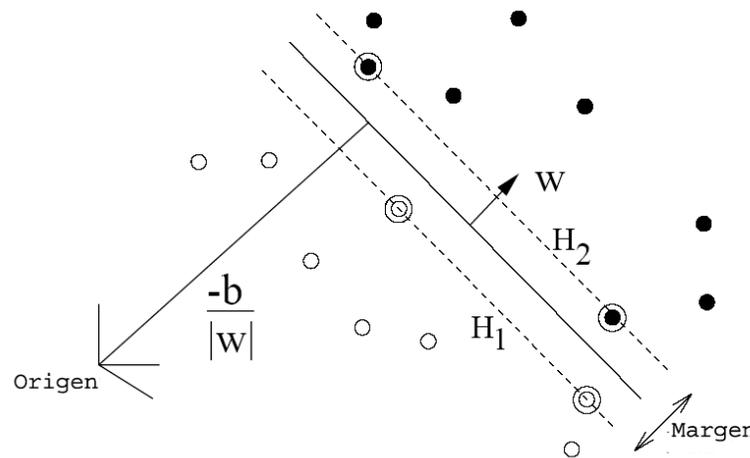


Figura 1.4. Ajuste para una SVM lineal.

Maximizar el margen es un problema de programación cuadrática (QP) y puede ser resuelto por su problema dual introduciendo multiplicadores de Lagrange. La SVM encuentra el hiperplano óptimo utilizando el producto punto con funciones en el espacio de características que son llamadas kernels.

2.3.1 Caso Linealmente Separable

Supongamos que se tiene un conjunto S de puntos etiquetados (puntos blancos y negros en la figura 1.4)

$$S = (y_1, x_1), \dots, (y_N, x_N) \quad (10)$$

Cada punto de entrenamiento $x_1 \in \mathbb{R}^n$ pertenece a alguna de dos clases y se le ha dado una etiqueta $y_1 \in \{1, -1\}$. El objetivo es buscar o encontrar un hiperplano de separación donde aquellos puntos con igual etiqueta queden al mismo lado del hiperplano.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Esto significa encontrar un w y un b tal que:

$$y_i(w'x_i + b) > 0, i = 1, \dots, N \quad (11)$$

Si existe un hiperplano que separe a la ecuación anterior, se dice que los datos son *linealmente separables* y el hiperplano estará descrito por un vector normal w , con una distancia al origen dada por $b/\|w\|$.

La ecuación 10, puede reescribirse como:

$$y_i(w'x_i + b) \geq 1, i = 1, \dots, N \quad (12)$$

Con esto se busca que las distancias H_1 y H_2 representados en la figura 1.4, sea igual a $2/\|w\|$. Por consiguiente, para encontrar el máximo margen es necesario minimizar la norma de w , es decir:

$$\text{Minimizar} \quad \frac{1}{2}w'.w \quad (13)$$

$$\text{Sujeto a} \quad y_i(w'x_i + b) \geq 1 \quad \forall i \quad (14)$$

Mientras que $w'.w$ sea convexo, el problema 13 tendrá una posible solución utilizando los multiplicadores de Lagrange. Sean $\alpha = \{\alpha_1, \dots, \alpha_N\}$, los N multiplicadores de Lagrange no negativos.

Para minimizar 13 hay que encontrar el punto de silla de la función de Lagrange:

$$L(w, b, \alpha) = \frac{1}{2}w'w - \sum_{i=1}^N \alpha_i [y_i(w'x_i + b) - 1] \quad (15)$$

Para encontrar el punto de silla, hay que minimizar la función 14 respectos a w y b , después maximizarla sobre los multiplicadores de Lagrange $\alpha_i \geq 0$. El punto encontrado debe satisfacer las condiciones de Karush-Kuhn-Tucker (KKT) donde demuestra que la mayoría de los coeficientes de Lagrange son cero, y que solo puede ser distintos de cero

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

los vectores de soporte, los vectores que se encuentran exactamente en la distancia marcada por el margen. Al dualizar el modelo de maximización del margen, la función se transforma en una función cuadrática convexa sujeta a restricciones lineales.

Cuando se ha encontrado la solución de la función, el hiperplano de separación tiene la siguiente forma:

$$w_0 = \sum_{i=1}^N \alpha_i^0 y_i x_i \quad (16)$$

Mientras que b es encontrado a partir de las condiciones de (KKT):

$$b_0 = y_i - w_0' x_i \quad (17)$$

El problema de clasificar un nuevo punto x , es resuelto examinando el signo de:

$$w_0' x_i + b_0 \quad (18)$$

2.3.2 Caso Linealmente no Separable

En muchas aplicaciones se pueden encontrar que los datos son linealmente no separables, el análisis previo puede ser generalizado introduciendo variables no negativas o “slack” (de relajación) (ξ_1, \dots, ξ_N) , $\xi_i \geq 0$, de tal modo que la expresión 12 se puede escribir como:

$$y_i(w' x_i + b) \geq 1 + \xi_i, \forall i \quad (19)$$

$\xi_i \geq 0$ en 19 son aquellos puntos que los x_i no satisfacen a 12. Entonces el término $\sum_i \xi_i$ puede tomarse como algún tipo de medida de error en la clasificación.

El problema del hiperplano óptimo es entonces redefinido como la solución del siguiente problema:

$$\text{Minimizar} \quad \frac{1}{2} w' w + C \sum_{i=1}^N \xi_i \quad (20)$$

$$\text{Sujeto a} \quad y_i(w' x_i + b) \geq 1 + \xi_i \text{ y } \xi_i \geq 0 \forall i \quad (21)$$

La primera función es minimizada para poder controlar los errores cometidos en la etapa de aprendizaje del clasificador.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

El segundo término permite tener el control del número de clasificaciones erróneas. El valor del parámetro C lo elige el usuario de tal manera que un valor muy grande equivale a asignar una alta penalización a los errores, implicando a la creación del llamado margen blando (soft margin); que permite algunos errores en la clasificación a la vez que los penaliza. Haciendo una comparación con el caso linealmente separable, la utilización de los multiplicadores de Lagrange deriva en el siguiente problema de optimización:

$$\text{Minimizar} \quad \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i' x_j \quad (22)$$

$$\text{Sujeto a} \quad \sum_{i=1}^N \alpha_i y_i = 0 \text{ y } 0 \leq \alpha_i \leq C, \forall i \quad (23)$$

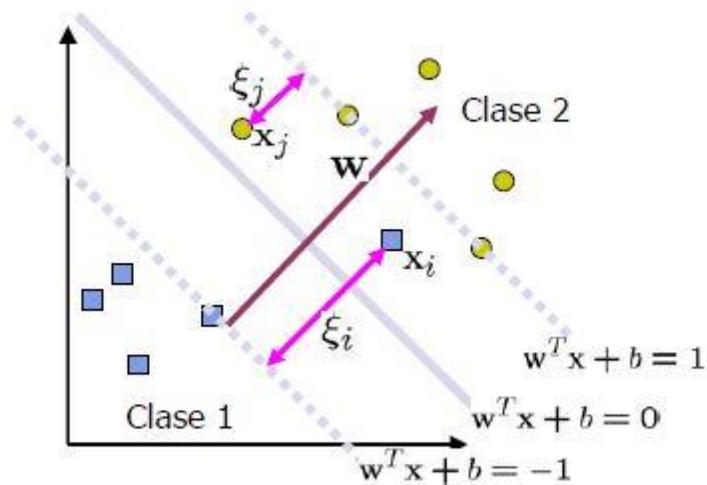


Figura. 1.5. Aparición del parámetro de error ξ_i en el error de clasificación. (Tomada de Betancourt A., 2005).

2.3.3 Espacio de alta dimensionalidad – Kernels.

Cuando se habla de espacio de alta dimensionalidad, se habla de las Máquinas de Vectores de Soporte no lineales; cuyo principio consiste en mapear el espacio de los datos de entrada a un espacio de representación mayor o de dimensión alta a través de una función no lineal. Existe una buena propiedad que posee SVM. Esta propiedad es una función llamada *Kernel* (Schölkopf, 2002); calcula el producto punto de los puntos de

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

entrada en el espacio de representación Z , el mapeo $Z = \phi(x)$, de manera que reemplazando x por $\phi(x)$ en la siguiente ecuación:

$$\text{Maximizar} \quad \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i' x_j \quad (24)$$

$$\text{Sujeto a} \quad \sum_{i=1}^N \alpha_i y_i = 0 \text{ y } \alpha_i \geq 0, \forall i \quad (25)$$

Se obtiene:

$$\text{Maximizar} \quad \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \phi(x_i)' \phi(x_j) \quad (26)$$

$$\text{Sujeto a} \quad \sum_{i=1}^N \alpha_i y_i = 0 \text{ y } \alpha_i \geq 0, \forall i \quad (27)$$

De lo anterior, el entrenamiento del algoritmo solo depende de la función *Kernel*, es decir, funciones de la forma $\phi(x_i)' \phi(x_j)$. Dada la función Kernel simétrica K tal manera que $K(x_i, x_j) = \phi(x_i)' \phi(x_j)$, de este modo el entrenamiento solo depende de K .

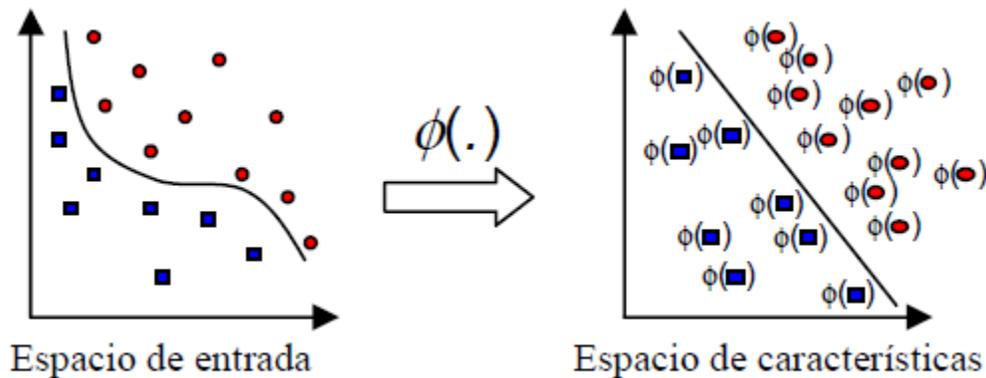


Figura. 1.6. Uso de un tipo de Kernel para la transformación del espacio de representación de los datos. (Tomada de Betancourt A., 2005).

Las Funciones que satisfacen el teorema de Mercer pueden ser usadas como productos punto y por ende pueden ser usadas como kernels.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Sin ningún conocimiento del mapeo, la SVM encuentra el hiperplano óptimo utilizando el producto escalar con funciones en el espacio de características que se llaman kernels. La solución de la hiperplano óptimo se puede escribir como una combinación de unos pocos puntos de entrada que son llamados vectores de soporte. Para extender la solución a problemas de frontera no lineales, funciones que expresan el producto escalar de dos vectores en el espacio de entrada (denominado kernels) se pueden introducir. En este trabajo, el ampliamente conocido Función de base radial (RBF) kernel es utilizado:

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} \quad (28)$$

Donde dispersión $\sigma \in \mathbb{R}^+ > 0$ controla el ancho de la función RBF. Tanto la dispersión del kernel y los parámetros de compensación del clasificador, C , son sintonizados para la prueba con una optimización meta-heurística de enjambre de partículas que es el método bio-inspirado utilizado para la determinación de parámetros de SVM (Lin et al., 2008).

2.3.4 Optimización por Enjambres de Partículas.

Optimización de enjambre de partículas (PSO) es un algoritmo de optimización global de aleatorio. Este pensamiento básico proviene de un comportamiento inteligente del enjambre. Emula principalmente las características de la migración y la recolección de las aves durante la búsqueda de alimento (Eberhart y Shi, 2001). Las soluciones potenciales en PSO se denominan partículas. Estas partículas vuelan hiperespacio a fondo del problema siguiendo mejores partículas. Todas las partículas tienen valores asignados por la función de evaluación, y sus velocidades dirigen las partículas voladoras. PSO se inicializa con un grupo de partículas aleatorias y buscan el punto óptimo a través de un proceso de actualización sucesivas denominadas generaciones (Lin y Hsieh, 2009). En cada iteración, las partículas son actualizadas por los siguientes mejores valores. La primera de ellas es la mejor solución que ha encontrado una partícula hasta el momento (pb) y la segunda es la mejor solución global obtenida hasta ahora en todas las partículas de la

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

población (gb) (Eberhart y Shi, 2001). La velocidad y la posición de las partículas (28) se actualizan con las ecuaciones:

$$v_i^{t+1} = v_i^t + c_1 r_1 (pb_i^t - pb_i^t) + c_2 r_2 (gb^t - p_i^t) p_i^{t+i} = p_i^t + v_i^{t+1} \quad (29)$$

Donde p_i , v_i son la enésima posición y velocidad de las partículas (solución), respectivamente. pb_i^t y gb^t Ya están definidas, r_1 y r_2 son números aleatorios entre (0,1) y c_1 y c_2 son el factor de aprendizaje (Eberhart y Shi, 2001).

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3. MARCO EXPERIMENTAL

En este capítulo se describe la metodología propuesta para la Identificación de firmas dinámicas por métodos no lineales usando máquinas de vectores de soporte.

El proceso a emplear para encontrar la metodología de identificación y verificación de la firma dinámica (on-line), parte desde la captura de la información utilizando dispositivos digitales especializados, con capacidad de capturar la evolución temporal de las señales generadas por el lápiz en el plano establecido para firmar, al momento de interactuar con dicho dispositivo. En ese momento es cuando se capturan las características comportamentales fundamentales de estudio por cada firmante, dicha información es procesada y almacenada en la bases de datos para el respectivo análisis de identificación biométrica. La identificación es un procedimiento que se realiza combinando procesos uno a muchos (1:N). Es decir, una muestra perteneciente a una clase es comparada con las otras muestras que se encuentran almacenadas en una base de datos de firmas dinámicas, registros que se tiene de guardar toda la información de cada uno de los firmantes o de todos los usuarios para su respectiva identificación. De esta manera, el objetivo de la identificación biométrica es el de clasificar una realización determinada de un rasgo biométrico de identidad desconocida como perteneciente a uno de entre un conjunto de N posibles individuos (Ahmedt-Aristizabal & et al., 2011).

En este capítulo presentamos la base de datos de firma dinámica sobre la que se centra este trabajo de grado. En la segunda parte del capítulo mostramos los resultados obtenidos en ella después de caracterizar y clasificar las firmas encontradas en la base de datos.

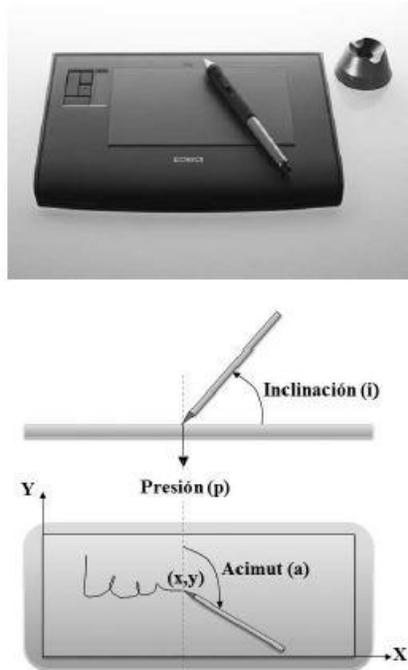


Figura.3.1. Tablet gráfica utilizada para la adquisición de las firmas. (Tomada de Gaspar Pascual, 2010)

En la figura 3.2, se presenta con un poco más en detalle la metodología propuesta en este trabajo de grado.

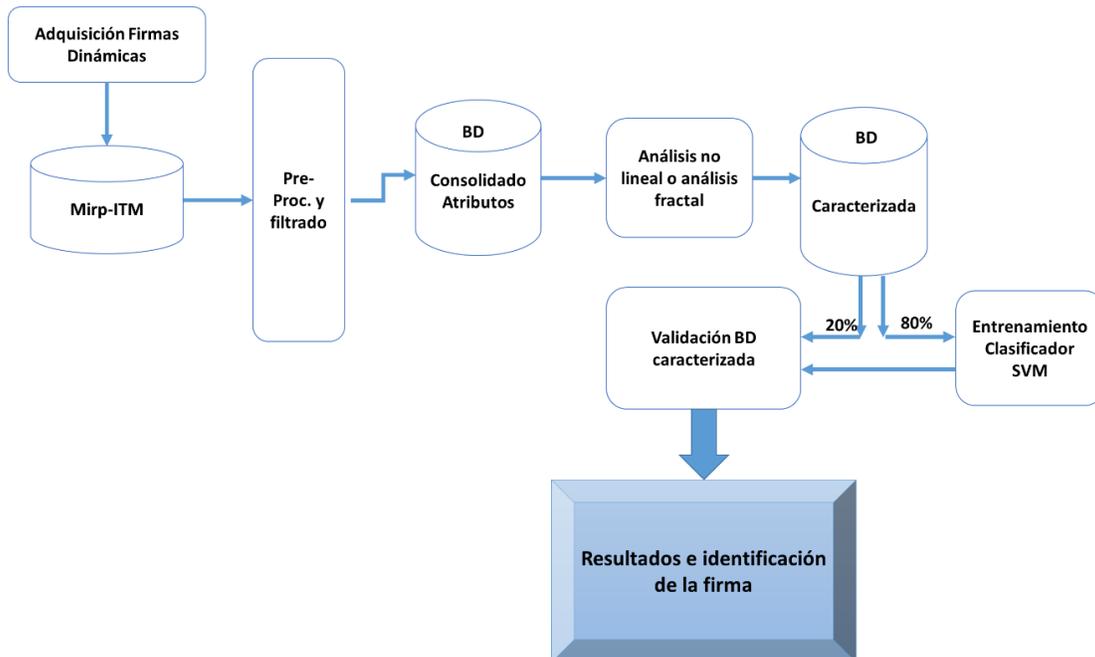


Figura. 3.2. Diagrama de bloques de la Metodología utilizada en este trabajo.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3.1. Base de Datos

La disponibilidad pública de bases de datos biométricas de referencia permite a los investigadores evaluar sus propuestas con los mismos conjuntos de datos, aunque la falta de protocolos estándar de evaluación puede provocar algunas diferencias significativas entre sus resultados. En consecuencia, las competiciones o evaluaciones objetivas realizadas por terceros establecen reglas y protocolos de evaluación que los autores adoptan para medir el rendimiento de sus sistemas. La combinación de ambos recursos permite con el tiempo avanzar el estado del arte a nivel científico y tecnológico. En la construcción de sistemas automáticos de identificación biométrica, es beneficioso el empleo de estas bases de datos, por cuanto estas contribuyen a la evaluación de los métodos de análisis y procesamiento desarrollados de una forma reproducible, cuantitativa y estándar. A continuación se presenta la base de datos de firmas dinámicas, con la que se evalúa el rendimiento o desempeño de la metodología planteada en la figura 3.2.

3.1.1. Base de Datos (ITM-MIRP-SIGN-01 – ITMMS01)

Esta base de firmas (sólo firmas auténticas), llamada ITM-MIRP-SIGN-01 (ITMMS01) el nombre de la universidad y el grupo de investigación (Ahmedt-Aristizábal et al., 2011). Contiene 800 firmas recogidas a partir de 40 voluntarios diferentes. Entre estos sujetos, había 23 mujeres, dos zurdas, con edades que oscilan entre 20 y 50. Se le pidió a cada sujeto a contribuir con 20 firmas recogidas en dos sesiones. Hubo aproximadamente dos semanas de tiempo entre las sesiones. Diez firmas fueron recolectadas de cada sujeto durante cada sesión.

No hubo restricciones, por lo que los sujetos firmaron en su forma más natural; en una orientación arbitraria. Por lo tanto, hubo una significativa deformación intra-clase y la variación entre las firmas pertenecientes al mismo sujeto. Las firmas se recolectaron con una tableta digitalizadora (WACOM Intuos 4), sin retroalimentación visual. Nosotros

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

diseñamos un mecanismo que fue utilizado por los autores para interconectar los sensores biométricos utilizando la Plataforma .NET y Component Object Model (COM) (plataforma de Microsoft para componentes de software introducida por la empresa en 1993) como una interfaz para el dispositivo. Cada firma de la base de datos se caracterizó por coordenadas X e Y de presión, el acimut y la información de la altitud.

En este trabajo, como se habló anteriormente se ha utilizado la tableta digitalizadora (marca WACOM) de modo que posee un conjunto de características generales que facilita la forma de visualizar los atributos extraídos de la firma. Este tipo de tableta gráfica proporciona las cinco características siguientes (Figura 3.1):

- Posición en el eje X: entre 0 y 12700 (0 - 127 mm).
- Posición en el eje Y: entre 0 y 9700 (0 - 97 mm).
- Presión: entre 0 y 1023
- Acimut: entre 0 y 3600 (0 – 360°)
- Inclinación: entre 300 y 900 (30° – 90°)

En la tabla 3.1 se resume algunas de las características básicas de la composición de la base de datos utilizada. La cifra mostrada en la tabla no se muestra en su totalidad, razón que llevó en nuestros experimentos a eliminar algunos usuarios por falta de datos.

Base de Datos	Origen	N° Autenticas	N° usuarios	Total Firmas
ITMMS01	Colombia	20	40	800

Tabla 3.1. Resumen de la composición de la base de datos utilizada.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

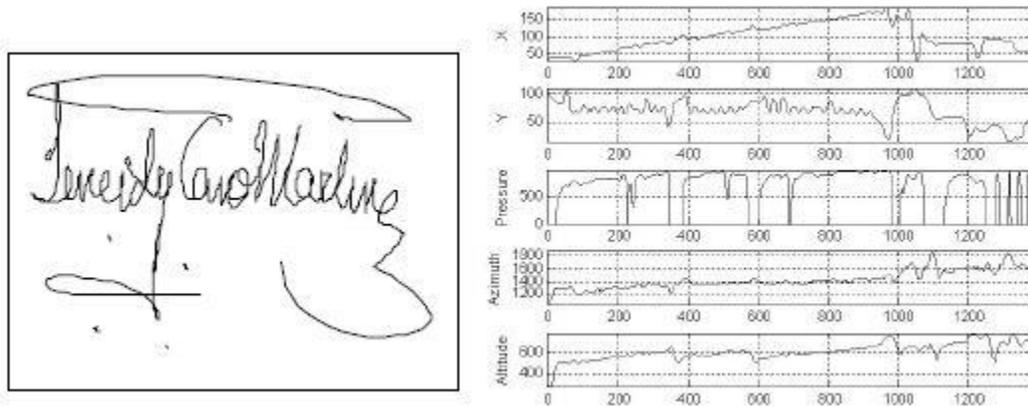


Figura. 3.3. Series de tiempo extraídas de la firma dinámicas. (Tomada de Ahmedt-Aristizabal et al., 2011)

Teniendo en cuenta que el firmante generaba la firma sin realimentación visual. Fue necesario crear una plataforma o una herramienta de desarrollo de apoyo a la investigación de este proyecto en lenguaje de programación Visual Basic .NET., La interfaz se generó con el fin de visualizar las firmas dinámicas como se muestra en la Figura 3.4, permitiendo analizar la firma dinámica con sus respectivos atributos extraídos con la que se evaluará el desempeño del sistema. Las señales o atributo extraídos son la posición en X, la posición en Y, acimut, altitud (altura del lápiz respecto a la superficie de escritura de la Tablet), y los cambios de presión, como se observó en la Figura 3.3. Es importante aclarar que en la adquisición no se presenta ruido, aunque en el cuadro visualizador de la firma para el usuario aparente tenerlo.

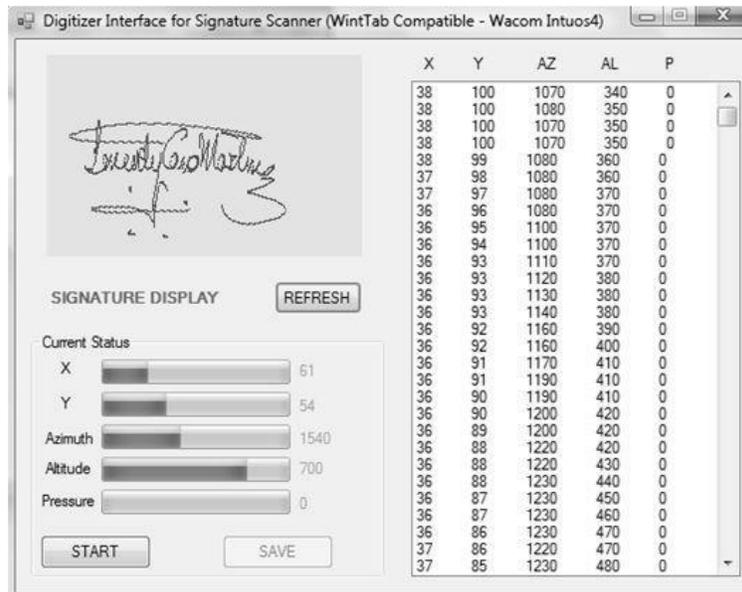


Figura 3.4. Ventana de la interfaz desarrollada para la visualización de las firmas dinámicas y de las series temporales.

3.2. Estimación de Características

Para la estimación de las características no lineales a cada de uno de los atributos extraídos a la firma dinámica, se aplican cada medida de complejidad descritas en la sección 2.2 de manera teórica y matemática. Entonces, para la conformación de la base de datos caracterizada; se tiene la base de datos propia conformada por 40 firmantes o clases y cada una representada por 20 muestras para una totalidad de 800 registros por 25 características extraídas.

Las medidas de complejidad utilizadas en este trabajo o proyecto que se aplicaron a cada uno de los atributos extraídos de las firmas dinámicas son las siguientes:

- Complejidad de Lempel-Ziv.
- Dimensión de Correlación.
- Exponente de Hurst.
- Entropía de Shannon.
- Entropía de Kolmogorov.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Las técnicas de dinámica no lineal son aplicadas a cada una de las series temporales que representan la dinámica intrínseca del proceso de la firma. La información temporal da a conocer cómo fue generada la firma y establece un marco de referencia que permite medir el grado de dependencia entre variables, o semejanza del comportamiento entre individuos.

La implementación de los algoritmos utilizados que calculan las medidas de complejidad de cada una de las muestras y sus respectivas series temporales están basadas en los Toolbox que pueden ser ejecutados en MATLAB. Las medidas de complejidad que se realiza a cada uno de los atributos extraídos de la firma dinámica. Los algoritmos utilizados para calcular las medidas de complejidad, encontramos a Lempel-Ziv; que se encuentra descrito en el artículo (A. Lempel and J. Ziv, 1976), el Exponente de Hurst que lo podemos encontrar (T. Di Matteo et al. 2003, 2005), (T. Di Matteo, 2007), (R. Morales et al., 2012), la entropía de Kolmogorov descrita en (Stephen Faul, 2005), la entropía de Shannon descrita (Shannon Claude E., 1948) y la dimensión de correlación (Grassberger and Proccacia, 1984).

3.3. Selección de los parámetros para SVM

Para que el modelo planteado funcione correctamente, se debe tener en cuenta los siguientes parámetros: C (Valor que establece la medida del modelo entre los errores de entrenamiento y los márgenes definidos por ξ) y el kernel $K(x, y)$ mapea el espacio de entradas X a un nuevo espacio de características de mayor dimensionalidad (Hilbert), para poder encontrar la línea o el plano de separación entre las clases.

Para calcular el parámetro C . Este se puede estimar por la intervención de una persona conocedora del tema, o por un algoritmo de búsqueda bio-inspirados en la naturaleza.

El parámetro C compensa la clasificación errónea de ejemplos de entrenamiento contra la simplicidad de la superficie decisión. Un bajo C hace que el clasificador sea más estricto, en la medida que se clasifica las muestras; mientras que un alto C tiene como objetivo clasificar todos los ejemplos de entrenamiento correctamente.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

En el caso del Kernel, para este proyecto de grado se escogió el kernel RBF (Radial Basis Function, por sus siglas en inglés) gaussiano descrito en la metodología de (Wang et al., 2003), por mostrar excelentes resultados en comparación con el Kernel gaussiano clásico (Scholkopf et al., 1997). El parámetro a calcular en el Kernel RBF es (γ). El algoritmo encargado de realizar o calcular dicho parámetro es la Optimización por enjambre de partículas (PSO, por sus siglas en inglés) descrita en la sección (2.4). Teniendo en cuenta la metodología utilizada en (Wang et al., 2003), la SVM varía la capacidad de su espacio de generalización con el Kernel RBF gaussiano, formando un espacio de búsqueda semiconvexo, dependiente de γ .

3.3.1. Selección y Sintonización del Kernel Gaussiano (γ).

El Kernel gaussiano puede ser usado en muchas aplicaciones que proporcionan un puente entre la linealidad y no-linealidad para algoritmos que pueden ser expresados en términos de un producto punto.

El Kernel gaussiano utilizado en este proyecto se define o está dado por:

$$K(x, y) = \exp(-\gamma \|x - y\|^2) \quad (30)$$

Donde γ se expresa como:

$$\gamma = \frac{1}{2\sigma^2} \quad (31)$$

El $\sigma > 0$ determina el ancho del Kernel Gaussiano y es el parámetro que debe ser sintonizado para el buen funcionamiento del clasificador.

El parámetro *sigma* ajustable juega un papel importante en el rendimiento del kernel, y debe ser cuidadosamente ajustado para el problema en cuestión. Si es sobrestimado, la exponencial se comportará de forma casi lineal y la proyección de más dimensiones comenzará a perder su poder no lineal. En el otro lado, si es subestimado, la función carecen de regularización y la frontera de decisión será muy sensible a ruido en los datos de entrenamiento.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Para evitar los problemas presentados anteriormente, es necesario utilizar un algoritmo de búsqueda que minimice el error de generalización de la SVM. Por lo tanto, se propone para este trabajo el algoritmo de búsqueda inspirado en el comportamiento social de individuos dentro de enjambres en la naturaleza. Básicamente usa la experiencia y el conocimiento de sus vecinos para desplazarse en un espacio de búsqueda con distintos grados de confianza. El algoritmo planteado es PSO y está descrito en la sección (2.4).

3.3.2. Factor de penalización de error.

El coeficiente C determina el compromiso entre la complejidad del modelo y el grado en que las desviaciones mayores a ξ son toleradas (J. Jaramillo, 2007). Cuando se generaliza la separación, los vectores de soporte presentan una expansión respecto a los errores cometidos por el clasificador. Entonces, se puede asociar el parámetro C al rango de salida de los datos de entrada (Mattera et al., 1999), (Mangasarian, 1999).

3.4. Técnica de Evaluación

La aplicación de técnicas de evaluación a los algoritmos de aprendizaje, buscan la forma de evaluar un análisis matemático o estadístico aplicado a un problema de reconocimiento de patrones. Se busca principalmente, la certeza sobre alguna inferencia hecha a un proceso o a un problema de la vida cotidiana.

En la actualidad, existen distintas formas de representar el modelo. Representación puede ser, representación proposicional, árboles de decisión, reglas de decisión, listas de decisión, reglas con excepciones, reglas jerárquicas de decisión, reglas difusas y probabilidades, redes neuronales, están entre las estructuras más utilizadas (Bishop, 2006).

El comparador también llamado módulo de clasificación, mide la semejanza entre el patrón de entrada y la información previamente almacenada en la base de datos. Para posteriormente entregar una decisión sobre si el patrón encontrado pertenece a un patrón existente en la base de datos.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Teniendo ya caracterizadas las bases de datos, y que éstas nos permitan establecer la capacidad discriminadora, es necesario validar la metodología propuesta para establecer su funcionamiento en un entorno realista. Se dispone a partir o dividir la base de datos en conjuntos utilizando una estrategia de validación para sistema de aprendizaje supervisado.

3.4.1 Validación Cruzada con K – Particiones.

Es una técnica de validación de un sistema de aprendizaje y se utiliza para evaluar a las metodologías de clasificación.

Para esta técnica, la base de datos se divide en K subconjuntos (folds) de datos de igual tamaño. En cada evaluación, se deja un subconjunto como datos de prueba y se entrena el clasificador con los $(K-1)$ folds restantes. El proceso de validación depende del número de particiones o de $(K - \text{folds})$ escogidos para la evaluación del sistema, por lo tanto, la validación se repite K iteraciones. El error (E) se calcula como la media aritmética de los errores de cada iteración (E_i) para obtener un único resultado, es decir:

$$E = \frac{1}{K} \sum_{i=1}^K E_i \quad (32)$$

Se realiza la sumatoria de los errores cometidos en cada iteración, dividido el número o valor de K .

3.5 Metodología de Clasificación.

Como se menciona en la introducción, en el marco teórico y los objetivos de este trabajo, el objetivo de este proyecto de grado está direccionada a la caracterización; utilizando el análisis fractal descrito en el marco teórico (Sección 2.2). Por esta razón, para obtener una mayor generalización de los datos se ha optado por la metodología de clasificación SVM. En esta etapa se genera un análisis del desempeño del sistema.

Como se puede visualizar a continuación (Figura 3.5), se presenta (en bloques) un poco más detallada, es decir, más específica la metodología de clasificación utilizada en este proyecto.

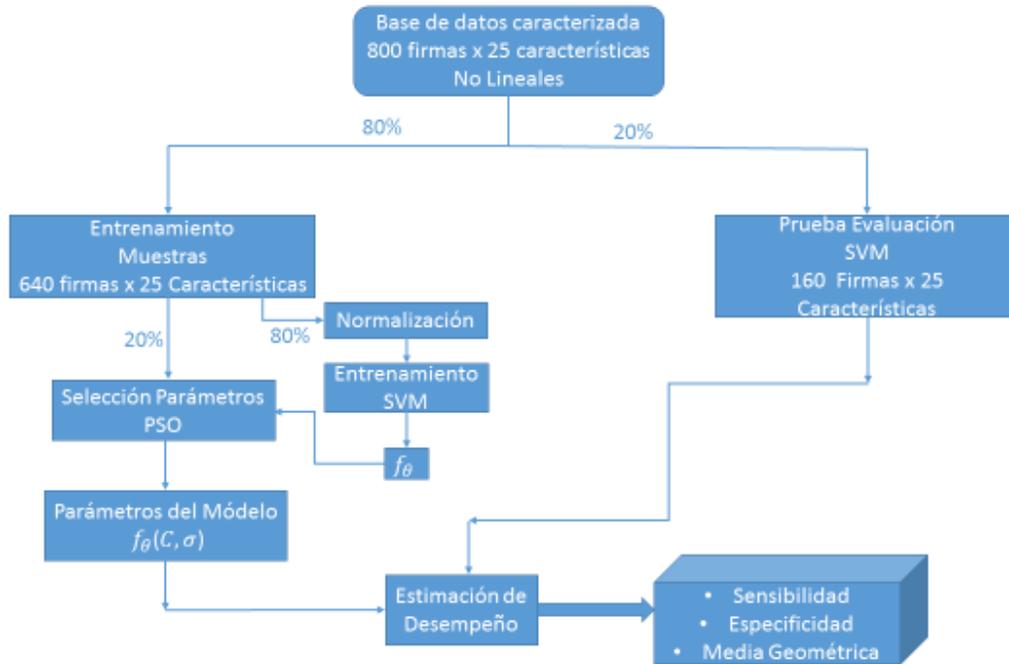


Figura. 3.5. Metodología de Clasificación.

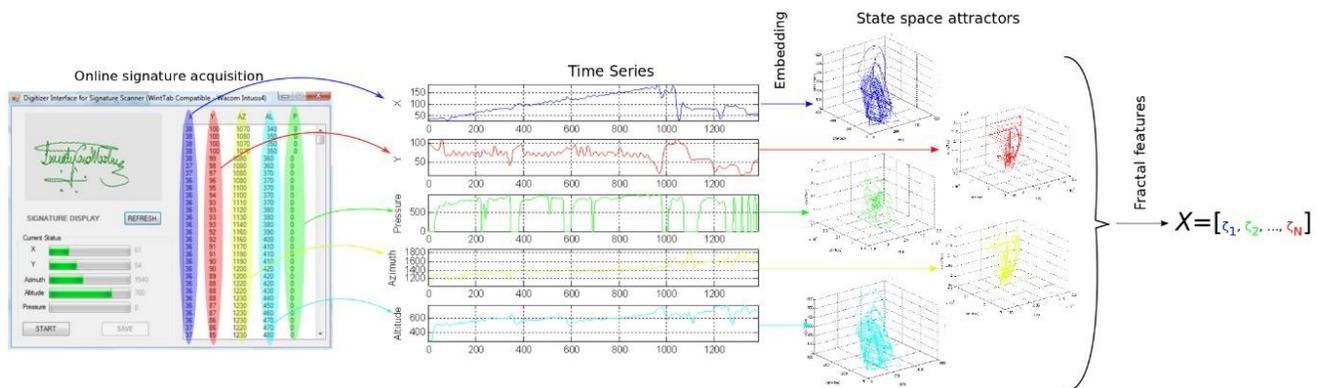


Figura. 3.6. Esquema general con el enfoque propuesto para la clasificación de las firmas Dinámicas.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Como se puede observar en la figura 3.5, se tiene en más detalle la metodología completa diseñada para la identificación de firmas dinámicas utilizando características no lineales o análisis fractal, propuesto en este trabajo de grado.

Una vez se adquiere los atributos de cada firmante, se obtienen 25 características por cada uno de los registros existentes (aplicando las características no lineales), almacenando toda la información obtenida en una nueva base de datos. A partir de ésta última base de datos se tomará el 80% de los registros para el entrenamiento del clasificador (SVM). Luego se procederá a realizar las respectivas validaciones con en el 20% de los registros que se reservaron para la validación del clasificador; la validación cruzada.

Una vez que se tiene el clasificador entrenado se genera la estrategia de validación cruzada descrita en la sección 3.4.1. Se utiliza el número de *folds* o número de particiones igual a 10.

Finalmente se obtiene las medidas estadísticas de desempeño como son: la sensibilidad, especificidad y la media geométrica. Cabe anotar que la identificación determina si la firma que se esté evaluando se encuentre en la base de datos propuesta.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3 EXPERIMENTOS Y RESULTADOS

En esta sección se presentan los resultados del trabajo de grado a partir de los métodos que se implementaron para la extracción de las características no lineales o el análisis Fractal y la validación de la metodología de clasificación con su respectivo análisis de desempeño.

La sensibilidad y la especificidad de desempeño del clasificador SVM son estimadas, y se evalúan mediante el uso de un procedimiento de validación cruzada convencional, que consiste en dividir la base de datos en 10 *folds*; cada uno con un número igual de señales por clase.

<i>Desempeño del Clasificador SVM</i>		
Media Geométrica	Sensibilidad	Especificidad
93.11	91.50	94.75

Tabla.4.1. Resultados de clasificación de firmas, en términos de medidas de rendimiento del clasificador considerado (a_{gm} , a_{sen} , a_{esp}).

Los datos dados anteriormente, son extraídos de la matriz de confusión (Kohavi and Provost, 1998) que se construyó a partir de la validación realizada a la SVM.

La matriz de confusión contiene información acerca de las clasificaciones actuales y predichas, y que son realizadas por un sistema de clasificación. El rendimiento de estos sistemas, se evalúa comúnmente usando los datos en la matriz. La siguiente tabla muestra la matriz de confusión para un clasificador de dos clases.

		<i>Clases Adivinadas</i>	
		Negativo	Positivo
<i>Clases Actuales o Reales</i>		<i>a</i>	<i>b</i>
	Negativo	<i>a</i>	<i>b</i>
	Positivo	<i>c</i>	<i>d</i>

Tabla.4.2. Matriz de Confusión. (Kohavi and Provost, 1998)

Letra **a** = Es el número correcto de predicciones que un caso sea negativo.

Letra **b** = Es el número incorrecto de predicciones que un caso sea positivo.

Letra **c** = Es el número incorrecto de predicciones que un caso sea negativo, y

Letra **d** = Es el número correcto de predicciones que un caso sea positivo.

Para la Sensibilidad, se calcula con los datos extraídos de la matriz de confusión. La fórmula con la cual se calcula la sensibilidad, es la siguiente:

$$\text{Sensibilidad} = \frac{d}{c+d} \quad (33)$$

Para la Especificidad, también se calcula con los datos extraídos de la matriz de confusión.

La fórmula con la cual se calcula la Especificidad, es la siguiente:

$$\text{Especificidad} = \frac{c}{c+d} \quad (34)$$

La media geométrica se calcula con la raíz *n*-ésima del producto de los *n* números (para este caso, los datos para calcular la media geométrica son la Sensibilidad y la Especificidad). Por tanto, la fórmula para la media geométrica es dada por:

$$MG = \sqrt[n]{(X_1)(X_2), \dots, (X_n)} \quad (35)$$

Los resultados de identificación de grupo cerrado se presentan en el porcentaje de firmas que fueron asignadas correctamente a cada usuario. Las medidas de clasificación promedio calculados de rendimiento se presentan en la Tabla 4.2 para el experimento considerado. Para la mayor parte de los firmantes, la especificidad es mayor que la sensibilidad.

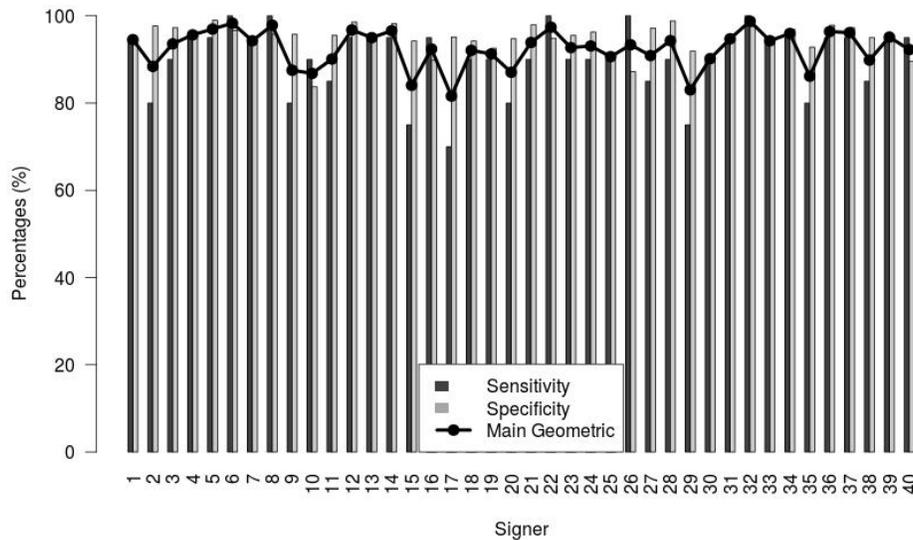


Figura. 4.1. Resultados de la Clasificación en la identificación de firmantes.

Como se ve en la figura 4.2, se indica el rendimiento entregado por el clasificador, para este caso, SVM; de derecha a la izquierda se organizan las características fractales de la más alta, al rendimiento más bajo. Además, la figura muestra el rendimiento obtenido por cada característica con relación a los firmantes que se encuentran en la Base de Datos. El cuarto firmante, para el sistema propuesto es el firmante más fácil de reconocer o identificar, caso contrario sucede con el firmante número 9, por tener o poseer el rendimiento más bajo, por tanto, es el firmante más difícil de reconocer por el sistema. Sin embargo, el índice más bajo de rendimiento presentado por el sistema es superior al 60%, por lo tanto, el sistema tiene un índice de identificación eficiente para la identificación de firmas dinámicas propuesto en este proyecto.

También se puede observar que la característica Dimensión de Correlación, dentro del conjunto de características propuesto, presenta el rendimiento más bajo, sólo que en este caso, se muestran algunas excepciones; el firmante número 32 con respecto a la característica (Dimensión de Correlación) presenta un índice de rendimiento bueno, en comparación con otros firmantes como: 20, 35, entre otros.

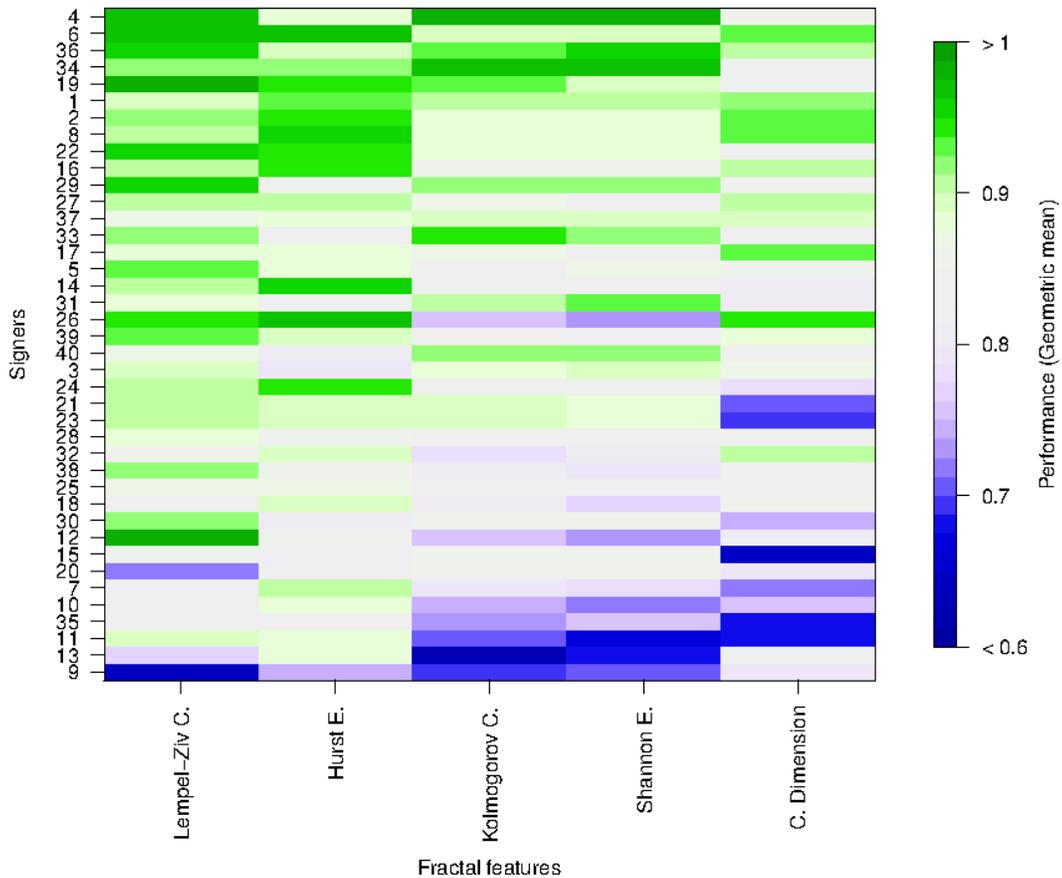


Figura. 4.2. Desempeño obtenido por el sistema propuesto, haciendo una organización con las características no-lineales más relevantes para la identificación de firmantes.

La presión se asocia con la dinámica psicomotoras de los firmantes, por lo tanto, es un patrón discriminante para el reconocimiento de firmantes, como se ve en la figura 4.3.

La figura 4.3 representa la clasificación de rendimiento obtenido, después de extraer las características fractales propuestos en este trabajo. Cabe señalar, ese resultado se muestra con el propósito de proporcionar las prestaciones de las características fractales (entropía de Shannon, la dimensión de correlación, el exponente de Hurst, la complejidad de Kolmogorov, la complejidad de Lempel-Ziv) con respecto a cada atributo de la firma de cada persona o individuo. Estos atributos se obtuvieron del dispositivo de adquisición WACOM descrito en la sección (3.1.1). El atributo de presión es el más discriminante, en comparación con los otros atributos propuestos; con respecto al firmante número 4 es el

más fácil de reconocer, y el firmante 9 es el más difícil de reconocer por el algoritmo propuesto.

La figura 4.3 muestra que el peor atributo para identificar cualquier firmante es la altitud, ya que presenta un bajo rendimiento con respecto a las otras. Pero, con algunas excepciones, como el firmante número 27.

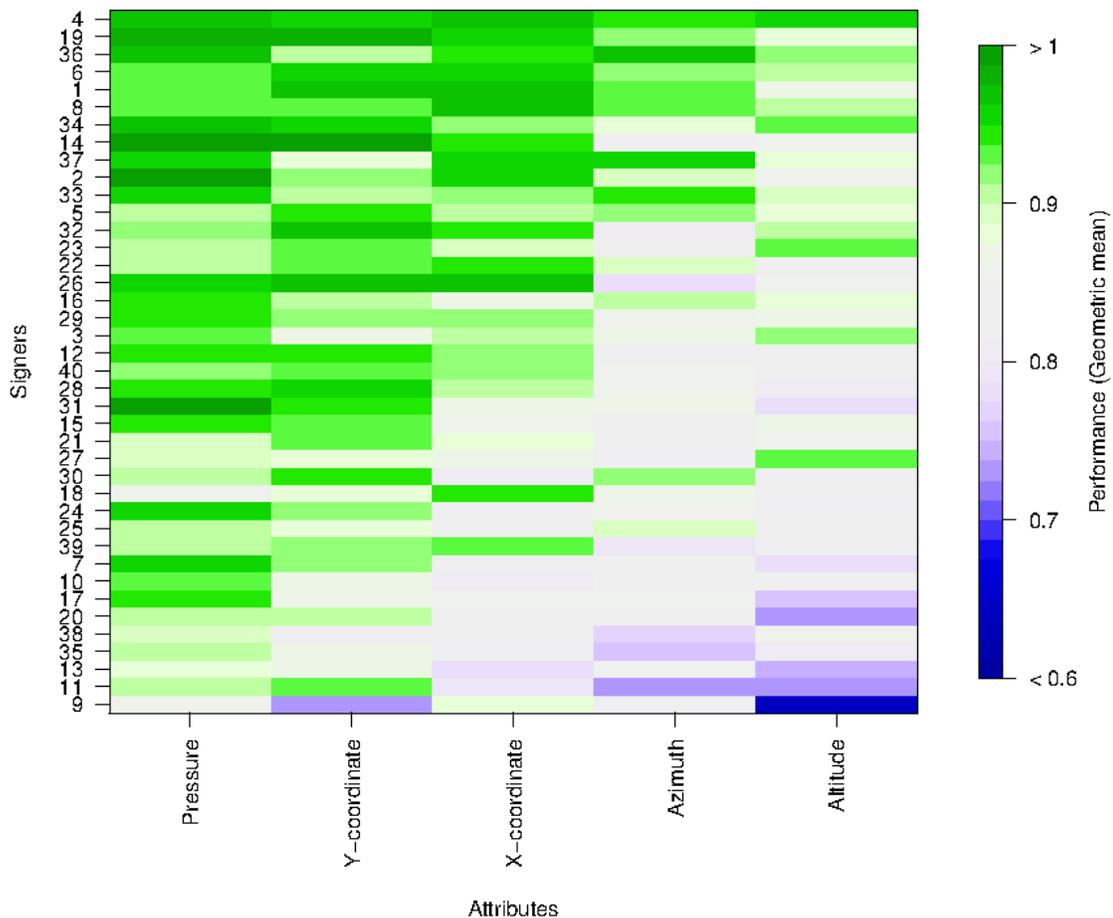


Figura. 4.3. Desempeño obtenido por el sistema propuesto, haciendo una organización con los atributos más relevantes para la identificación de firmantes.

También se puede apreciar que en la figura 4.2, LZ (Lempel-Ziv Complejidad) presenta los mejores resultados; esta característica fractal es capaz de detectar o localizar patrones únicos en una serie de tiempo, lo que permite estimar qué tan compleja o irregular es una

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

serie de tiempo que representa la dinámica intrínseca del firmante. En este sentido, para una serie temporal, la complejidad de LZ se utiliza como una medida de la estructura subyacente de la serie temporal.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

4 CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO

Se desarrolló una metodología para la identificación de firmas dinámicas utilizando características fractales y máquinas de vectores soporte. Como la información biométrica contiene operadores no lineales embebidos, se llevó a cabo para el tratamiento propuesto el uso de técnicas de dinámica no lineal, que pueden capturar la dinámica intrínseca del firmante para detectar información subyacente en la serie de tiempo ayudando a entender el proceso de firma por un individuo. Por lo tanto, se analizó y desarrolló modelos que representan la firma dinámica generada por la posición en los ejes (x, y), la presión, el acimut y la inclinación; utilizando las siguientes medidas fractales: la complejidad de Lempel-Ziv, dimensión de correlación, la entropía de Shannon, Kolmogorov entropía y el Exponente de Hurst. La estimación de características contiene la suficiente consistencia para la clasificación de los individuos.

En definitiva los resultados obtenidos parecen ser prometedores en el área de la seguridad con el fin de permitir la identificación del firmante, con un equilibrio general entre especificidad y sensibilidad calculada con la media geométrica de 92,5%, optando que el análisis de los sistemas caóticos o fractales aplicados al comportamiento dinámico de la firma pueden generar resultados muy prometedores.

Para la propuesta presentada se evaluó una estrategia de validación (*Cross Validation k-folds*), en busca de observar los resultados de manera más representativa, es decir, observar la poderosa capacidad de generalización que tiene las Máquinas de Vectores de Soporte (SVM); que para nuestro caso es el clasificador presentado y que busca contener la suficiente consistencia de las características en la identificación de individuos. Para la validación de la metodología se utiliza la base de datos propia presentada en este proyecto y descrita en la sección 3.1.1. Con la base de datos propia se obtuvo un

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

desempeño de identificación con una Sensibilidad del 91,50% y una Especificidad del 94.75% respectivamente. La medición del desempeño del clasificador se realizó con la matriz de confusión, y que, con base a esos datos se extrajeron la sensibilidad y la especificidad ya mencionadas anteriormente.

En definitiva, se puede establecer que la metodología presentada tiene buenos resultados y que la en la gran mayoría de los casos (más del 90%) se pudo identificar todas las clases existentes en la base de datos, además, el algoritmo identifica la firma perteneciente a un individuo, y está en la capacidad de mostrar que una firma sólo pertenece a un firmante y no a varios. Esto es demostrado con la especificidad. No obstante el desempeño obtenido con la metodología propuesta puede compararse con los métodos reportados en la literatura, resulta necesario ampliar la población de estudio para verificar la metodología propuesta.

Como trabajo futuro se propone incluir el mayor máximo Exponente de Lyapunov y otras medidas de entropía como entropía Aproximada, la entropía Muestreada, Entropía Gaussiana y Entropía Difusa para mejorar la caracterización de la dinámica no lineal de las firmas dinámicas y, probablemente, aumentar el rendimiento de la metodología propuesta.

Desarrollar una metodología utilizando clasificadores de una sola clase en la identificación de firmas dinámicas para sistemas abiertos, utilizando análisis Fractal a partir de las características biométricas comportamentales del individuo firmante.

Proponer una metodología con las características planteadas en este proyecto en el estudio de las señales de electromiografía, para la adecuada fabricación y puesta en marcha de prótesis para personas que hayan perdido alguna extremidad.

Proponer una metodología con características no lineales al estudio de enfermedades mentales, tal como la epilepsia; causada por el desequilibrio en la actividad eléctrica de las neuronas en alguna zona del cerebro. Estas señales se pueden asociar a series caóticas por contener información subyacente en sus patrones definidos por tal enfermedad.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

REFERENCIAS

Ahmedt, D. E., Trejos, E. D., & Vargas, J. F. (2011). Generación de espacios de representación de firmas dinámicas: una revisión enfocada al análisis de complejidad. *Scientia et Technica*, 2(48), 244-252.

Ahmedt-Aristizabal, D., Delgado-Trejos, E., Vargas-Bonilla, J. F., & Jaramillo-Garzon, J. A. (2011, October). Dynamic signature for a closed-set identification based on nonlinear analysis. In *Biometrics (IJCB), 2011 International Joint Conference on* (pp. 1-8). IEEE.

Alariki, A. A., & Manaf, A. A. (2014). Behavioral Biometrics Authentication Using Gesture-Based. *Advanced Science Letters*, 20(2), 492-495.

Al-Mayyan, W., Own, H. S., and Zedan, H. (2011). Rough set approach to online signature identification. *Digital Signal Processing*, 21(3):477–485.

Bailey, K. O., Okolica, J. S., & Peterson, G. L. (2014). User identification and authentication using multi-modal behavioral biometrics. *Computers & Security*, 43, 77-89.

Betancourt, G. A. (2005). Las máquinas de soporte vectorial (SVMs). *Scientia et Technica*, 1(27).

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 12). New York: springer.

Bolle, R., Connell, J., Pankanti, S., Ratha, N., & Senior, A. (2005). The Relation between the ROC Curve and the CMC. In *Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, 15-20.

Carvajal, R., Vallverdú, M. and Caminal, P. (2000). Análisis no lineal de la variabilidad de la frecuencia cardíaca en casos normales y cardiopatías. *Revista Mexicana de Ingeniería Biomédica*, 21(2), 29-34.

Casper. D., Becker-Szendy. R. et al, (1991). Measurement of atmospheric neutrino composition with the IMB-3 detector, *Phys. Rev. Lett*, V. 66,issue 20, American Physical Society, May, p. 2561-2564.

Castellanos, G. (2005). Identificación de estados funcionales en bioseñales. *Voz, ECG y Fonocardiografía*. Tech. Rep. CS-96-09, Universidad Nacional de Colombia, Manizales, Colombia.

Chen, W., Wang, Z., Xie, H. & Yu, W. (2007). Characterization of surface EMG signal based on fuzzy entropy. *IEEE Transactions on neural systems and rehabilitation engineering*, 15(2), 266-72.

Chi, Z., Wu, W., Haga, Z., Hatsopoulos, N. G., & Margoliash, D. (2007). Template-based spike pattern identification with linear convolution and dynamic time warping. *Journal of neurophysiology*, 97(2), 1221-1235.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Cappelli, R., Maio, D., Maltoni, D., Wayman, J. L., & Jain, A. K. (2006). Performance evaluation of fingerprint verification systems. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(1), 3-18.

Delgado, E., Castellanos, G. & Vallverdú, M. (2009). Análisis de Relevancia en Espacios de Representación orientado al Soporte de Diagnóstico. Editorial Universidad Nacional de Colombia, Manizales, Colombia.

Delgado-Trejos, E., Perera-Lluna, A., Vallverdú-Ferrer, M., Caminal-Magrans, P., & Castellanos-Domínguez, G. (2009). Dimensionality reduction oriented toward the feature visualization for ischemia detection. *Information Technology in Biomedicine, IEEE Transactions on*, 13(4), 590-598.

Dessimoz, D., Richiardi, J., Champod, C., and Drygajlo, A. (2007). Multimodal biometrics for identity documents (). *Forensic science international*, 167(2):154–159.

Di Martino. Matías, Decia. Federico, Molinelli. Juan y Fernández. Alicia. (2012). Improving electric fraud detection using class imbalance strategies, *International Conference on Pattern Recognition Applications and Methods, ICPRAM*, p.135-141.

Di Matteo, T. (2007). Multi-scaling in finance. *Quantitative finance*, 7(1), 21-36.

Dullink, H., van Daalen, B., Nijhuis, J., Spaanenburg, L., Zuidhof, H., & EFRIE, F. (1995). Implementing a DSP kernel for online dynamic handwritten signature verification using the TMS320 DSP family. EFRIE, France December.

Dumas, B., Pugin, C., Hennebert, J., Humm, D.P.D.A., Evequoz, F., Ingold, R. & von Rotz, D. (2005). MylDea - Multimodal biometrics database, description of acquisition-protocols. In *Proceedings of Third COST 275 Workshop (COST 275)*, 59–62, Hatfield (UK).

Eberhart, R. C. and Shi, Y. (2001). Particle swarm optimization: developments, applications and resources. In *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, volume 1, pages 81–86. IEEE.

Faul S., Boylan, G., Connolly, S., Marnane, W., & Lightbody, G. (2005, September). Chaos theory analysis of the newborn EEG-is it worth the wait?. In *IEEE Int. Workshop on Intelligent Signal Processing* (pp. 381-386).

Faundez-Zanuy, M. (2007). On-line signature recognition based on VQ-DTW. *Pattern Recognition*, 40(3), 981-992.

Fan, J. C. S., Hastie, T. J., & Kishon, E. (1992). U.S. Patent No. 5, 111, 512. Washington, DC: U.S. Patent and Trademark Office.

Fierrez-Aguilar, J., Nanni, L., Lopez-Peñalba, J., Ortega-García, J., and Maltoni, D. (2005). An on-line signature verification system based on fusion of local and global information. In *Audio-and video-based biometric person authentication*, pages 523–532. Springer.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Fierrez, J., Ortega-Garcia, J., Ramos, D., & Gonzalez-Rodriguez, J. (2007). HMM-based on-line signature verification: Feature extraction and signature modeling. *Pattern Recognition Letters*, 28(16), 2325-2334.

Garcia-Salicetti, S., Beumier, C., Chollet, G., Dorizzi, B., les Jardins, J. L., Lunter, J., & Petrovska Delacretaz, D. (2003, January). BIOMET: A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities. In *Audio-and Video-Based Biometric Person Authentication* (pp. 845-853). Springer Berlin Heidelberg.

Grassberger, P., & Procaccia, I. (1983). Characterization of strange attractors. *Physical review letters*, 50(5), 346.

Goldberger, A. L. et al. (1990). Chaos and fractals in human physiology. *Scientific American*, 262(2):42-49.

Hamilton, D. J., Whelan, J., McLaren, A., MacIntyre, I., & Tizzard, A. (1995). Low cost dynamic signature verification system.

Hangai, S., Yamanaka, S., & Hanamoto, T. (2000). On-line signature verification based on altitude and direction of pen movement. In *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on* (Vol. 1, pp. 489-492). IEEE.

Huang, C. L., & Wang, C. J. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with applications*, 31(2), 231-240.

Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1), 4-20.

Jain, A., Bolle, R. & Pankanti, S. (1999). *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers.

Kantz, H. & Schreiber, T. *Nonlinear Time Series Analysis*. Cambridge University Press, 2002.

Kameya, H., Mori, S., & Oka, R. (2003, August). Figure-based writer verification by matching between an arbitrary part of registered sequence and an input sequence extracted from on-line handwritten figures. In *2013 12th International Conference on Document Analysis and Recognition* (Vol. 2, pp. 985-985). IEEE Computer Society.

Kaspar, F. & Schuster, H. (1987). Easily calculable measure for the complexity of spatiotemporal patterns. *Physical Review A*, 36(2), 842-848.

Kato, M., & Kawashima, Y. (1993). Signature verification using online data such as pen pressure and velocity.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Kumar, D., & Unikrishnan, P. (2013). Class specific feature selection for identity validation using dynamic signatures. *J Biomet Biostat*, 4(160), 2.

Lauchenbruch, P. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, 23:639 – 645.

Lempel A, & Ziv J. (1976). On the complexity of finite sequences. *IEEE Trans. Inf. Theory*; 22, 75-81.

Lin, S.-W., Ying, K.-C., Chen, S.-C., and Lee, Z.-J. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, 35(4):1817–1824.

Lin, C. J. and Hsieh, M.-H. (2009). Classification of mental task from eeg data using neural networks based on particle swarm optimization. *Neurocomputing*, 72(4):1121– 1130.

Longstaff, M. G. and Heath, R. A. (1999). A nonlinear analysis of the temporal characteristics of handwriting. *Human Movement Science*, 18(4):485–524.

Mackenzie. Ian R.A., Neumann. Manuela, Baborie. Atik, Sampathu. Deepak M., Plessis. Daniel, Jaros. Evelyn, Perry. Robert H., Trojanowski. John Q., Mann. David M.A. y Lee. Virginia. (2011). A harmonized classification system for FTLD-TDP pathology, *Acta Neuropathologica*, Springer-Verlag, July, Volume 122, Issue 1, pp 111-113.

Mangasarian, O. L. (1999). Generalized support vector machines. *Advances in Neural Information Processing Systems*, 135-146.

Martens, R., & Claesen, L. (1997, August). Dynamic programming optimisation for on-line signature verification. In *Document Analysis and Recognition, 1997.*, Proceedings of the Fourth International Conference on (Vol. 2, pp. 653-656). IEEE.

Matheus. David, Villazana. Sergio y Seijas. Cesar. (2010). Detección de Arritmias en Señales de ECG Basado en Máquinas de Vectores de Soporte. *Revista Ingeniería A UC*, Vol. 17, Nro. 3, Dic, p. 49 – 56.

MATLAB (Matrix Laboratory): Is a language of technical computing, is a programming environment for algorithm development, data analysis, visualization, and numeric computation. Developed by MathWorks Inc, Natick, Massachusetts, U.S.A, 1984. URL <http://www.mathworks.com/products/matlab/>. English Version ISBN: 978-0-9896-140-23.

Mattera, D., Palmieri, F., & Haykin, S. (1999). An explicit algorithm for training support vector machines. *Signal Processing Letters, IEEE*, 6(9), 243-245.

Morales, R., Di Matteo, T., Gramatica, R., & Aste, T. (2012). Dynamical generalized Hurst exponent as a tool to monitor unstable periods in financial time series. *Physica A: Statistical Mechanics and its Applications*, 391(11), 3180-3189.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Nanni, L. and Lumini, A. (2006). Advanced methods for twoclass problem formulation for on-line signature verification. *Neurocomputing*, 69(7):854–857.

Nelson, W., & Kishon, E. (1991, October). Use of dynamic features for signature verification. In *Systems, Man, and Cybernetics, 1991. 'Decision Aiding for Complex Systems, Conference Proceedings., 1991 IEEE International Conference on* (pp. 201-205). IEEE.

Ohishi, Taguchi., Komiya, Y., & Matsumoto, T. (2000, September). On-line signature verification using pen-position, pen-pressure and pen-inclination trajectories. In *Pattern Recognition, International Conference on* (Vol. 4, pp. 4547-4547). IEEE Computer Society.

Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez-Zanuy, M., Espinosa, V. et al. (2003). MCVT baseline corpus: A bimodal biometric database. *IEEE Proceedings Vision, Image and Signal Processing*, 150(6), 395-401.

Ortega-Garcia, J., Fierrez, J., Alonso-Fernandez, F., Galbally, J., Freire, M., Gonzalez-Rodriguez, J., Garcia-Mateo, C., Alba-Castro, J.L., Gonzalez-Agulla, E., Otero-Muras, E., Garcia-Salicetti, S., Allano, L., Ly-Van, B., Dorizzi, B., Kittler, J., Bourlai, T., Poh, N., Deravi, F., Ng, M., Fairhurst, M., Hennebert, J., Humm, A., Tistarelli, M., Brodo, L., Richiardi, J., Drygajlo, A., Ganster, H., Sukno, F., Pavani, S.K., Frangi, A., Akarun, L. & Savran, A. (2009). The multi-scenario multi-environment biosecure multimodal database (bmdb). *IEEE Trans. on Pattern Analysis and Machine Intelligence*.

Pascual-Gaspar, J. M., Cardeñoso-Payo, V., & Vivaracho-Pascual, C. E. (2009). Practical on-line signature verification. In *Advances in Biometrics* (pp. 1180-1189). Springer Berlin Heidelberg.

Pascual Gaspar, Juan Manuel., (2010), *Uso de la Firma Manuscrita Dinámica para el Reconocimiento Biométrico de Personas en Escenarios Prácticos*, Doctorado, Universidad de Valladolid, Escuela Técnica Superior de Ingeniería Informática - Departamento de Informática, Valladolid, España.

Pascual-Gaspar, J. M., Faundez-Zanuy, M., & Vivaracho, C. (2011). Fast on-line signature recognition based on VQ with time modeling. *Engineering Applications of Artificial Intelligence*, 24(2), 368-377.

Phillips, P. J., Flynn, P. J., Beveridge, J. R., Scruggs, W. T., O’toole, A. J., Bolme, D., & Weimer, S. (2009). Overview of the multiple biometrics grand challenge. In *Advances in Biometrics* (pp. 705-714). Springer Berlin Heidelberg.

Phillips, P. J., Grother, P., & Micheals, R. (2011). *Evaluation methods in face recognition* (pp. 551-574). Springer London.

Phillips P. J., P.J.R., H. Moon & Rizvi, S. (2006). The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 1090–1104. 2.

Phillips, P.J. (2006a). *Face and iris evaluations at NIST. CardTech/SecurTech. 2.*

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Phillips, P.J. (2006b). FRGC and ICE workshop.

http://iris.nist.gov/ICE/ICE_2005_Results_30March2006.pdf. 24

Plamondon, R., & Lorette, G. (1989). Automatic signature verification and writer identification—the state of the art. *Pattern recognition*, 22(2), 107-131.

Pallikari, F., & Boller, E. (1999). A Rescaled Range Analysis of Random Events. *Journal of Scientific Exploration*, 13 (1).

Provost, F. J., Fawcett, T., & Kohavi, R. (1998, July). The case against accuracy estimation for comparing induction algorithms. In *ICML* (Vol. 98, pp. 445-453).

Przybocki, M.A., Martin, A.F. & Le, A.N. (2006). NIST speaker recognition evaluation chronicles - part 2. In *Speaker and Language Recognition Workshop, IEEE Odyssey*, 1–6. 2, 24.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL <http://www.R-project.org/>. ISBN 3-900051-07-0. 5.

Schmidt, T., Riffo, V., & Mery, D. (2011). Dynamic signature recognition based on fisher discriminant. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (pp. 433-442). Springer Berlin Heidelberg.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT press.

Schölkopf B., K.-K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, “Comparing support vector machines with Gaussian kernels to radial basis function classifiers,” *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2758–2765, 1997. [Online]. Available: citeseer.ist.psu.edu/sch97comparing.html.

Shafiei, M. M., & Rabiee, H. R. (2003, August). A new online signature verification algorithm using variable length segmentation and hidden Markov models. In *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on* (pp. 443-446). IEEE.

Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Tech. J*, 27(7-10), 379-423.

Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Tech. J*, 27(7-10), 379-423.

Small, M. (2005). *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*. Nonlinear Science Series A, 52.

Sakamoto, D., Morita, H., Ohishi, T., Komiya, Y., & Matsumoto, T. (2001, May). On-line signature verification algorithm incorporating pen position, pen pressure and pen inclination trajectories. In

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Acoustics, Speech, and Signal Processing, IEEE International Conference on (Vol. 2, pp. 993-996). IEEE.

Solé, R. V., Bascompte, J., Delgado, J., Luque, B., and Manrubia, S. C. (1996). Complejidad en la frontera 7 del caos. *Investigación y Ciencia: Edición Española de Scientific American*, (236):14–21.

Sprott, J. and Rowlands, G. *Chaos Data Analyzer: The Professional Version*. American Institute of Physics, Physics Academic Software, New York, 1995.

Sprott, J. (2003). *Chaos and Time-Series Analysis*. Oxford University Press.

Sprott, J. C. and Sprott, J. C. (2003). *Chaos and time-series analysis*, volume 69. Oxford University Press Oxford.

Subias, J. (1991). *Introducción a la teoría del caos. Representando atractores por orbitales*. University of Zaragoza, Spain.

Tsai, M. Y., & Lan, L. S. (2005, October). Online writer identification using the point distribution model. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on* (Vol. 2, pp. 1264-1268). IEEE.

Thumwarin, P., & Matsuura, T. (2004, October). On-line writer recognition for thai based on velocity of barycenter of pen-point movement. In *Image Processing, 2004. ICIP'04. 2004 International Conference on* (Vol. 2, pp. 889-892). IEEE.

Vapnik, V. N., & Vapnik, V. (1998). *Statistical learning theory* (Vol. 1). New York: Wiley.

Walters, P. (1982). *An Introduction to Ergodic Theory*. G. T. in Mathematics, Ed. Springer, 79.

Wang W., Z. Xu, W. Lu, and X. Zhang, "Determination of the spread parameter in the gaussian kernel for classification and regression," *Neurocomputing*, vol. 55, no. 3-4, pp. 643–663, 2003. [Online]. Available: [http://dx.doi.org/10.1016/S0925-2312\(02\)00632-X](http://dx.doi.org/10.1016/S0925-2312(02)00632-X).

Wu, Q. Z., Lee, S. Y., & Jou, I. C. (1997). On-line signature verification based on split-and-merge matching mechanism. *Pattern Recognition Letters*, 18(7), 665-673.

Yeung, D., Chang, H., Xiong, Y., George, S., Kashi, R., Matsumoto, T. & Rigoll, G. (2004). SVC2004: First international signature verification competition. In *Proc. of the First International Conference on Biometrics Authentication (ICBA 2004)*, 16–22.

Zhang, X., Zhu, Y., Thakor, N. & Wang, Z. (1999). Detecting Ventricular Tachycardia and Fibrillation by Complexity Measure. *IEEE Transactions on Biomedical Engineering*, 46(5), 548-55.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

<p>FIRMA ESTUDIANTES _____</p> <p>_____</p> <p>_____</p> <p>FIRMA ASESOR _____</p> <p style="text-align: right;">FECHA ENTREGA: _____</p>

<p>FIRMA COMITÉ TRABAJO DE GRADO DE LA FACULTAD _____</p> <p>RECHAZADO ___ ACEPTADO ___ ACEPTADO CON MODIFICACIONES _____</p> <p style="text-align: right;">ACTA NO. _____</p> <p style="text-align: right;">FECHA ENTREGA: _____</p>

<p>FIRMA CONSEJO DE FACULTAD _____</p> <p style="text-align: right;">ACTA NO. _____</p> <p style="text-align: right;">FECHA ENTREGA: _____</p>
--