



Universidad Autónoma de San Luis
Potosí

Facultad de Ciencias

**Extracción y análisis de características
con la Transformada Wavelet para
el reconocimiento de imágenes**

TESIS QUE PARA OBTENER EL TÍTULO DE

MAESTRO EN CIENCIAS

PRESENTA:

I.E. JORGE DANIEL CÁRDENAS AMAYA

ASESORES:

DR. JOSÉ SALOMÉ MURGUÍA IBARRA

DRA. MARCELA MEJÍA CARLOS

San Luis Potosí, S. L. P., Agosto del 2018

Extracción y análisis de características con la Transformada Wavelet para el reconocimiento de imágenes

Nombre del estudiante:

Jorge Daniel Cárdenas Amaya

Comité que acepta la tesis

Dr. José Salomé Murguía Ibarra (*Asesor*) _____

Dra. Marcela Mejía Carlos (*Asesor*) _____

Dr. Marco Tulio Ramírez Torres (*Sinodal*) _____

Dr. Raúl E. Balderas Navarro (*Sinodal*) _____

Dedicatoria

A mis padres Jorge Cárdenas Estrada y Genoveva Amaya Mendoza por sus consejos, sus valores, por la motivación constante que me ha permitido ser una persona de bien y por ser siempre un ejemplo de perseverancia y constancia, todo esto es gracias a ustedes que me enseñaron el valor del trabajo y que aún en mis momentos más difíciles me ayudaron a seguir adelante, ustedes que siempre estuvieron presentes en cada etapa de mi desarrollo educativo y personal mostrándome siempre el camino correcto, pero sobre todo, por su amor. Todo lo que soy se los debo a ustedes por lo que siempre les estaré agradecido, los amo.

A mis hermanas Nallely Roxana y Saraly Nohemí por estar conmigo y apoyarme siempre a lo largo de todos estos años, el amor y cariño que nos tenemos como familia nos ha fortalecido para ser lo que nuestros padres siempre desearon y el apoyo que ustedes me dieron espero poder devolvérselos de igual manera. El amor de hermanos es el que nos permite hacer grandes cosas con nuestras vidas y por eso les estoy agradecido.

Finalmente a toda mi familia, profesores, amigos y compañeros quienes participaron directa o indirectamente en la elaboración de este trabajo, nada de esto hubiese sido posible sin su apoyo ya que a su lado compartí buenos momentos y supere todo los malos, ustedes fueron para mi una guía que me ayudó a alcanzar todas mis metas y que sin duda me alentaron a proponerme retos cada vez mayores. Por todo esto, les doy las gracias.

Agradecimientos

A la Universidad Autónoma de San Luis Potosí, por ser la sede en la que realicé todos mis estudios profesionales, a la Facultad de Ciencias, al Instituto de Investigación en Comunicación Óptica y a todo su personal tanto académico como administrativo por haber compartido conmigo todos sus conocimientos a lo largo de toda mi preparación académica.

A mis asesores, Dr. José Salomé Murguía Ibarra, Dra. Marcela Mejía Carlos, por todo el tiempo invertido y paciencia que me brindaron a lo largo de la realización de este trabajo, gracias a sus consejos y su orientación me guiaron a la culminación de esta investigación. Toda mi admiración hacia ustedes.

A cada uno de los profesores que me transmitieron sus conocimientos a lo largo de todo el programa de la Maestría en Ciencias Aplicadas, su sentido de la responsabilidad y consejos me orientaron a culminar mis estudios.

A mis sinodales, Dr. José Salomé Murguía Ibarra, Dra. Marcela Mejía Carlos, Dr. Marco T. Ramírez Torres, Dr. Raúl E. Balderas Navarro, por el tiempo y las atenciones brindadas en la realización del presente trabajo de tesis.

A CONACYT por el apoyo recibido durante toda mi maestría, sin los recursos brindados nada esto hubiese sido posible.

A mis amigos Francisco Javier, Ricardo, José, Raúl Iván, Raúl Machuca, Juan Carlos, Gonzalo, Diomar, Gerardo, Erika, Iyari, Caro, Emanuel, Daniel Medina, Luis, Andrés, Rosario, Claudia, Iván Israel, Miguel Cardona, Miguel Silva, Cruz, Ociel, Leonardo, César, Abel, Everardo, Óscar, Cynthia, Etna, Yadira y a todas las personas que estuvieron presentes a lo largo de mi vida ya que me brindaron en algún momento su apoyo y amistad.

A Sol, por haberme brindado todo su cariño y apoyo durante todo este tiempo y por convertirse en una fuente de inspiración, por estar a mi lado compartiendo todos los buenos momentos y apoyándome en los más difíciles. Te quiero mucho.

Resulta muy difícil agradecer a todas las personas que estuvieron involucradas directa o indirectamente, sin embargo, pido disculpas por las omisiones que obviamente fueron involuntarias.

Resumen

En los últimos años el reconocimiento facial ha sido una parte importante en la mayoría de las investigaciones en el campo de la visión computacional. Dentro de las múltiples aplicaciones que existen en los sistemas de reconocimiento, la extracción de características es considerada una de las partes más importantes. La transformada wavelet en su forma discreta bidimensional, debido a sus propiedades, ha demostrado ser además de un buen extractor de características y un excelente método para minimizar el costo computacional de cualquier proceso en cuestión del tiempo de ejecución. En este trabajo se presenta la creación de diferentes algoritmos para la extracción de características en imágenes faciales con la intención de diseñar e implementar un sistema de identificación de personas. Se estudian y se prueban algunas de las funciones Wavelet dentro de un sistema de reconocimiento facial y se proponen el uso de sistemas de transformación híbrida wavelet. Las primeras pruebas se llevaron a cabo con base a la energía contenida en las características extraídas y se variaron las condiciones de adquisición para las imágenes de prueba, en las siguientes se utilizaron los valores de todas las características extraídas y de igual forma fueron variadas las condiciones de adquisición. Los resultados obtenidos permitieron determinar la eficacia de cada una de las funciones wavelet en las etapas de extracción de características y a su vez logrando implementar sistemas de reconocimiento de personas de forma robusta y de alto desempeño en términos de el porcentaje de éxito y tiempo de ejecución de las operaciones de clasificación.

Índice general

1. INTRODUCCIÓN	1
1.1. Objetivos	2
1.2. Organización	3
2. DETECCIÓN Y PRE-PROCESADO DE IMÁGENES FACIALES	4
2.1. Detección de rostros	4
2.1.1. Algoritmo Viola-Jones	5
2.2. Pre-procesado de imágenes	8
3. EXTRACCIÓN DE CARACTERÍSTICAS POR TRANSFORMADAS WAVELET	10
3.1. Preliminares	10
3.1.1. Expansión en series de señales	11
3.1.2. Transformada wavelet	12
3.1.3. Transformada wavelet discreta	13
3.1.4. Análisis Multi-resolución	15
3.2. Transformada de Haar	17
3.3. Transformada Daubechies	20
3.4. Transformación wavelet packet	23
3.5. Transformada wavelet híbrida	25
3.6. Transformada wavelet Bidimensional	26
3.7. Estimación de la calidad de los coeficientes extraídos	29
4. CLASIFICACIÓN DE IMÁGENES	33

4.1. Algoritmos de clasificación	35
4.1.1. Análisis de Componentes Principales	35
4.1.2. Análisis Discriminante	36
4.1.3. Máquinas de Soporte Vectorial	37
4.1.4. K- Vecindarios Cercanos	38
4.1.5. Clasificadores en Conjunto	40
4.2. Evaluación de los clasificadores	42
4.2.1. Prueba 1: Análisis por energía	43
4.2.2. Prueba 2: Análisis por energía	49
4.2.3. Prueba 3: Análisis de coeficientes extraídos	54
4.2.4. Prueba 4: Análisis de coeficientes extraídos	61
5. CONCLUSIONES Y TRABAJO FUTURO	68
Bibliografía	70

Índice de figuras

1.1. Diagrama del sistema de identificación de personas implementado y basado en la extracción de características por TW.	2
2.1. Descriptores rectangulares de Haar para bordes (a) horizontales y (b) verticales y para líneas (c) horizontales y (d) verticales.	6
2.2. Ilustración de una imagen integral.	7
2.3. Esquema de un clasificador en cascada.	8
2.4. Detección de (a) rostro y (b) ojos, nariz y boca a través del algoritmo Viola-Jones.	8
2.5. Segmentación del área de los ojos, nariz y boca.	9
2.6. Ejemplo de interpolación bicúbica de vecindario 4×4 aplicada al área de los ojos.	9
3.1. Funciones wavelet madre de (a) Haar y (b) Daubechies de cuatro coeficientes.	13
3.2. Algoritmo piramidal de descomposición.	16
3.3. (a) Análisis mediante Bancos de Filtros, (b) descomposición del espectro en frecuencia de la señal al ser procesada en tres niveles de transformación y (c) BF para el análisis y síntesis respectivamente de una señal.	18
3.4. Funciones de (a) escala y (b) wavelet de Haar, respectivamente.	19
3.5. Funciones Daubechies, expresadas de derecha a izquierda, función de escalamiento y función wavelet para (a) soporte D2, (b) soporte D4 y (c) soporte D6 respectivamente.	21
3.6. Diagrama esquemático de la descomposición por wavelet packets, donde 'V' y 'W' de izquierda a derecha denota el filtro por el que ha pasado previamente la señal.	23
3.7. Transformada wavelet Packet de segundo nivel desde el enfoque de los BF, es mostrado en (a) el banco de análisis y en (b) el banco de síntesis	24
3.8. Diagrama de descomposición híbrido (Banco de análisis) piramidal para señales en una dimensión visto con el enfoque de los Bancos de Filtros.	25

3.9. Diagrama a bloques del sistema híbrido de 2 niveles de dos funciones wavelet; en el primer caso se muestra la TW Híbrida Haar-Db4 y en el segundo la TW Híbrida Db4-Haar.	26
3.10. Diagrama de descomposición (Banco de análisis) piramidal para señales en dos dimensiones visto con el enfoque de los Bancos de Filtros.	28
3.11. Esquema de los tres primeros niveles de transformación para una imagen. En este se muestra el acomodo de cada una de las imágenes resultantes de acuerdo con las operaciones aplicadas	29
3.12. Error relativo de norma 2 a) sub-imágenes promedio, b) sub-imágenes de detalles horizontales, c) sub-imágenes de detalles verticales y d) sub-imágenes de detalles diagonales.	31
4.1. Diseño de una máquina de aprendizaje, donde los resultados de la evaluación pueden requerir la repetición de algunos pasos para obtener los mejores resultados.	34
4.2. Ejemplo de A) matriz de dispersión de las clases y B) matriz de dispersión entre las clases.	37
4.3. Ejemplo de una máquina de soporte vectorial de clasificación binaria.	38
4.4. Diagrama a bloques de un sistema de clasificación en conjunto.	40
4.5. Gráficas de dispersión usando PCA del espacio de características generado por a) TW Packet de 3 niveles en el área de la boca, b) TW Haar de un nivel en el área de la nariz, c) TW Haar-DB4 de dos niveles en el área de los ojos y d) TW DB4 de 3 niveles en el área del rostro.	44
4.6. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4.	46
4.7. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar.	46
4.8. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4-Haar.	47
4.9. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar-DB4.	47
4.10. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de Haar.	48
4.11. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de DB4.	48

4.12. Comparación entre las gráficas de dispersión generadas por el PCA de la TW de Haar en un nivel del área de los ojos de a) la Prueba 1 de experimentación y b) de la Prueba 2.	50
4.13. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4.	51
4.14. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar.	52
4.15. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4-Haar.	52
4.16. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar-DB4.	53
4.17. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de Haar.	53
4.18. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de DB4.	54
4.19. Gráficas de dispersión generadas por el PCA de a) la versión promedio de la TW DB4, b) la versión de detalles horizontales de la TW DB4 , c) la versión promedio de la TW Haar y d) la versión de detalles horizontales de la TW Haar, todas en un primer nivel de transformación del área de los ojos.	56
4.20. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4.	58
4.21. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar.	58
4.22. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4-Haar.	59
4.23. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar-DB4.	59
4.24. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de Haar.	60
4.25. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de DB4.	60
4.26. Gráficas de dispersión generadas por el PCA de a) la versión promedio de la TW DB4, b) la versión de detalles horizontales de la TW DB4 , c) la versión promedio de la TW Haar y d) la versión de detalles horizontales de la TW Haar, todas en un primer nivel de transformación del área de los ojos.	62

4.27. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4.	64
4.28. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar.	64
4.29. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4-Haar.	65
4.30. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar-DB4.	65
4.31. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de Haar.	66
4.32. Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de DB4.	66

Índice de tablas

- 3.1. Dimensión de las sub-imágenes obtenidas por la TW 30

- 4.1. Descripción de las pruebas realizadas 43
- 4.2. Resultados de la Prueba 1 45
- 4.3. Resultados de la Prueba 2 51
- 4.4. Resultados Prueba 3 57
- 4.5. Resultados Prueba 4 63

CAPÍTULO 1

INTRODUCCIÓN

Durante los últimos años la identificación de rostros humanos ha sido el tema central en la mayoría de las investigaciones en las áreas de los sistemas de reconocimiento de patrones, el procesamiento de imágenes y la visión computacional [22], debido a las múltiples aplicaciones que pueden derivarse, tal como los sistemas de seguridad y vigilancia, el control de acceso o incluso en la detección de patologías. Existen múltiples algoritmos capaces de identificar y clasificar los rostros humanos y cuyo modelo generalmente puede ser definido por una serie de pasos o etapas. En la Figura 1.1 se ilustra un ejemplo, en el cual se consideran un cierto número de pasos o etapas. En este caso y después de adquirirse las imágenes, el primer paso en la construcción de un sistema de identificación es encontrar en cada imagen la región que abarque el área de interés a ser clasificada, para posteriormente poder aplicarse algún tipo de operación de pre-procesamiento y así corregir la posible existencia de variaciones en la captura o adquisición de las imágenes. El segundo paso consiste, en la extracción de características que en otras palabras es el proceso en el cual la información contenida en una imagen es transformada en una serie de características propias con las cuales puede ser distinguida del resto. Se ha demostrado, que la eficiencia en los sistemas de reconocimiento depende principalmente de esta etapa [3] en muchas investigaciones han sido propuestas diversas técnicas para poder extraer dichas características sin afectar el rendimiento de los sistemas, siendo una de las más populares la Transformada wavelet (TW) [2, 15, 4]. La TW es una herramienta matemática cuyas propiedades permiten analizar y manipular la información de una señal de forma que pueden ser obtenidos detalles que al ingresarlos a los sistemas de clasificación estos pueden volverse cada vez más robustos y eficientes. Por último la etapa de clasificación es una técnica probabilística que busca determinar a que grupo pertenece cada elemento de un conjunto de datos y se divide en un proceso de entrenamiento y uno de validación con los cuales pueden ser implementados los sistemas de identificación de personas.

Los métodos de identificación de rostros humanos pueden ser divididos en dos enfoques distintos, los holísticos y los locales. En el caso del enfoque holístico las características extraídas son usadas para modelar la relación que existe entre ellas y por lo tanto se requiere de un gran conjunto de datos ya que no es posible realizar una identificación con una sola muestra ya que se perdería la capacidad de generalización [22]. Por otra parte el enfoque local busca extraer las características en ciertos componentes del área de la cara como lo son los ojos, nariz o boca

en donde las ventajas incluyen la reducción en las dimensiones de los datos y fortalecen los sistemas en contra de variaciones en la iluminación y corrupción por ruido de las imágenes al momento de su adquisición. Sin embargo, la identificación basada solamente en el enfoque local destruye completamente la información holística y por lo tanto es de suma importancia el poder combinar ambos enfoques para así obtener los mejores resultados.

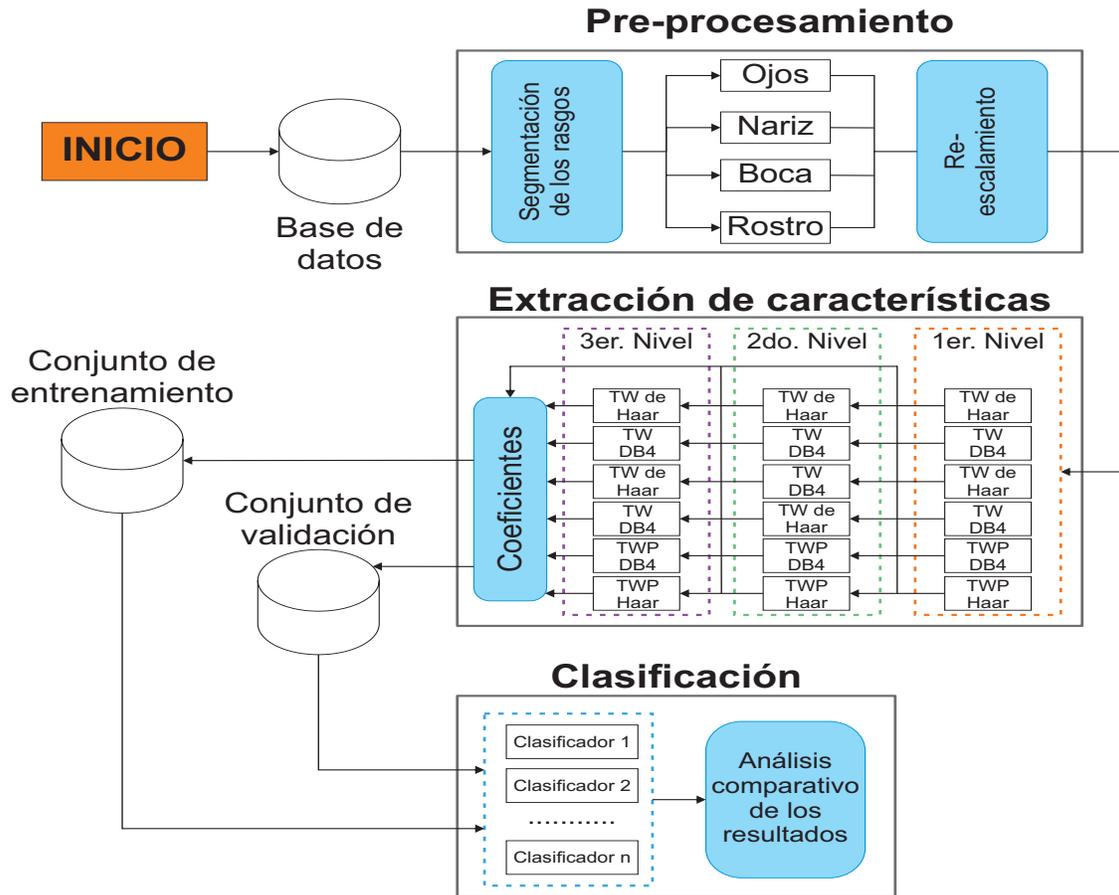


Figura 1.1: Diagrama del sistema de identificación de personas implementado y basado en la extracción de características por TW.

1.1 Objetivos

Debido a la gran importancia que tiene la extracción de características en los sistemas de clasificación de imágenes de rostros humanos el objetivo de este trabajo es la implementación y verificación de un sistema de identificación facial que busca ser robusto a las diferentes variaciones presentes en las imágenes de personas y específicamente para las áreas faciales. La técnica propuesta es basada en la teoría wavelet para la transformación de señales [1, 7, 11, 10] y consiste en tomar algunas de las más conocidas funciones wavelet y combinar las propiedades de ambas para así formar un sistema dinámico definido como “sistema híbrido” con el cual se busca no solo extraer ciertos aspectos que no pueden ser observados con otras técnicas sino

además reducir el tiempo total de cómputo en la ejecución de cualquier sistema [9]. De manera particular se plantea implementar un sistema completo para la identificación de personas y realizar un análisis comparativo entre los resultados para poder determinar la eficacia de utilizar las TW en su forma híbrida como extractor de características. Para llevar a cabo lo anterior se pretende:

1. Seleccionar un conjunto de imágenes de prueba que contengan rostros humanos e implementar un sistema que logre identificar y segmentar las áreas de los principales rasgos de un rostro (ojos, nariz y boca).
2. Diseñar un conjunto de técnicas que permitan mejorar la visualización de las imágenes adquiridas y que a su vez puedan facilitar las operaciones de clasificación al disminuir las variaciones.
3. Extraer las características de las imágenes y llevar a cabo el análisis comparativo. La extracción de características se llevará a cabo con la TW de manera discreta y en dos dimensiones en su forma no estándar, en donde se contemple usar diferentes funciones wavelet.
4. Implementar una serie de sistemas de clasificación para evaluar los datos extraídos y determinar los parámetros que otorguen el mayor porcentaje de éxito.

1.2 Organización

En el primer capítulo se ofrece la definición de los objetivos generales y particulares así como la justificación de este trabajo de tesis. Dentro del capítulo 2 se encuentran los principios y procedimientos para realizar la detección y pre-procesamiento de rostros como primera etapa en los sistemas de identificación de personas. El capítulo 3 contiene los preliminares de la teoría wavelet y una explicación para todas y cada una de las funciones utilizadas en los procedimientos de extracción de características, así como los algoritmos necesarios para su implementación en una y dos dimensiones de manera discreta y no estándar. En el capítulo 4 se expone la teoría necesaria para entender los sistemas de clasificación implementados y los resultados obtenidos para cada uno son utilizados para realizar un análisis comparativo y lograr establecer el éxito de la experimentación. Por último el capítulo 5 aborda las conclusiones generales a las que se llegaron al finalizar la investigación, los objetivos alcanzados así como los problemas encontrados y finalizando con el trabajo futuro dentro de esta misma línea de investigación.

CAPÍTULO 2

DETECCIÓN Y PRE-PROCESADO DE IMÁGENES FACIALES

El reconocimiento de rostros humanos es un tema de investigación que ha recibido una gran atención en las últimas décadas. Los sistemas de reconocimiento de imágenes han sufrido una gran expansión en su estudio e investigación durante los últimos años debido a que han sido incluidos en muchos campos de visión computacional, control de acceso, seguridad y vigilancia, entre otros. Especialmente los sistemas para el reconocimiento de imágenes faciales, se encuentran en una gran cantidad de aplicaciones tales como los sistemas inteligentes de análisis, la compresión de información e incluso en la detección de patologías.

Los algoritmos para la identificación de rostros en conjunto con una apropiada etapa de pre-procesamiento, garantizan mejores resultados en la clasificación de imágenes de esta naturaleza, debido a que en la adquisición de dichas imágenes se encuentran contenidas un gran número de variaciones, que pueden obstaculizar el funcionamiento del sistema. Es por esto, que en este capítulo se describen los métodos utilizados, para tratar de tener las imágenes lo más homogéneo posible, para que se puedan utilizar en las etapas posteriores del reconocimiento de imágenes faciales.

2.1 Detección de rostros

En términos del procesamiento y análisis de imágenes, una imagen puede ser representada de forma digital como una señal discreta en dos dimensiones, donde los valores de la señal están contenidos en arreglos y a dichos elementos se les denominan “píxeles” [14]. El valor de cada pixel corresponde a la irradiancia proveniente de las señales bidimensionales adquiridas y dichos valores pueden presentar múltiples variaciones en cada una de las diferentes imágenes, lo cual representa un problema de gran importancia dentro de los sistemas de reconocimiento, debido a que una misma persona puede tener considerables diferencias en cada una de las adquisiciones lo cual puede obstaculizar el proceso de discriminación. Lo anterior puede ser categorizado por dos grupos de variaciones dentro las imágenes de rostros humanos: los factores intrínsecos y los extrínsecos [16]. Los factores intrínsecos son aquellos que se encuentran en la naturaleza de cada una de las personas y que no depende del observador o de la adquisición de la imagen, y para estas variaciones se consideran dos clases, la intrapersonal y la interpersonal.

Los factores intrapersonales son debidos a las variaciones en la apariencia de cada persona, tales como el color y estilo de cabello o si la persona hace uso de anteojos o cosméticos así como la edad del sujeto. Mientras que los factores interpersonales son aquellos causados por el sexo o raza al que pertenece la persona. Finalmente los factores extrínsecos son los que dependen completamente de la adquisición de la imagen, tales como las de las condiciones de iluminación, escala, resolución, ruido o enfoque con el que fueron tomadas las imágenes [16, 18, 19]. Debido a lo anterior es necesario seleccionar un conjunto de muestras adecuado para poder llevar a cabo el reconocimiento, tal es el caso de la base de imágenes *BioID* [20] de la cual fueron seleccionadas un total de 208 imágenes con 22 sujetos diferentes. Dichas imágenes cumplen con los factores extrínsecos necesarios para su adecuado manejo; no obstante fueron tomados en cuenta factores intrínsecos de diferente naturaleza con la intención de poseer un amplio espectro en los resultados de la detección y clasificación. El tamaño de las imágenes contenidas en la base *BioID* es de 384 pixeles de ancho por 286 pixeles de alto y son representadas en una escala de grises (monocromáticas), lo cual quiere decir que la representación de los valores de los pixeles es en forma entera y en un rango de 0 para representar el color blanco y hasta 255 el cual representa el color negro [17].

2.1.1 Algoritmo Viola-Jones

Una vez seleccionadas las muestras es posible detectar objetos específicos dentro de estas, tal es el caso de los rasgos en un rostro humano. Existen diferentes formas de diferenciar un rostro de cualquier otro y principalmente es debido a los factores intrínsecos, por eso es de vital importancia poder detectarlos para así facilitar el proceso de identificación. Generalmente los ojos, nariz y boca de una persona brindan una gran cantidad de información para su discriminación, es por esto que han sido implementados diversos algoritmos capaces de identificar estos y algunos otros rasgos en imágenes que contienen rostros humanos. El algoritmo propuesto por Paul Viola y Michael Jones también conocido como el algoritmo Viola-Jones [21] resulta ser uno de los mas poderos para realizar esta clase de tareas, puesto que tiene un coste computacional muy bajo y garantiza una discriminación rápida de los elementos, además de tener una probabilidad de verdaderos positivos del 99.9 % y una probabilidad de falsos positivos del 3.33 %. Por lo anterior, se puede decir que el algoritmo Viola-Jones representa la mejor opción al momento de detectar rostros o rasgos y dado que el método procesa información en imágenes en escala de grises es totalmente adaptable a la base de datos *BioID*. El proceso de detección planteado por Viola-Jones puede ser dividido en los siguientes cuatro pasos:

1. Extracción de características.
2. Generación de imágenes integrales.
3. Implementación de clasificadores AdaBoost.
4. Implementación de clasificadores en cascada.

En la detección de objetos, las imágenes son clasificadas con base al valor de las características simples, esto quiere decir que cada pixel de la imagen cuenta con sus propias características y puede resultar en una rápida ejecución de los algoritmos. Por lo que el primer paso en el algoritmo Viola-Jones es la extracción de características por medio de las funciones wavelet de Haar, las cuales son una secuencia de funciones re-escaladas y de forma cuadrada propuestas por Alfred Haar en 1909, cuyo objetivo es lograr la resolución espacial necesaria para la detección y aumentar el poder expresivo del modelo a través de los descriptores rectangulares, los cuales son una extensión de la función wavelet de Haar de dos dimensiones (Figura 2.1) los detalles respecto a este tema pueden ser estudiados en [23].

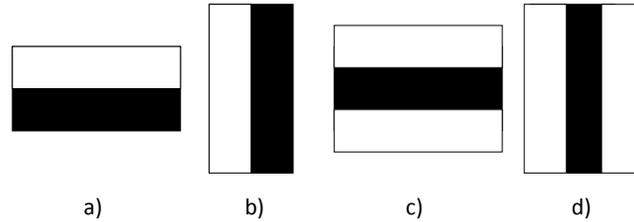


Figura 2.1: Descriptores rectangulares de Haar para bordes (a) horizontales y (b) verticales y para líneas (c) horizontales y (d) verticales.

Al extraer las características de las imágenes los descriptores rectangulares permiten detectar los diferentes cambios de intensidad en los pixeles y así detectar los bordes o líneas de los rasgos buscados. No obstante, el problema radica en los diferentes tamaños de kernel de las funciones de Haar y su localización ya que esto representa una gran cantidad de operaciones (aproximadamente 180,000 descriptores rectangulares en sub-imágenes de 24x24 pixeles) y por lo tanto un tiempo de ejecución muy grande.

Los descriptores rectangulares pueden ser procesados de manera más eficiente a través del segundo paso en el algoritmo Viola-Jones, el cual consiste en realizar una representación intermedia de la imagen, de ahí que se le denomine como proceso de imagen integral. En una imagen integral los puntos (x, y) contienen la suma de los pixeles que se encuentran por encima e izquierda de ellos tal que,

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'), \quad (2.1)$$

donde $ii(x, y)$ e $i(x, y)$ representan la imagen integral e imagen original, respectivamente. En la Figura 2.2 es representada la idea básica de una imagen integral, en donde el punto 1 es la suma de los pixeles en el rectángulo A, el valor de 2 la suma de A+B, el punto 3 es igual a A+C y finalmente el punto 4 es la suma de A+B+C+D.

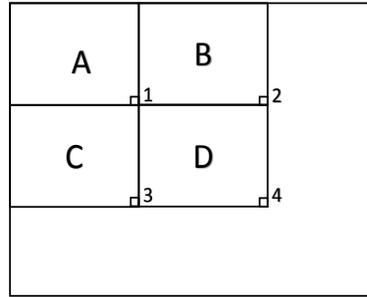


Figura 2.2: Ilustración de una imagen integral.

Con esto se logra minimizar las operaciones necesarias para computar las características de Haar, ya que todos los valores necesarios son calculados previamente por la imagen integral, además se garantiza una amplia eficiencia en la detección de objetos. La conjunción de los dos primeros pasos permite la detección de algunos de los diferentes rasgos contenidos en un rostro a través de las máquinas de aprendizaje. El algoritmo de aprendizaje AdaBoost utiliza un conjunto de características y un conjunto de entrenamiento de imágenes positivas y negativas, AdaBoost es utilizado para poder impulsar el rendimiento de la clasificación de un algoritmo de aprendizaje simple o débil [21], esto se logra al tomar un muy pequeño número de características y combinándolas para formar un clasificador fuerte y efectivo. El proceso consiste en tomar un algoritmo de aprendizaje débil para seleccionar el descriptor rectangular que tenga la mejor separación entre ejemplos de imágenes positivas y negativas. Por cada característica, el algoritmo débil determina el umbral óptimo de la función de clasificación que minimice el número de malas clasificaciones. De esta manera un clasificador débil $h_j(x)$ contiene una característica f_j , un umbral θ_j y una paridad p_j tal que

$$h_j(x) = \begin{cases} 1 & \text{si } p_j f_j(x) < p_j \theta_j, \\ 0 & \text{si no,} \end{cases} \quad (2.2)$$

donde x es una sub-ventana de 24×24 píxeles de una imagen. Por lo tanto el algoritmo AdaBoost obtiene el peso de los clasificadores simples antes descritos y asigna un peso específico a cada clasificación de manera que sea favorecida la detección de rasgos dándoles mayor peso a esto. El peso de un ejemplo clasificado correctamente es modificado en cada clasificación, mientras que el peso asignado a uno mal clasificado permanece constante, de esta forma el clasificador con menor error es seleccionado y así es formado un fuerte clasificador final.

Finalmente y como una forma de aumentar el desempeño en la detección de objetos es implementado un algoritmo de clasificación en cascada con el cual además se obtiene una notable disminución en el tiempo de ejecución, reduciendo así el costo computacional. En la Figura 2.3 es ilustrado el proceso de una clasificación en cascada la cual esta basada en los clasificadores AdaBoost, donde el resultado del primer clasificador en el caso de ser positivo (detección de rostro) pasa a ser evaluado por un segundo clasificador AdaBoost y así sucesivamente. Mientras que para el caso en el que la salida sea negativa (sin detección de rostros) la sub-ventana es descartada. En cada etapa de la clasificación en cascada la

complejidad es aumentada y el tiempo de ejecución crece con ella, por lo que son definidos 3 aspectos de optimización los cuales son:

1. El número de etapas de clasificación.
2. El número de características por etapa.
3. El umbral con el cual será determinada la complejidad en cada etapa.

Dichos aspectos son adaptados para el entrenamiento de las características de acuerdo con los rasgos a detectar.

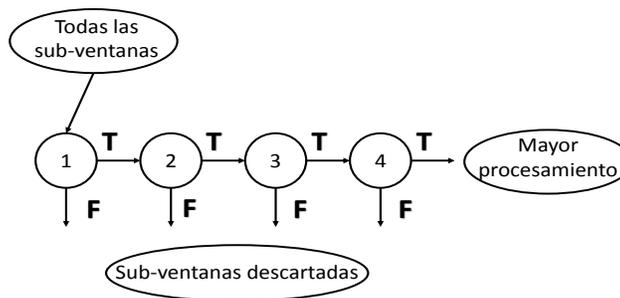


Figura 2.3: Esquema de un clasificador en cascada.

2.2 Pre-procesado de imágenes

El pre-procesamiento de una imagen comprende un conjunto de técnicas cuyo fin es el de mejorar aspectos o características que faciliten las operaciones del procesado sobre la imagen. Lo anterior brinda una solución a los factores extrínsecos que pueden afectar procesos posteriores, ejemplos de esto es la reducción del área de trabajo, eliminación de ruido, realce de contraste o el re-escalamiento de la imagen. El resultado de aplicar el algoritmo Viola-Jones a la base de imágenes *BioID* genera un conjunto de áreas dentro de los rostros (Figura 2.4), las cuales contienen los tres rasgos más importantes del rostro planteados al principio de este capítulo (ojos, nariz y boca), dicho algoritmo facilita la segmentación de estas áreas.



Figura 2.4: Detección de (a) rostro y (b) ojos, nariz y boca a través del algoritmo Viola-Jones.

No obstante debido a los factores extrínsecos de escala [19], el conjunto de sub-imágenes generadas a partir de la detección de los rasgos resultan ser de tamaños muy distintos. Por lo que es llevado a cabo un pre-procesamiento a las sub-imágenes con el objetivo de generar una base de imágenes consistente en sus dimensiones, esto quiere decir que todos los elementos cuenten con las mismas características de tamaño y forma. Este proceso consiste en el re-escalamiento del área de los ojos, nariz y boca, en donde el primer paso es la segmentación de las tres áreas a partir de tomar el tamaño del rostro y el rasgo detectado que va a segmentarse en cada una de las imágenes, para así formar el área de la nueva sub-imagen (Figura 2.5).

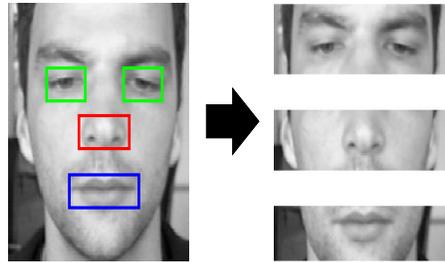


Figura 2.5: Segmentación del área de los ojos, nariz y boca.

Una vez obtenidos los nuevos conjuntos de imágenes, es definido un tamaño estándar de imagen a partir de tomar el mayor y menor tamaño de área y calculando la media existente entre estas dos. El re-escalamiento de las imágenes es llevado a cabo a través de la interpolación, que es un proceso matemático que calcula los datos necesarios, para formar una imagen de mayor o menor tamaño a partir de los datos originales, esto nuevos datos son los valores de los pixeles que representaran de la mejor manera a la original. La interpolación bicúbica de vecindario 4×4 es una de las técnicas con mejores resultados para el re-dimensionamiento de las imágenes aunque supone ser un poco mas costosa en términos de costo computacional, ya que toma la información de los 16 vecinos en cada punto y los promedia para tomar la mejor aproximación, basándonos en los resultados de este método y no en el costo computacional se seleccionó para realizar la etapa de re-escalamiento. Es así que el conjunto de sub-imágenes resultantes (Figura 2.6) cumple con los parámetros necesarios para poder llevar a cabo las operaciones de reconocimiento de imágenes, garantizando una consistencia en las características contenidas en cada una de ellas. El tamaño promedio para las imágenes contenidas en la base *BioID* resulta ser de 168×48 pixeles para cada uno de los rasgos detectados, por lo que con el re-escalamiento las imágenes son adaptadas a dichas dimensiones.



Figura 2.6: Ejemplo de interpolación bicúbica de vecindario 4×4 aplicada al área de los ojos.

CAPÍTULO 3

EXTRACCIÓN DE CARACTERÍSTICAS POR TRANSFORMADAS WAVELET

La aplicación de la TW en ciencias e ingeniería comenzó a despegar a principios de la década de 1990, aunque sus primeros estudios datan de 1910, con un rápido crecimiento en el número de investigadores que centraron su atención en el análisis wavelet durante esa década. En la actualidad, se puede constatar en la literatura que la TW ha tomado un papel importante en gran cantidad de aplicaciones que cubren diferentes disciplinas. Se puede decir que la TW es una herramienta matemática que ahora es parte indispensable en las cajas de herramientas de muchos analistas de datos [1].

En este capítulo se estipulan los principios básicos de la teoría wavelet para la descomposición de señales en una y dos dimensiones. En particular se considera la TW bidimensional en su forma no estándar, utilizando dos de las más conocidas funciones wavelet, la de Haar y la de Daubechies de cuatro coeficientes.

3.1 Preliminares

En un sistema de tiempo discreto se considera que una transformación, es aquel operador que toma una secuencia $\{x(n)\}$, llamada de excitación, dando como resultado la secuencia $\{y(n)\}$, llamada respuesta [6]. En ese sentido, la TW es usada para transformar señales y así poder manipularlas y analizarlas. Lo anterior se puede interpretar como la convolución de una señal con un conjunto de funciones wavelet. No se debe olvidar que el analizar y procesar señales en distintos dominios resulta beneficioso al obtener, en algunos casos, detalles que no pueden ser observados con técnicas convencionales como lo son los clásicos métodos de Fourier. Sin embargo, se hace hincapié en que la TW es una “alternativa” a la Transformada de Fourier más que una sustitución [5].

Los espacios a considerar en este trabajo son los llamados espacios de Hilbert, cuyas ventajas de considerarlos se explican en [5]. En particular, se consideran los espacios enteros (denotados como \mathbb{Z}^n) de secuencias cuadradas sumables que se denotan como $l_2(\mathbb{Z})$ y/o los

espacios reales (denotados como \mathbb{R}^n) de funciones cuadradas integrables que se denotan como $L_2(\mathbb{R})$, las cuales cumplen con $\sum |x[n]|^2 < \infty$ y $\int |x(t)|^2 dt < \infty$, respectivamente. En el procesamiento de señales en tiempo discreto dada una secuencia $x[n]$ la cual representa un vector en el espacio de Hilbert $l_2(\mathbb{Z})$ tiene un producto interno tal que

$$\langle x, y \rangle = \sum_{n=-\infty}^{\infty} x[n]^* y[n], \quad (3.1)$$

y su norma es

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{n \in \mathbb{Z}} |x[n]|^2}. \quad (3.2)$$

Mientras que el producto interno para el caso en tiempo continuo dada una función $f(t)$ definida en el espacio de Hilbert $L_2(\mathbb{R})$ es dado como

$$\langle f, g \rangle = \int_{t \in \mathbb{R}} f(t)^* g(t) dt, \quad (3.3)$$

con norma igual a

$$\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_{t \in \mathbb{R}} |f[t]|^2}. \quad (3.4)$$

Dada cualquier señal $f(x)$ de algún espacio E de dimensión finita es posible encontrar un conjunto de vectores $S = \{\varphi_k\}_{k \in \mathbb{Z}}$ tal que $f(x)$ pueda representar el espacio como una combinación de lineal

$$f(x) = \sum_k a_k \varphi_k, \quad (3.5)$$

la cual es llamada una expansión o serie para $f(x)$, donde a $\{\varphi_k\}$ se le conoce como **base vectorial ortonormal**.

3.1.1 Expansión en series de señales

En términos del procesamiento de señales en tiempo discreto una expansión en series puede ser definida de diversas maneras, por ejemplo una de las expansiones más simples es la llamada expansión de Haar.

Una expansión ortonormal de una secuencia arbitraria cuadrada sumable $x[n]$, esto es $x[n] \in l_2(\mathbb{Z})$, es representada como

$$x[n] = \sum_{k \in \mathbb{Z}} \langle \varphi_k[l], x[l] \rangle \varphi_k[n] = \sum_{k \in \mathbb{Z}} X[k] \varphi_k[n], \quad (3.6)$$

donde

$$X[k] = \langle \varphi_k[l], x[l] \rangle = \sum_l \varphi_k^*[l] x[l], \quad (3.7)$$

es la transformada de $x[n]$. El símbolo “*” denota el complejo conjugado. Las funciones base φ_k satisfacen las condiciones de ortonormalidad y son funciones completas esto quiere decir que el producto interno es tal que

$$\langle \varphi_k[n], \varphi_l[n] \rangle = \delta[k - l], \quad (3.8)$$

donde $\delta[k]$ es la función delta o función impulso la cual es igual a 1 si $k = 0$ y cero en cualquier otro caso, entonces cualquier señal en $l_2(\mathbb{Z})$ puede ser representada como en (3.6). Una importante propiedad en las expansiones ortonormales es la de la conservación de la energía,

$$\|x\|^2 = \|X\|^2. \quad (3.9)$$

3.1.2 Transformada wavelet

Mientras en el análisis de Fourier una señal es representada por medio de funciones sinusoidales, en el análisis wavelet el conjunto base no es único y está formado por funciones llamadas wavelets, las cuales son pulsos bien localizados en el dominio del tiempo lo que les permite proporcionar la información espectral correspondiente a posiciones temporales diferentes de una señal [9]. Como ha sido mencionado la TW es un método con el cual es posible convertir una señal en otra haciendo que ciertas características de la señal original sean más fáciles de estudiar, para calcular una *Transformada wavelet* es necesario seleccionar una función wavelet base, llamada en algunas ocasiones función “wavelet madre” $\psi(t)$, donde $\psi(t) \in L^2(\mathbb{R})$ y cumple con

$$\int_{-\infty}^{\infty} \psi(t) dt = 0, \quad (3.10)$$

estas funciones están dadas por escalamientos y traslaciones en el dominio del tiempo. Estos factores modifican las funciones wavelet madres y dan origen a versiones trasladadas y escaladas de la misma, también llamadas funciones “wavelet hijas” y que son de la forma

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \quad (3.11)$$

donde a y b son el factor de escala y de traslación, respectivamente. El factor \sqrt{a} es una constante de normalización considerada para preservar la energía [1]. Algunas de las funciones más utilizadas son la de Haar y la familia de funciones Daubechies nombradas así en honor a sus descubridores y cuyas gráficas se muestran en la Figura 3.1 las cuales representan funciones wavelet madres en tiempo continuo.

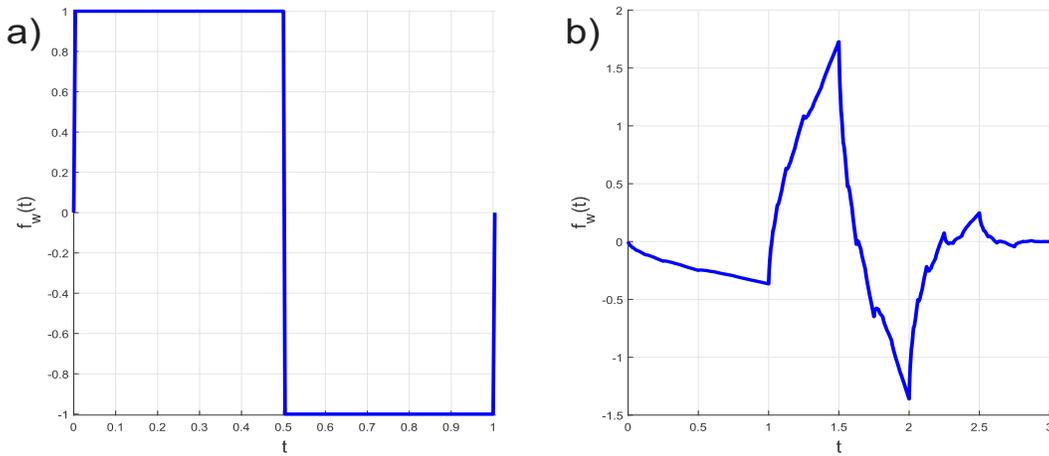


Figura 3.1: Funciones wavelet madre de (a) Haar y (b) Daubechies de cuatro coeficientes.

3.1.3 Transformada wavelet discreta

Las imágenes son interpretadas como señales bidimensionales en tiempo discreto, es por esto que es preciso considerar la TW de carácter bidimensional para señales de energía finita como lo son estas. La Transformada wavelet discreta (TWD) es una herramienta poderosa al ser implementada como un algoritmo rápido de descomposición. Como ha sido mostrado, una función wavelet en el espacio $L^2(\mathbb{R})$ es representada como una función continua y por definición, dada una función $x(t)$ la Transformada wavelet Continua (TWC) es de la forma

$$TWC(a, b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)dt, \quad (3.12)$$

donde el mapeo va de una señal continua unidimensional a una función de dos variables reales continuas (a y b). Para obtener la TWD es necesario en primer lugar, llevar a cabo la discretización de las señales analógicas y una de las maneras más sencillas para realizar este proceso es registrar su valor de muestreo $x(nT)_{n \in \mathbb{Z}}$ para un intervalo de tiempo T . Lo anterior resulta en una aproximación de la señal $x(t)$ en forma discreta. Comúnmente el muestreo de la TWC es diádico lo cual garantiza la eficiencia en la implementación de los parámetros de escala y traslación, derivando en que

$$a = 2^{-m} \text{ y } b = n2^{-m} \text{ donde } m, n \in \mathbb{Z}. \quad (3.13)$$

Estos parámetros, al ser sustituidos en (3.11) resulta en una familia de wavelets de la forma diádica tal que

$$\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n). \quad (3.14)$$

La ecuación (3.14) puede ser sustituida en (3.12) para obtener los coeficientes de la serie wavelet para la función continua $x(t)$, teniendo en cuenta que los índices de dilatación y traslación son expresados como m y n respectivamente, derivando en la siguiente expresión

$$d_{m,n} = TWC(2^{-m}, n2^{-m}) = \int_{-\infty}^{\infty} x(t)\psi_{m,n}(t)dt. \quad (3.15)$$

Dentro de este marco de trabajo, donde el conjunto de las funciones wavelet $\psi_{m,n}(t)$ constituyan una base ortonormal, una señal de energía finita se puede expresar como

$$x(t) = \sum_m \sum_n d_{m,n}\psi_{m,n}(t). \quad (3.16)$$

La implementación numérica de la TWD una vez que se definió la expansión para señales en tiempo continuo puede ser realizada a través del análisis multi-resolución (Sección 3.1.4)[11]. Mediante este método, a la función wavelet $\psi(t)$ se le asocia una función de escala $\varphi(t)$, la cual al aplicar las operaciones de traslación y escalamiento se tiene

$$\varphi_{m,n}(t) = 2^{m/2}\varphi(2^m t - n), \quad (3.17)$$

Es así que la expansión de $x(t)$, considerando que (3.14) y (3.17) forman una base ortonormal, es

$$x(t) = \sum_n \left(a_{m0,n}\varphi_{m0,n}(t) + \sum_{m=m0}^{M-1} d_{m,n}\psi_{m,n}(t) \right), \quad (3.18)$$

donde $a_{m,n}$ son los coeficientes de escala y $d_{m,n}$ los coeficientes wavelets necesarios para llevar a cabo la transformación de una señal $x(t)$. Dichos coeficientes se calculan mediante el producto punto de la función $x(t)$ con las funciones de escala y wavelet, respectivamente, de la siguiente forma

$$a_{m,n} = \sum_t x(t)\varphi_{m,n}(t), \quad (3.19a)$$

$$d_{m,n} = \sum_t x(t)\psi_{m,n}(t). \quad (3.19b)$$

Una de las propiedades más importantes de las transformaciones wavelet es la de la **conservación de la energía**, donde para la versión discreta $x[n] = x(t)|_{t=nT}$ la energía es expresada como la suma de los cuadrados de sus valores. Lo anterior puede ser expresado como

$$E_x = \sum_{j=1}^N x[j]^2, \quad (3.20)$$

donde N es la longitud de $x[n]$. Es entonces que la conservación de la energía en las transformadas wavelet es justificada gracias a que de acuerdo con la Ecuación 3.20 la energía contenida en los coeficientes $a_{m,n}$ y $d_{m,n}$ que comprenden la TWD de $x(t)$ puede ser calculada por

$$E_{a_{m,n}|b_{m,n}} = \sum_{i=1}^N (W_i)^2, \quad (3.21)$$

donde W_i representa un vector que contiene los coeficientes $a_{m,n}$ y $d_{m,n}$, dando como resultado que $E_{(a|d)}$ es igual a E_x , la demostración de estas afirmaciones es mostrada en [12, 1]. Por último es importante recalcar la redistribución de la energía de una señal al ser procesada por una TW, esto quiere decir que la energía contenida en una señal $x[n]$ no es distribuida uniformemente a través de las funciones de escala y wavelet, sino redistribuida tal que la energía contenida en las sub-señales resultantes de las funciones de escala contienen el mayor porcentaje de la energía total, a este fenómeno se le conoce como “compactación de la energía”. Mientras que el resto de la energía es distribuida por las sub-señales resultantes de las funciones wavelet, no obstante, estos valores pueden resaltar características de suma importancia (Sección 4.2.1 y 4.2.2).

3.1.4 Análisis Multi-resolución

El objetivo de aplicar el análisis multi-resolución a una función es el de representar esta mediante una aproximación promedio a través de proyecciones ortogonales con una resolución 2^{-j} en un espacio $V_m \subset L^2(\mathbb{R})$ y de una versión de detalles (la diferencia entre la señal original y su aproximación promedio). El objetivo principal de modificar la resolución en una señal es permitir procesar los detalles relevantes para una tarea en particular [11]; este proceso se aplica recursivamente sobre la versión promedio generando una colección de detalles más un paquete de versiones promedio de menor resolución, tal y como se ilustra en la Figura 3.2.

Formalmente una secuencia $\{V_m\}_{m \in \mathbb{Z}}$ de sub-espacios cerrados en $L^2(\mathbb{R})$ debe cumplir ciertas propiedades definidas por primera vez por Mallat y Meyer [11] las cuales son:

1. $\bigcup_{m \in \mathbb{Z}} V_m$ es denso en $L^2(\mathbb{R})$.
2. $\bigcap_{m \in \mathbb{Z}} V_m = 0$.
3. Invariabilidad en escala: para cada $m \in \mathbb{Z}$, $x(t) \in V_m$ es equivalente a $x(2t) \in V_{m+1}$.
4. Invariabilidad bajo corrimiento: para cada $x(t) \in V_0$ y para cada $n \in \mathbb{Z}$, $x(t - n) \in V_0$.
5. Existencia de una base $\{\varphi(t - n)\}_{n \in \mathbb{Z}}$ es una base ortonormal para V_0 , donde φ es llamada función de escala.

Dentro del análisis wavelet una señal $x(t) \in V_{m+1}$, donde $V_{m+1} \in L^2(\mathbb{R})$ puede generar una base para los espacios V_m que represente el promedio de la señal a través de la función de escala (3.17) y los espacios W_m pueden ser generados por la función wavelet (3.14) para representar los detalles de la señal. En otras palabras, es posible descomponer una señal por medio de un algoritmo piramidal (Figura 3.2), que desde el punto de vista de los bancos de filtros corresponde a un arreglo de filtros espejo en cada nivel de resolución, donde el primero de ellos es asociado a un filtro pasa bajas que proporciona las aproximaciones y el segundo a un filtro pasa altas para el caso de los detalles como se muestra en la Figura 3.3a. El primer nivel de descomposición de una función genera aproximaciones suavizadas y detalles de menor resolución y sucesivamente en cada nivel de descomposición. Dichos filtros representan a la función de escala y la función wavelet, respectivamente y que la aplicación iterativa producen la división del ancho de banda de la señal. En la Figura 3.3b se muestra gráficamente el comportamiento del espectro en frecuencia de la señal.

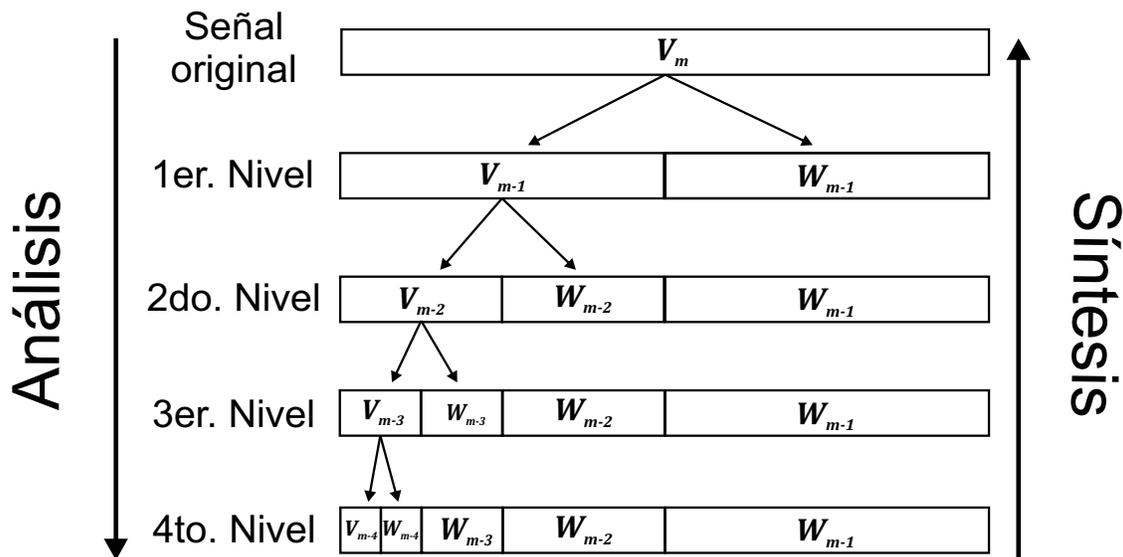


Figura 3.2: Algoritmo piramidal de descomposición.

La recursividad en la TW es posible gracias a que las ecuaciones (3.22a) y (3.22b), las cuales ilustran la relación entre la expansión de los coeficientes de un nivel menor en términos de otros de mayor escala y viceversa [7],

$$a_m[k] = \sum_n h[n - 2k]a_{m+1}[n], \quad (3.22a)$$

$$d_m[k] = \sum_n g[n - 2k]a_{m+1}[n]. \quad (3.22b)$$

La expresión (3.22) representa el cálculo de los coeficientes de escala y wavelet, respectivamente, de la TW en una escala m , donde $h[k]$ es un filtro pasa bajas y $g[k]$ un filtro pasa altas del banco de filtros, que al aplicar la convolución con los coeficientes de escala en la próxima escala más fina $m+1$ seguida por una reducción mediante un proceso de sub-muestreo o downsampling a través de operadores llamados *diezmadores* en un factor de 2, representan un **banco de análisis** (Figura 3.3c).

La TW es reversible, es decir que la señal una vez que fue descompuesta en sus aproximaciones y detalles, puede ser reconstruida de manera conveniente gracias a que al combinar los filtros escalables y wavelet de manera inversa al orden que fueron aplicados genera la señal original. Esto es debido a que dichos filtros son derivados de bases ortonormales y por consecuencia se traduce en una suma de dos convoluciones como se muestra a continuación:

$$a_{m+1}[k] = \sum_n (a_m[n]h[k - 2n] + d_m[n]g[k - 2n]). \quad (3.23)$$

En la Figura 3.3c se ilustra el proceso de reconstrucción, cuyo primer paso es aplicar un sobremuestreo o upsampling a los coeficientes sub-banda generados por el banco de análisis a través de operadores llamados *interpoladores* en un factor de 2, seguido de la suma de convoluciones como en (3.23). A esto se le conoce como **banco de síntesis**.

3.2 Transformada de Haar

En cuestión de wavelets ortonormales, existen diversas funciones con sus propias características así como una representación para el caso de tiempo continuo y discreto. La wavelet de Haar siempre ha resultado ser el principal ejemplo dada su sencillez y fácil implementación; es la más recurrida por los analistas de datos y está presente en la mayoría de estudios, por lo que suele ser el punto de partida en cualquier análisis de señales. Ya que su ecuación de escala resulta ser la más simple de las familias wavelet existentes, en esta solo dos coeficientes de escala son distintos de cero. Gráficamente en la Figura 3.4a se puede observar la solución a la ecuación de escalamiento de Haar, donde los coeficientes son definidos como un pulso cuadrado derivando en que

$$\varphi(x) = \begin{cases} 1 & \text{si } 0 \leq x < 1, \\ 0 & \text{si no.} \end{cases} \quad (3.24)$$

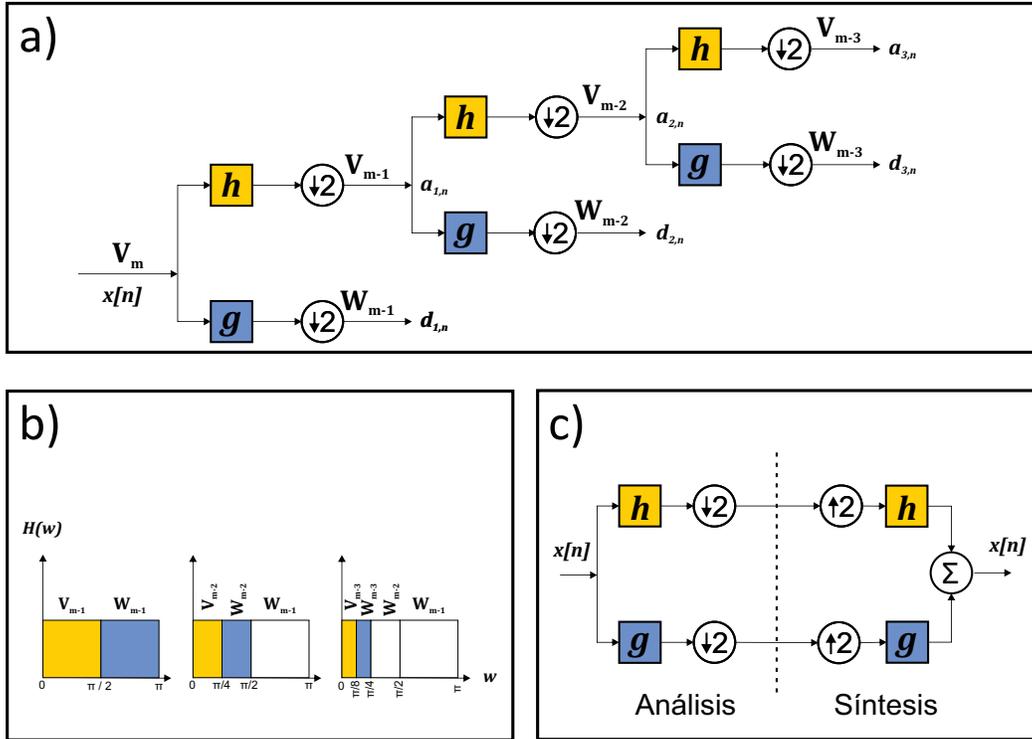


Figura 3.3: (a) Análisis mediante Bancos de Filtros, (b) descomposición del espectro en frecuencia de la señal al ser procesada en tres niveles de transformación y (c) BF para el análisis y síntesis respectivamente de una señal.

Así que el filtro escalable $\varphi(x)$ se puede expresar como una ecuación bi-escala en tiempo discreto y dicha función puede escribirse como una combinación lineal de funciones base en un espacio $V_m \in l_2(\mathbb{R})$ [7], teniendo que

$$\varphi(x) = \sqrt{2} \sum_{k=-\infty}^{\infty} a[k] \varphi(2x - k), k \in \mathbb{Z}, \quad (3.25)$$

donde los coeficientes $a[k]$ son dados por

$$a[k] = \sqrt{2} \int_{-\infty}^{\infty} \varphi(x) \varphi^*(2x - k) dx, \quad (3.26)$$

de esta manera el filtro escalable $\varphi(x)$ puede ser expresado como

$$\varphi(x) = \varphi(2x) + \varphi(2x - 1). \quad (3.27)$$

De acuerdo con [10] el filtro wavelet de Haar resulta de la construcción de la función wavelet expresada como

$$\psi(x) = \sqrt{2} \sum_k b[k] \varphi(2x - k), \quad (3.28)$$

con coeficientes iguales a

$$b[k] := (-1)^k a^*[1 - k] \quad (3.29)$$

y que puede escribirse como

$$\psi(x) = \varphi(2x) - \varphi(2x - 1). \quad (3.30)$$

En la Figura 3.4b se muestra el resultado de esta función y los coeficientes de la wavelet de Haar se pueden escribir de manera equivalente como

$$\psi(x) = \begin{cases} 1 & \text{si } 0 \leq x < \frac{1}{2}, \\ -1 & \text{si } \frac{1}{2} \leq x < 1, \\ 0 & \text{si no.} \end{cases} \quad (3.31)$$

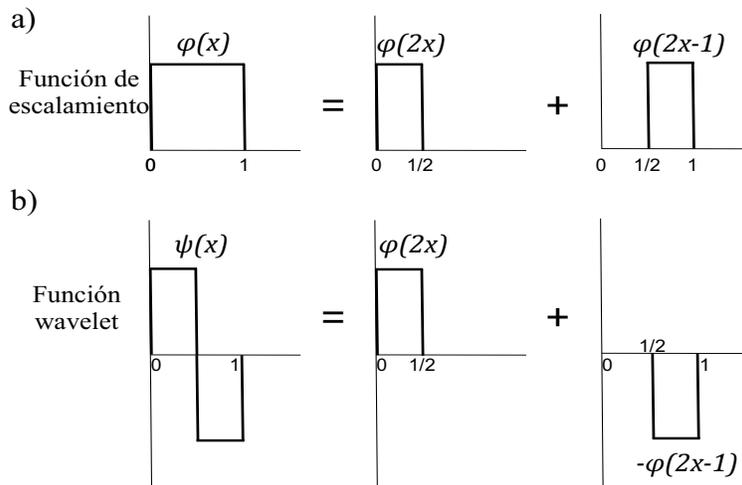


Figura 3.4: Funciones de (a) escala y (b) wavelet de Haar, respectivamente.

Para el caso de señales en tiempo discreto la TW puede ser interpretada desde el punto de vista de los Bancos de Filtros, los cuales son filtros lineales constituidos usualmente por un filtro pasa-bajas (**h**) y un filtro pasa-altas (**g**). Aplicando el algoritmo piramidal de descomposición se obtiene el banco de análisis al aplicar un filtraje seguido de las operaciones de diezmación con factor de 2. Esto quiere decir que una señal $a_0[k] = X[k]$ es procesada a través de (3.32) para un primer nivel de descomposición, donde $a_{-1}[k]$ es la señal promedio resultante y $d_{-1}[k]$ son los detalles.

$$a_{-1}[k] = \sum_n h[n - 2k]a_0[n] = \frac{1}{\sqrt{2}}(a_0[2k] + a_0[2k + 1]), \quad (3.32a)$$

$$d_{-1}[k] = \sum_n g[n - 2k]a_0[n] = \frac{1}{\sqrt{2}}(a_0[2k] - a_0[2k + 1]). \quad (3.32b)$$

Finalmente el banco de síntesis, que representa la reconstrucción de una señal que ha sido descompuesta por (3.32), es implementada como el proceso inverso de filtraje con la diferencia de que los operadores de diezmación son cambiados por los de interpolación. Por lo tanto

$$X[k] = \sum_{n \in Z} [a_{-1}[n]h[k - 2n] + d_{-1}[n]g[k - 2n]]. \quad (3.33)$$

3.3 Transformada Daubechies

La Transformada wavelet Daubechies, nombrada así en honor a Ingrid Daubechies [8], al igual que la TW de Haar es calculada a través del producto escalar entre una señal y las funciones de escala y wavelet. Estas transformaciones tienen un soporte ligeramente más largo que para el caso de la función wavelet de Haar, aunque el número de coeficientes de escala N_k sigue siendo de carácter finito. Sin embargo, este pequeño cambio produce una gran mejora al obtener los promedios y detalles utilizando solo unos cuantos valores de la señal. La wavelet de Haar se encuentra dentro de la familia de wavelets Daubechies como el miembro más simple de esta, debido a que los coeficientes de Haar forman el soporte más sencillo que pueden tener las funciones wavelet de Daubechies. Esto quiere decir que todas las funciones dentro de la familia cumplen con las condiciones de ortonormalidad de la Sección 3.1 con un número de coeficientes finitos [1]. El suavizado causado por la wavelet de Daubechies es comúnmente asociado con la condición de *momento*, la cual se expresa en términos de los coeficientes de escalamiento como

$$\sum_{k=0}^{N_k-1} (-1)^k a_k k^m = 0, \quad (3.34)$$

donde los momentos de desvanecimiento son $N_k/2$, brindando así una buena representación del comportamiento de señales polinomiales. En las funciones wavelet de Daubechies el largo de sus soportes se puede expresar como $N_k/2 - 1$, que como se ilustra en la Figura 3.5. Por ejemplo la función wavelet de Daubechies, Db2, tiene un soporte de largo 1, mientras que para la Db4 el largo de su soporte es igual a 3, y así sucesivamente para cada uno de los momentos de desvanecimiento. Las siglas ‘Db’ corresponden a la familia de wavelets de Daubechies y el número consecuente a estas representa los coeficientes distintos a cero. N_k , es notable que las secuencias de wavelets es dada por momentos iguales a $N_k/2$ para cada caso.

En (3.27) se expresa la ecuación correspondiente al filtro escalable para la wavelet

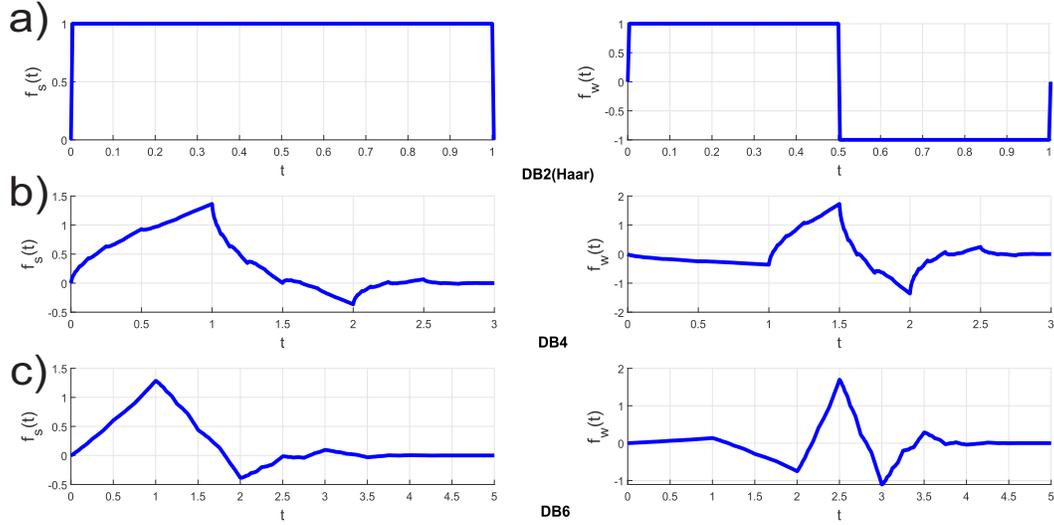


Figura 3.5: Funciones Daubechies, expresadas de derecha a izquierda, función de escalamiento y función wavelet para (a) soporte D2, (b) soporte D4 y (c) soporte D6 respectivamente.

de Haar, la cual es derivada de (3.25). Tomando esto en cuenta una función wavelet de 4 coeficientes (Db4) es igual a

$$\varphi(x) = \alpha_1\varphi(2x) + \alpha_2\varphi(2x - 1) + \alpha_3\varphi(2x - 2) + \alpha_4\varphi(2x - 3). \quad (3.35)$$

Los coeficientes α_j corresponden a los dados en (3.26) y son los llamados coeficientes de escalamiento cuyo resultado es

$$\alpha_1 = \frac{1 + \sqrt{3}}{4\sqrt{2}}, \alpha_2 = \frac{3 + \sqrt{3}}{4\sqrt{2}}, \alpha_3 = \frac{3 - \sqrt{3}}{4\sqrt{2}}, \alpha_4 = \frac{1 - \sqrt{3}}{4\sqrt{2}}. \quad (3.36)$$

Mientras que la función wavelet Db4 de acuerdo con (3.29) y (3.30) es dada como

$$\psi(x) = \beta_1\varphi(2x) + \beta_2\varphi(2x - 1) + \beta_3\varphi(2x - 2) + \beta_4\varphi(2x - 3). \quad (3.37)$$

Los coeficientes wavelet β_j de acuerdo con (3.29) y el resultado de (3.36) son

$$\beta_1 = \frac{1 - \sqrt{3}}{4\sqrt{2}}, \beta_2 = \frac{\sqrt{3} - 3}{4\sqrt{2}}, \beta_3 = \frac{3 + \sqrt{3}}{4\sqrt{2}}, \beta_4 = \frac{-1 - \sqrt{3}}{4\sqrt{2}}. \quad (3.38)$$

La Transformada Daubechies al igual que la Transformada de Haar puede ser calculada a través de las funciones de escala y wavelet, utilizando el análisis multi-resolución para implementar un algoritmo de descomposición [7, 11]. De esta manera dada una señal $a_m[k] = X[k]$ de N valores en un primer nivel de descomposición la señal promedio resultante es igual al producto escalar

$$a_{m-1} = a_m \cdot \mathbf{V}_m^1, \quad (3.39)$$

Donde \mathbf{V}_m^1 es el vector que contiene los coeficientes de escalamiento para el primer nivel de transformación, con un desplazamiento a través de la señal de la forma

$$\begin{aligned} V_1^1 &= (\alpha_1, \alpha_2, \alpha_3, \alpha_4, 0, 0, \dots, 0) \\ V_2^1 &= (0, 0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, 0, 0, \dots, 0) \\ V_3^1 &= (0, 0, 0, 0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, 0, 0, \dots, 0) \\ &\vdots \\ V_{N/2-1}^1 &= (0, 0, \dots, 0, \alpha_1, \alpha_2, \alpha_3, \alpha_4) \\ V_{N/2}^1 &= (\alpha_3, \alpha_4, 0, 0, \dots, 0, \alpha_1, \alpha_2). \end{aligned} \quad (3.40)$$

De igual manera la señal de las fluctuaciones o detalles es el producto escalar de

$$d_m = a_m \cdot \mathbf{W}_m^1, \quad (3.41)$$

con \mathbf{W}_m^1 como el vector que contiene los coeficientes wavelet del primer nivel de la transformación de desplazamientos iguales a

$$\begin{aligned} W_1^1 &= (\beta_1, \beta_2, \beta_3, \beta_4, 0, 0, \dots, 0) \\ W_2^1 &= (0, 0, \beta_1, \beta_2, \beta_3, \beta_4, 0, 0, \dots, 0) \\ W_3^1 &= (0, 0, 0, 0, \beta_1, \beta_2, \beta_3, \beta_4, 0, 0, \dots, 0) \\ &\vdots \\ W_{N/2-1}^1 &= (0, 0, \dots, 0, \beta_1, \beta_2, \beta_3, \beta_4) \\ W_{N/2}^1 &= (\beta_3, \beta_4, 0, 0, \dots, 0, \beta_1, \beta_2). \end{aligned} \quad (3.42)$$

Como se puede observar en (3.40) y (3.42) las secuencias corresponden a traslaciones de dos unidades de tiempo de los coeficientes a través de la señal original. Este proceso presenta un inconveniente, el cual muestra una clara diferencia al proceso que conlleva la TW de Haar, debido a que existen 4 coeficientes ($\alpha_1, \alpha_2, \alpha_3, \alpha_4$) que al ser trasladados hasta $N - 2$ unidades de tiempo los correspondientes α_3 y α_4 desbordarían el largo de la señal original. La mejor manera para evitar este problema es la técnica de *wrap-around*, la cual consiste en “envolver” la señal tal y como se muestra en $V_{N/2}^1$ y $W_{N/2}^1$ este proceso garantiza que cualquier señal pueda ser descompuesta siempre y cuando cumpla que $N/2$ corresponda a un valor par.

Finalmente al aplicar el proceso inverso a la señal que ha sido procesada se puede obtener la reconstrucción de la misma, que de manera generalizada para los distintos niveles de transformación se define como

$$X[k] = \mathbf{A}^k + \mathbf{D}^k + \dots + \mathbf{D}^2 + \mathbf{D}^1, \quad (3.43)$$

donde \mathbf{A}^k para $N_k = N/2^k$ es igual a

$$\mathbf{A}^k = (a_m \cdot \mathbf{V}_1^k) + \dots + (a_m \cdot \mathbf{V}_{N_k}^k) \mathbf{V}_{N_k}^k \quad (3.44)$$

y \mathbf{D}^k corresponde a

$$\mathbf{D}^k = (a_m \cdot \mathbf{W}_1^k) + \dots + (a_m \cdot \mathbf{W}_{N_k}^k) \mathbf{W}_{N_k}^k \quad (3.45)$$

3.4 Transformación wavelet packet

Una transformada por wavelet packets (TWP) es una generalización de las transformaciones wavelets convencionales, pero que es asociada a las llamadas cajas de Heisenberg [11] las cuales muestran una expansión en el dominio de la frecuencia y una contracción en el dominio del tiempo donde la mayor parte de la energía de estas se ve concentrada. Las wavelet packets son combinaciones lineales, que de acuerdo al análisis multi-resolución la principal diferencia radica, en que los coeficientes de escalamiento y detalles son descompuestos en cada uno de los niveles. En la Figura 3.6 es mostrado el esquema del árbol de descomposición de las wavelet packets, donde en un primer nivel de transformación no es perceptible una diferencia al primer nivel mostrado en la Figura 3.2. Sin embargo, para niveles posteriores ambos coeficientes son procesados, produciendo así que la señal pueda ser descompuesta en un mayor número de coeficientes.

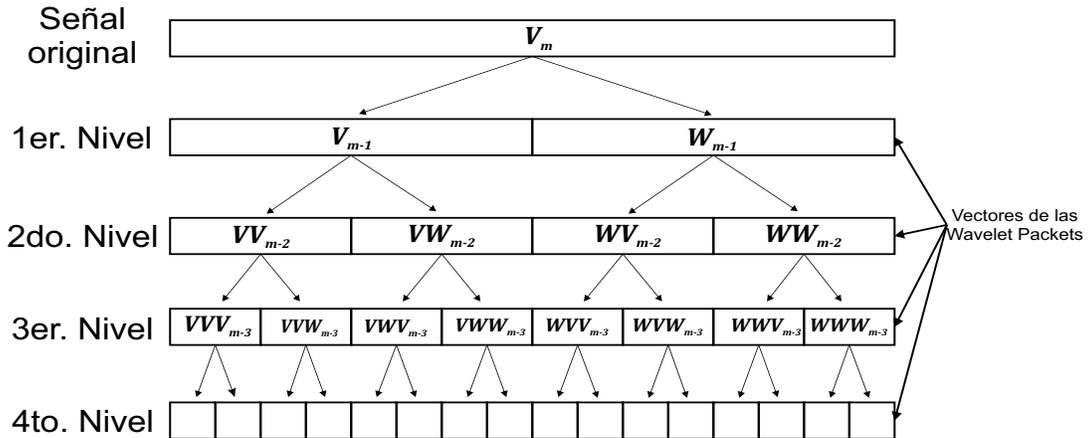


Figura 3.6: Diagrama esquemático de la descomposición por wavelet packets, donde ‘V’ y ‘W’ de izquierda a derecha denota el filtro por el que ha pasado previamente la señal.

El poder obtener no solo la descomposición de las versiones promedio de la señal original si no también de los detalles, permite analizar otro tipo de perspectivas, ya que se puede observar una diferencia entre los coeficientes que brinda el árbol de descomposición de la Figura 3.6 a comparación de los árboles de descomposición de otros métodos de transformación wavelet. Desde el punto de vista de los BF el algoritmo iterativo de filtros pasa-bajas y pasa-altas es generalizado y presentado en la Figura 3.7, donde es mostrado el conjunto de filtros espejo y los operadores de diezmación en la parte correspondiente al análisis. Mientras que para el caso de síntesis se presentan los operadores de interpolación junto a los mismos filtros espejo.

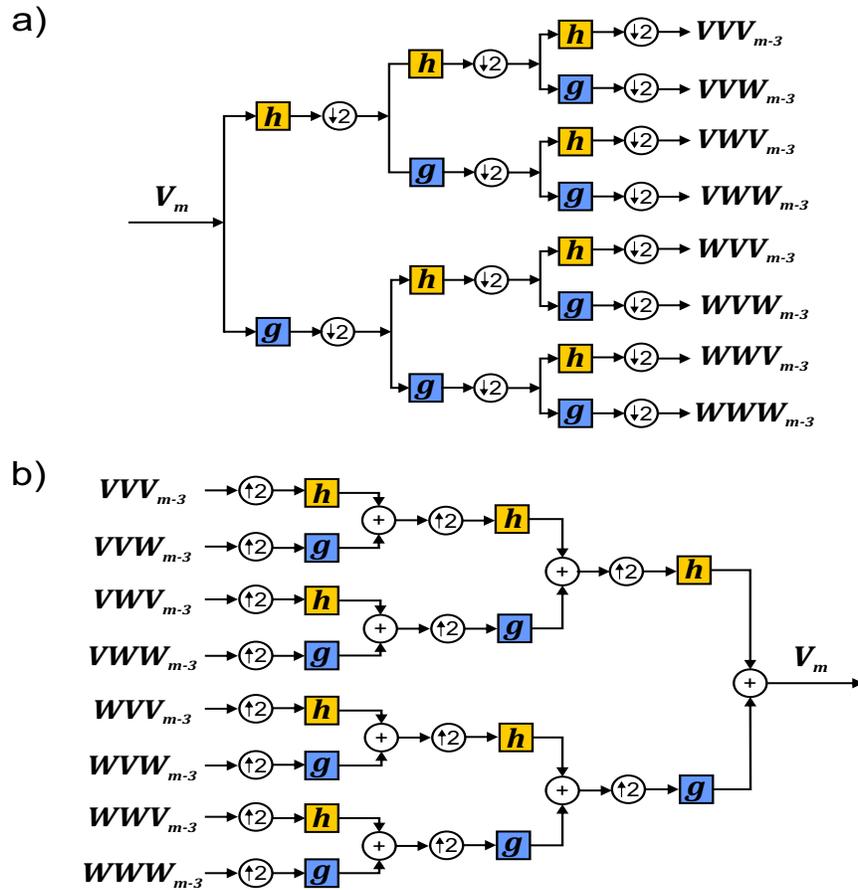


Figura 3.7: Transformada wavelet Packet de segundo nivel desde el enfoque de los BF, es mostrado en (a) el banco de análisis y en (b) el banco de síntesis

Al dividir la frecuencia de las señales en diversos intervalos las transformaciones a través de las wavelet packets, el número de operaciones aumenta en cada nivel calculado, pero que a su vez conlleva el poder aplicar nuevas técnicas tanto en el área de compresión de datos como en la extracción de características en señales de una o dos dimensiones.

3.5 Transformada wavelet híbrida

En términos generales, los sistemas híbridos son sistemas dinámicos que implican la interacción de diferentes tipos de dinámicas [13]. Que para el caso de las transformaciones wavelet se traduce en una interacción entre dos o más tipos de funciones wavelet, esto se logra a través de ir intercalando en cada uno de los niveles de digitalización alguna de las funciones. Tal y como se muestra en la Figura 3.8, podemos observar que para el caso en 1D los coeficientes resultantes de aplicar en primer lugar la TW de Haar, sirven como entrada para poder ser nuevamente procesados por la TW Db4 y así sucesivamente en un proceso iterativo en cascada.

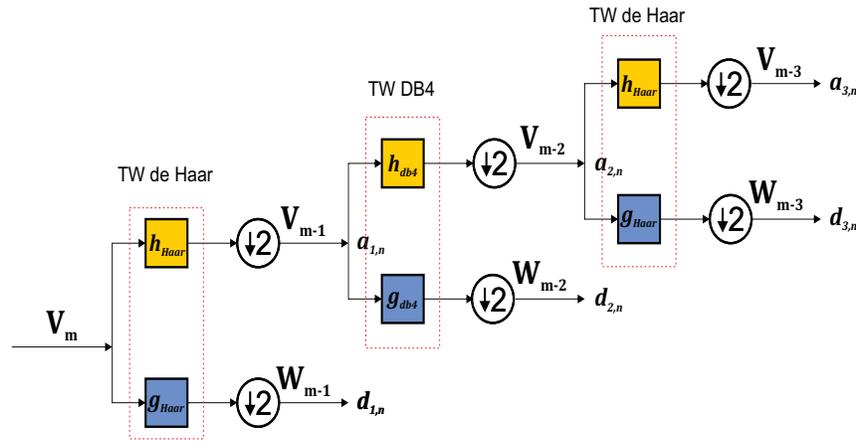


Figura 3.8: Diagrama de descomposición híbrido (Banco de análisis) piramidal para señales en una dimensión visto con el enfoque de los Bancos de Filtros.

La meta en este proceso, es el de poder aplicar y comparar los resultados en la extracción de características, en señales en 2D y así poder determinar el poder de los sistemas híbridos por sobre los métodos convencionales.

Es importante recalcar, que la variación de las dinámicas aplicadas pueden ser de formas muy distintas y estar compuestas por cualquiera de las funciones wavelet definidas hasta el momento. Sin embargo, para nuestro caso fue tomada la función wavelet de Haar y la de Daubechies 4 como la base del sistema a implementar para la extracción de características en imágenes de faciales.

Para ilustrar lo anterior, en la Figura 3.9 se ilustra un diagrama de bloques que nos auxilia para lograr una mejor comprensión del intercalado de transformación wavelet. Donde al tener una señal de entrada de la forma $c_0[k] = X[k]$, es posible obtener resultados diferentes al cambiar el orden de los filtros en el algoritmo y así ampliar el área de experimentación. Ya que las regiones de interés en una señal cambian dependiendo de lo buscado ya sea en aplicaciones de compresión de datos o específicamente como el tema central de este trabajo, en la extracción y clasificación de características en imágenes faciales.

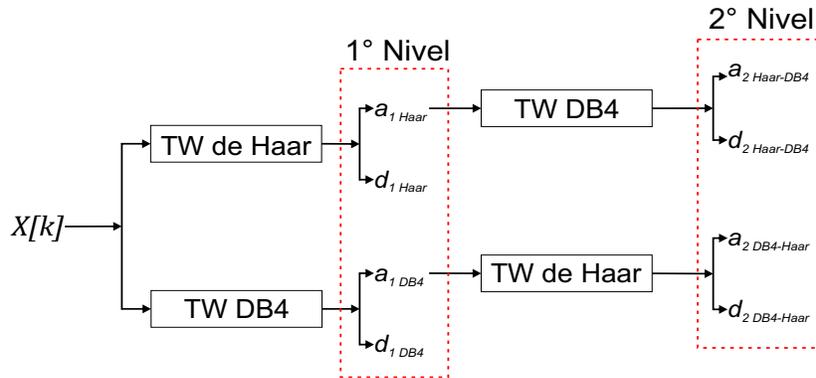


Figura 3.9: Diagrama a bloques del sistema híbrido de 2 niveles de dos funciones wavelet; en el primer caso se muestra la TW Híbrida Haar-Db4 y en el segundo la TW Híbrida Db4-Haar.

Como se puede apreciar, un sistema híbrido no presenta distinción a las ya estudiadas wavelet convencionales. Es por esto, que este tipo de análisis cobra sentido en niveles superiores de transformación en donde pueden ser combinados. Por último, cabe recalcar que existen pocos estudios, sobre la aplicación de wavelets híbridas en temas del procesamiento de imágenes y su beneficio será puesto a prueba en capítulos posteriores, con el objetivo de que ayuden a la clasificación de datos dentro de bases con gran tamaño.

3.6 Transformada wavelet Bidimensional

Como se ha mencionado, la TW en el tratamiento de señales unidimensionales es una gran herramienta de descomposición. No obstante esta puede ser calculada para el caso de múltiples dimensiones, específicamente el conjunto de datos en arreglos bidimensionales (imágenes) ha sido tema de interés en técnicas avanzadas del procesamiento de señales [1, 12]. La implementación de la llamada TWD bidimensional (TWD2D) se realiza al tomar todos los productos tensoriales posibles de las funciones unidimensionales. La forma más simple es usar la misma escala en las direcciones vertical y horizontal con el fin de obtener transformaciones cuadradas (aunque sin duda es posible obtener transformaciones rectangulares). Los siguientes pasos ejemplifican el proceso necesario para realizar los productos tensoriales, una vez seleccionada alguna TWD en 1D:

1. Una vez implementada una transformación en 1D, ésta se aplica en un primer nivel de descomposición a cada uno de los renglones de un arreglo bidimensional f .
2. El primer paso generara un nuevo arreglo al que nuevamente le será aplicada la transformación en 1D; en este caso transformando solo el primer nivel de descomposición de las columnas del arreglo.

Este proceso al igual que para el caso unidimensional, puede ser recursivo y reversible en los distintos niveles de transformación. Los productos tensoriales generan cuatro funciones base bidimensionales separables

$$\Phi(x, y) = \varphi(x)\varphi(y), \quad (3.46a)$$

$$\Psi^H(x, y) = \psi(x)\varphi(y), \quad (3.46b)$$

$$\Psi^V(x, y) = \varphi(x)\psi(y), \quad (3.46c)$$

$$\Psi^D(x, y) = \psi(x)\psi(y), \quad (3.46d)$$

donde $\Phi(x, y)$ corresponde a la función de escala y $\Psi^H(x, y)$, $\Psi^V(x, y)$, $\Psi^D(x, y)$ son funciones wavelet bidimensionales, brindando así una aproximación de menor resolución para el caso de la función de escala y una perspectiva de los detalles horizontales, verticales y diagonales respectivamente para cada una de las funciones wavelet. Todas estas funciones se definen, como el producto de la función de escala unidimensional φ y una correspondiente wavelet ψ . Al igual que para el caso en una dimensión, se pueden definir las funciones base de escala y wavelet como

$$\Phi_{j;m,n}(x, y) = 2^{j/2}\Phi(2^j x - m, 2^j y - n), \quad (3.47a)$$

$$\Psi_{j;m,n}^d(x, y) = 2^{j/2}\Psi^d(2^j x - m, 2^j y - n). \quad (3.47b)$$

La expansión de una función $f(x, y)$ de energía finita considerando las bases ortonormales para $L^2(\mathbb{R}^2)$ de (3.47) se puede expresar como

$$\begin{aligned} f(x, y) = & \frac{1}{\sqrt{MN}} \sum_m \sum_n \mathbf{a}_{j_0;m,n} \Phi_{j_0;m,n}(x, y) \\ & + \frac{1}{MN} \sum_{d=H,V,D} \sum_{j=j_0} \sum_m \sum_n \mathbf{d}_{j;m,n}^d \Psi_{j;m,n}^d(x, y), \end{aligned} \quad (3.48)$$

donde

$$\mathbf{a}_{j;m,n} = \int \int f(x, y), \Phi_{j;m,n}(x, y) dx dy, \quad (3.49a)$$

$$\mathbf{d}_{j;m,n}^d = \int \int f(x, y), \Psi_{j;m,n}^d(x, y) dx dy, \quad (3.49b)$$

corresponden a los coeficientes de escala y wavelet, respectivamente. El par de ecuaciones (3.48) y (3.49) constituyen la TWD2D para el caso de síntesis y análisis de la señal.

Una imagen discretizada, puede ser expresada como un arreglo matricial de M renglones y N columnas de números reales, lo cual abre un sinfín de posibilidades entorno al procesamiento

digital de imágenes a través de la TWD2D, debido a que retomando los bancos de filtros como una forma de implementar las expansiones en serie de una señal en tiempo discreto, una imagen puede ser descompuesta y analizada a través de sus principales componentes. Como se pudo observar en la Figura 3.3, el elemento clave de los bancos de filtros es el uso de los diezmadores e interpoladores los cuales se encargan de realizar los cambios en la frecuencia de muestreo. Al ser implementados en el banco de filtros mostrado en la Figura 3.10 la señal de entrada será diezmada en un factor de 2, con lo cual para el caso de imágenes resultara en una división de cuatro sub-imágenes cada una con dimensiones iguales a $M/2$ filas y $N/2$ columnas por cada filtraje que le sea aplicado acorde al análisis multi-resolución. El arreglo de filtros de acuerdo con el algoritmo piramidal de descomposición de Mallat [11] consiste en una entrada $x[m, n] = a_{j+1, m, n}$ que es sometida a un primer filtraje a lo largo de sus renglones por h y g que son filtros pasa-bajas y pasa-altas, respectivamente, seguido de la diezmación entorno a sus columnas. El proceso es repetido para las 2 nuevas señales obtenidas sometiendo en esta ocasión las columnas de la imagen y por ende diezmando la parte de los renglones, resultando en las 4 sub-imágenes con resolución de $M/2$ filas y $N/2$ columnas.

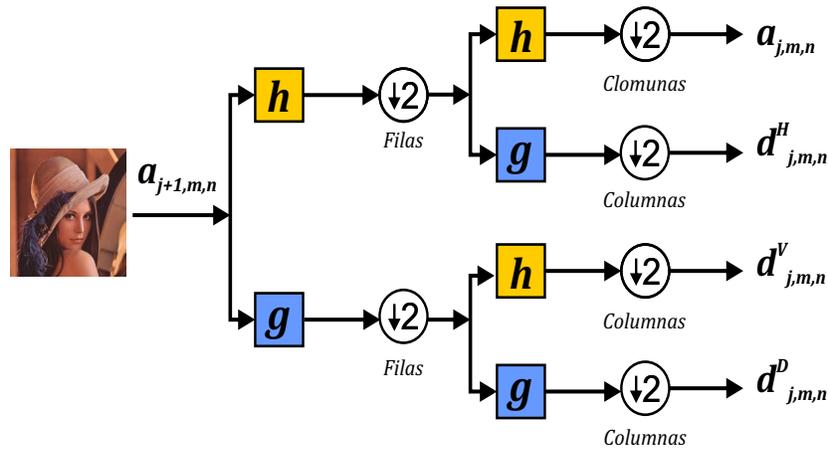


Figura 3.10: Diagrama de descomposición (Banco de análisis) piramidal para señales en dos dimensiones visto con el enfoque de los Bancos de Filtros.

A esto se le conoce como el primer nivel de transformación wavelet para señales bidimensionales, el cual de igual forma que para el caso unidimensional puede ser iterado tantas veces como la señal pueda ser dividida en un factor de 2. En la Figura 3.10 a los coeficientes calculados a partir de h (filtro pasa-bajas) se les conoce como “tendencias” y a los calculados por g (filtro pasa-altas) como “fluctuaciones”.

Acorde a la Figura 3.11 en un primer nivel de transformación la sub-imagen que aparece en el cuadrante superior izquierdo $a_{j, m, n}$ corresponde a una versión de menor resolución a la original conocida como el promedio de la señal, calculada a partir de las tendencias a lo largo de los renglones y columnas. Otra manera de expresar $a_{j, m, n}$ es como el producto escalar de la imagen con las funciones de escalamiento.

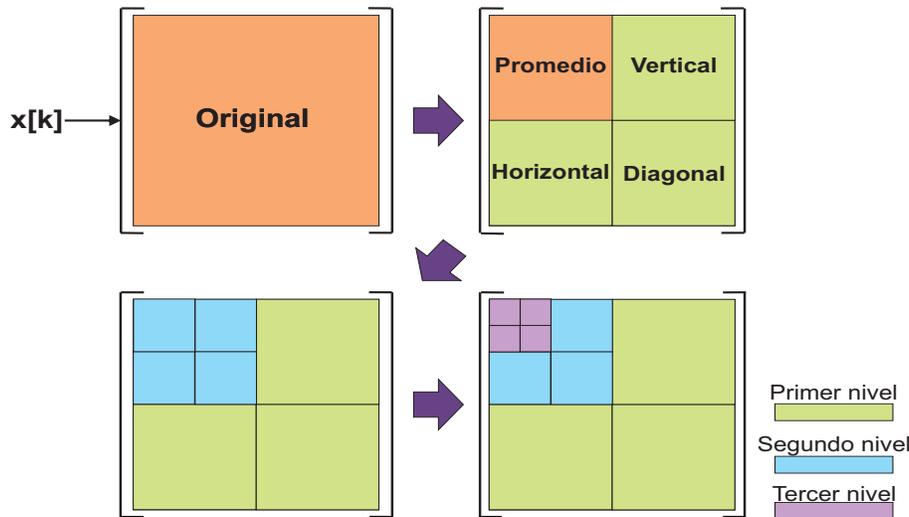


Figura 3.11: Esquema de los tres primeros niveles de transformación para una imagen. En este se muestra el acomodo de cada una de las imágenes resultantes de acuerdo con las operaciones aplicadas

La sub-imagen $d_{j,m,n}^H$ es obtenida tras calcular las tendencias alrededor de los renglones de la imagen original seguido de calcular las fluctuaciones a través de las columnas. Esto provoca que los bordes horizontales de la imagen sean enfatizados brindando los detalles de estos. En $d_{j,m,n}^V$ se encuentran los detalles verticales de la imagen, calculados de igual forma que $d_{j,m,n}^H$ con la diferencia de que el orden de las tendencias y fluctuaciones son invertidos. Finalmente la sub-imagen $d_{j,m,n}^D$ corresponde a los detalles diagonales y se obtiene de las fluctuaciones en renglones de la imagen original seguido de las fluctuaciones en las columnas. Esto comprende la TWD2D donde la extracción de características en imágenes resulta de gran utilidad, debido a que detalles que no pueden ser observados a simple vista son enfatizados para aplicaciones como el reconocimiento de imágenes y las versiones promedio en estudios de compresión de archivos.

3.7 Estimación de la calidad de los coeficientes extraídos

Con el objetivo de evaluar los coeficientes extraídos mediante los métodos de transformación wavelet, en esta sección las imágenes que contiene los coeficientes fueron procesadas a través del calculo de el *error de norma relativa 2*. Esta operación representa uno de los métodos más simples en el reconocimiento de imágenes y otorga una idea clara de que coeficientes presentaran un mayor porcentaje éxito en procesos de clasificación posteriores de mayor complejidad. Para esto fueron tomadas un total de 208 imágenes contenidas en la base *BioID* de 384×216 pixeles y fueron implementadas las transformadas wavelet mencionadas en 3.2, 3.3, 3.4 y 3.5 en su forma bidimensional y discreta. En el capítulo 2 se mencionó una etapa de pre-procesamiento para las imágenes y del cual se derivaron 4 nuevas categorías; rostro, ojos, nariz y boca. Como ya ha sido mencionado, a través del análisis multi-resolución es posible generar proyecciones cada vez más pequeñas de una imagen en cada uno de los niveles de transforma-

ción, por lo que a cada una de las imágenes de las nuevas categorías le fueron aplicados tres niveles de descomposición, debido a que el tamaño de las imágenes impide aplicar un mayor número de niveles ya que esto implicaría reducirlas hasta un punto en el que los coeficientes contenidos son tan pocos que diferenciarlos resultaría en una tarea sumamente difícil. Una vez concluido este proceso para la extracción de los coeficientes, fueron procesados a través de una técnica conocida como el error relativo de norma 2 y que puede ser expresado como

$$D(f, g) = \frac{\sqrt{\sum_{n=1}^N \sum_{m=1}^M (f_{n,m} - g_{n,m})^2}}{\sqrt{\sum_{n=1}^N \sum_{m=1}^M (f_{n,m})^2}}, \quad (3.50)$$

en donde $D(f, g)$ es el error que existe al calcular la diferencia entre los coeficientes de una imagen f y una imagen g y con el cual es posible poder determinar que tan parecida es una de otra. Todo esto con la intención de determinar las sub-imágenes con las que puede ser llevada una mejor discriminación en procesos posteriores.

Al aplicar una TWD de dos dimensiones a las imágenes contenidas en las diferentes categorías, el resultado es una descomposición en una versión promedio más los detalles horizontales, verticales y diagonales cuyas dimensiones se ven diezmadas en un factor de 2 por cada uno de los niveles aplicados. En la Tabla 3.1 son mostradas las dimensiones resultantes para cada una de las sub-imágenes obtenidas al ser transformadas en uno, dos y tres niveles.

Tabla 3.1: Dimensión de las sub-imágenes obtenidas por la TW

Rasgo	Dimensiones originales (píxeles)	Dimensión en un 1° Nivel de TW (píxeles)	Dimensión en un 2° Nivel de TW (píxeles)	Dimensión en un 3° Nivel de TW (píxeles)
Ojos	48 × 168	24 × 84	12 × 42	6 × 21
Nariz	48 × 168	24 × 84	12 × 42	6 × 21
Boca	48 × 168	24 × 84	12 × 42	6 × 21
Rostro	168 × 168	84 × 84	42 × 42	21 × 21
Imagen Original	384 x 286			

Como puede ser observado en la tabla 3.1 las sub-imágenes resultantes son de un tamaño menor al de la imagen original y contendrán los coeficientes calculados por las diferentes funciones wavelet. Gracias a esto el analizar cualquiera de las nuevas versiones resulta en un menor número de operaciones debido a la diferencia de dimensiones. No obstante, las sub-imágenes contienen coeficientes que resaltan detalles o aspectos que pueden ser utilizados para realizar una discriminación entre todas las imágenes contenidas en la base de datos. Lo anterior fue comprobado al medir el error de norma relativa 2 (Ecuación 3.50) para así poder determinar la diferencia existente entre los coeficientes de una imagen en comparación con el resto.

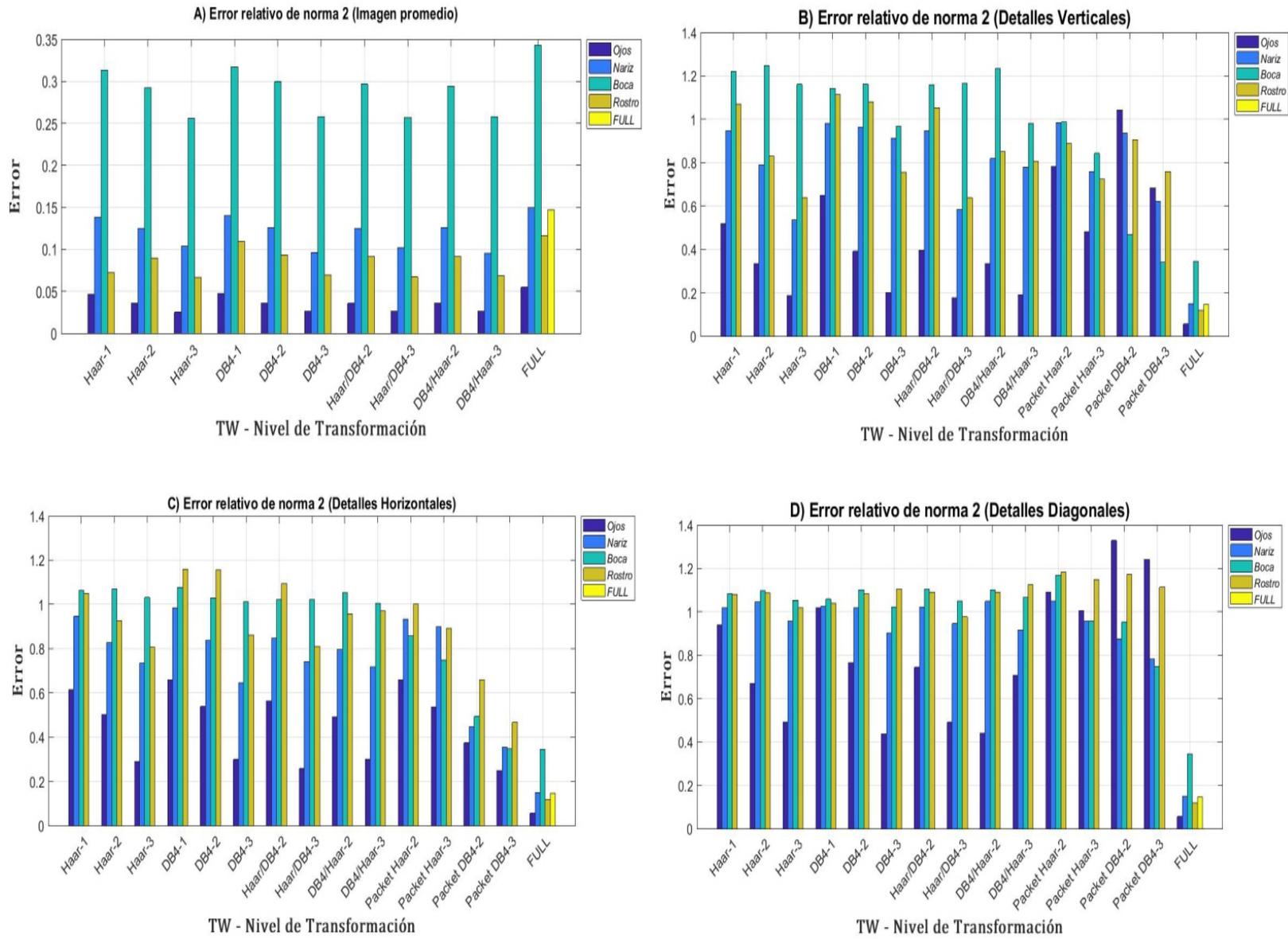


Figura 3.12: Error relativo de norma 2 a) sub-imágenes promedio, b) sub-imágenes de detalles horizontales, c) sub-imágenes de detalles verticales y d) sub-imágenes de detalles diagonales.

Para esto fueron seleccionadas un grupo de imágenes base f para cada una de las cuatro categorías en función de los factores intrínsecos y extrínsecos de iluminación presentada por las imágenes así como en el ángulo con el que fue capturado el rostro. En la Figura 3.12 son mostrados los resultados obtenidos en cada uno de los niveles de transformación para las diferentes TWD en dos dimensiones aplicadas a las cuatro categorías. Estos resultados corresponden al error encontrado entre las imágenes base f y un conjunto de imágenes g en las que se presentó el menor error.

Así mismo en la Figura 3.12 fueron incluidos, a manera de comparación, los resultados del error relativo de norma 2 entre una imagen base f y un conjunto de imágenes g a las cuales no les fue aplicado ningún tipo de procesamiento.

De acuerdo a los resultados obtenidos, no sólo es evidente, que el utilizar la TW disminuye el costo computacional al realizar cualquier tipo de procesamiento u operación de clasificación, debido a la disminución en las dimensiones de las sub-imágenes obtenidas al final de este proceso. Dichas sub-imágenes resultan ser versiones promedio y de detalles horizontales, verticales y diagonales de la imagen original, por lo tanto al llevar a cabo una comparación entre la eficacia de cada una al momento de realizar una discriminación de sus componentes, se puede predecir los rasgos y que nivel de transformación de cada una de las TWD presentaran mejores resultados en procesos posteriores de clasificación, debido a que entre menor sea el error que existe entre dos imágenes mayor tenderá a ser la eficacia al determinar si es que pertenecen a la misma clase. Estos resultados indican que para el área tanto de la boca como de la nariz se presentan los mayores errores e indican un menor poder de discriminación, mientras que para el área de los ojos y el rostro se presentan los mejores resultados. Por lo tanto, estos resultados ayudaran a facilitar la elección de los rasgos y el nivel de transformación necesarios para obtener el mejor rendimiento en la etapa de clasificación de las imágenes faciales. Por último, es evidente que las transformaciones wavelet no solo resultan ser una excelente técnica para disminuir el tiempo de ejecución en cualquier tipo de proceso, sino de igual forma un extractor de características apto para ser probado en diversos sistemas de clasificación.

CAPÍTULO 4

CLASIFICACIÓN DE IMÁGENES

Los sistemas de reconocimiento de patrones actualmente son sumamente utilizados para la identificación de huellas dactilares, secuencias de ADN (Ácido desoxirribonucleico), personas, etc. Estos sistemas están conformados por tres diferentes operaciones las cuales son el pre-procesamiento, la extracción de características y la clasificación. Específicamente, en los sistemas de reconocimiento de imágenes, la tarea de clasificación consiste en utilizar un vector de características extraídas de cada imagen, para determinar la clase a la que pertenecen [24]. No obstante, el mayor problema radica en que la clasificación tiene que ser discreta, esto quiere decir que cada imagen pertenece precisamente a una clase y el conjunto de clases cubre todo el espacio de salida posible. Estas dos restricciones no son necesariamente realistas; a veces las imágenes pueden pertenecer parcialmente a dos clases diferentes [25]. Es por eso que la clasificación perfecta no siempre es posible debido a que los elementos de una misma clase pueden presentar valores muy parecidos en sus características, dificultando la categorización de cada una, concluyendo en que la selección correcta de características o también llamadas predictores forma parte vital de los sistemas de clasificación ya que aumenta la probabilidad de determinar la clase correcta.

En términos de diseño e implementación, en un sistema de reconocimiento pueden ser usadas las **máquinas de aprendizaje**, las cuales son algoritmos que pueden modificarse o adaptarse para realizar predicciones probabilísticas en múltiples disciplinas y que en el caso de la clasificación de imágenes estos pueden ser seleccionados y evaluados para la obtención de sistemas complejos para el reconocimiento facial de personas. Las máquinas de aprendizaje se dividen en cuatro categorías las cuales son:

1. Aprendizaje Supervisado: El cual consiste en proveer al sistema, un conjunto de entrenamiento que contiene las respuestas correctas (categorías o clases) y que con base a estas, el algoritmo determina la respuesta a cualquier tipo de entrada.
2. Aprendizaje No Supervisado: En este caso las respuestas no son proporcionadas, por lo que el algoritmo identifica las respuestas con base a la similitud que existe entre las entradas.
3. Aprendizaje Reforzado: Este algoritmo es el punto medio entre el aprendizaje supervisado

y no supervisado, ya que las respuestas le son brindada, pero es necesario explorar diferentes posibilidades, hasta encontrar la respuesta correcta.

4. Aprendizaje Evolutivo: Estos algoritmos son capaces de adaptarse a las condiciones de las entradas para poder brindar la mejor solución.

El diseño de las máquinas de aprendizaje comúnmente suele ser representado como un ciclo que realiza ciertas tareas como se muestra en la Figura 4.1, donde la **recopilación de datos** comprende la selección de la base de datos o imágenes a utilizar dentro del sistema y es con base a esta, que son llevados a cabo los diferentes estudios. Tal es el caso del reconocimiento de personas, cuya selección en los datos es estudiada en el Capítulo 2. Mientras que la **selección de características** consiste en seleccionar los atributos con los cuales se pueda realizar una discriminación entre las distintas clases; dichos atributos forman el llamado “vector de características” x , donde x es un espacio Euclidiano p -dimensional \mathbb{R}^p y es definido como el *espacio de características* [24]. El proceso de extracción de estos vectores se presentó en el Capítulo 3.

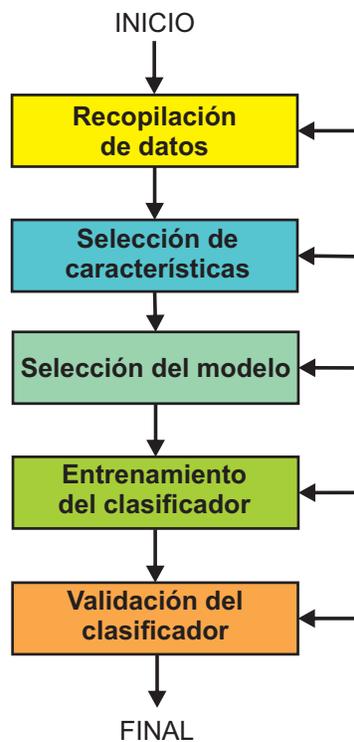


Figura 4.1: Diseño de una máquina de aprendizaje, donde los resultados de la evaluación pueden requerir la repetición de algunos pasos para obtener los mejores resultados.

La **selección del modelo** dado un conjunto de datos, consiste en escoger el algoritmo de clasificación que reduzca de mejor manera los errores. Existen muchos y diversos algoritmos, de los cuales se hablara y dará detalle en secciones posteriores. El **entrenamiento** y la **validación**, son los últimos pasos en el ciclo de las máquinas de aprendizaje y consisten en tomar una porción del conjunto de datos, que comúnmente en el caso del entrenamiento comprende un

intervalo entre el 70 % o 75 % del total de los elementos, y los elementos de este conjunto son usados en los algoritmos de clasificación para predecir la clase de nuevos datos de entrada. Finalmente la validación corresponde al 30 % o 25 % restante de los datos y son estos los que son usados como las nuevas entradas con el fin de evaluar el rendimiento del clasificador.

4.1 Algoritmos de clasificación

Un modelo de clasificación, es un algoritmo que forma parte de las máquinas de aprendizaje y representan la parte mas importante en la evaluación de los sistemas de clasificación. En términos del reconocimiento de imágenes, dichos clasificadores están basados en diferentes métodos para comparar las formas y apariencias de cada imagen. Los algoritmos basados en técnicas de clasificación y agrupamiento forman parte central de esta sección, es por eso que se presentan, las principales características y propiedades de los clasificadores de mayor uso, dentro del procesamiento y reconocimiento de imágenes.

4.1.1 Análisis de Componentes Principales

El Análisis de Componentes Principales (PCA por sus siglas en inglés) es un método de proyección estadística multivariante. Comúnmente, el PCA es utilizado para monitorear el desempeño operativo de grandes procesos continuos, para lo cual es necesario transformar un espacio de características lineales de alta dimensión a uno de baja dimensión, donde las variables obtenidas conservan la mayor parte de la información de los datos originales [26]. Esta técnica consiste en tomar una matriz $X(n \times m)$, donde n es el número de muestras y m es el número de variables del proceso. Para eliminar los errores causados por las distintas dimensiones de las variables, los datos de la matriz X son estandarizados por el cálculo de su covarianza, tal que

$$V = \frac{1}{m-1} X X^T. \quad (4.1)$$

El siguiente paso consiste en calcular los eigenvalores representados por λ_i y los eigenvectores p_i de la matriz de covarianza, donde $i \in [1, 2, \dots, m]$, para después ordenar de mayor a menor los eigenvectores por eigenvalores tal que $(\lambda_1 \geq \lambda_2 \geq \dots \lambda_m \geq 0)$. Una vez realizado esto, es tomado el primer k eigenvector de la matriz P expresada por (p_1, p_2, \dots, p_k) , generalmente el PCA de una matriz X puede ser representado por la suma de los productos de los m vectores, donde

$$X = T P^T = \sum_{j=1}^m t_j p_j^T = t_1 p_1^T + t_2 p_2^T + \dots + t_m p_m^T, \quad (4.2)$$

donde t_j está definido como el vector de los componentes principales de X . Entonces se puede descomponer el espacio de datos en: (1) el principal espacio compuesto por P y (2)

en el espacio residual compuesto por $[p_{k+1}, \dots, p_m]$. Por último, se puede decir que el primer componente principal t_1 contiene la mayor información de los datos originales y esta disminuye gradualmente a medida que aumenta el índice del componente principal. Por deducción es posible obtener el modelo que reduzca eficientemente la dimensión de los datos.

4.1.2 Análisis Discriminante

El análisis discriminante lineal (LDA por sus siglas en inglés) es una técnica de clasificación que surge como una estrategia alternativa al PCA. El PCA puede ser utilizado para la reducción de las características, pero es sumamente difícil discriminar las clases una vez llevada a cabo la reducción. El LDA es una estrategia en donde son seleccionadas las características, tal que la *dispersión de las clases* (Figura 4.2a) sea minimizada y la *dispersión entre clases* (Figura 4.2b) sea maximizada para la distribución de muestras en el espacio de características [32]. La matriz de dispersión de las clases se denota por Σ_W y es definida por

$$\Sigma_W = \sum_{k=1}^K P(k) \cdot \Sigma_k, \quad (4.3)$$

donde $P(k)$ es la probabilidad a priori de una muestra perteneciente a la clase k y Σ_k es la matriz de covarianza para la distribución de muestras de la clase k . La matriz de dispersión entre las clases Σ_B es igual a

$$\begin{aligned} \Sigma_B &= \sum_{k=1}^K P(k) (\bar{\mathbf{f}}_k - \bar{\mathbf{f}})(\bar{\mathbf{f}}_k - \bar{\mathbf{f}})^T, \\ \bar{\mathbf{f}} &= \sum_{k=1}^K P(k) \mathbf{f}_k. \end{aligned} \quad (4.4)$$

Esta matriz, mide la dispersión del vector de características promedio $\bar{\mathbf{f}}_k$, para las diferentes clases k con respecto al vector de características promedio $\bar{\mathbf{f}}$ de todo el conjunto de datos. Entonces se puede seleccionar un conjunto de ejes de características ortogonales, de modo que las variaciones combinadas de Σ_B y la inversa de Σ_W dada por la suma de elementos diagonales de su producto $tr(\Sigma_W^{-1} \Sigma_B)$ (donde tr es la traza de la matriz) sea la máxima [32]. Por lo tanto al igual que en el PCA, es posible calcular los eigenvectores y eigenvalores de $\Sigma_W^{-1} \Sigma_B$, en donde las características a lo largo de los eigenvalores pueden ser seleccionadas para una subsecuente clasificación.

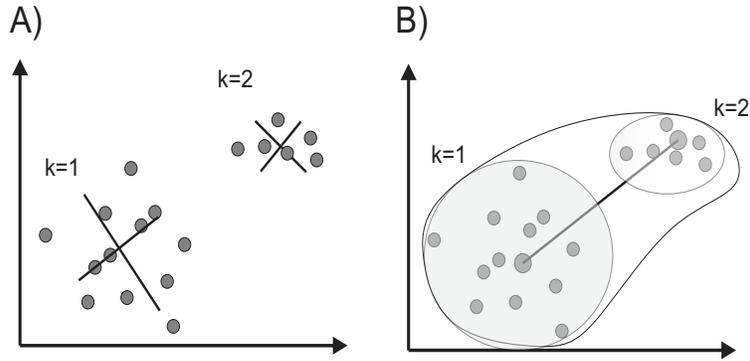


Figura 4.2: Ejemplo de A) matriz de dispersión de las clases y B) matriz de dispersión entre las clases.

4.1.3 Máquinas de Soporte Vectorial

Una Máquina de Soporte Vectorial (SVM por sus siglas en inglés), es uno de los algoritmos más populares en cuanto a máquinas de aprendizaje se refiere, y principalmente se debe a que presenta un alto desempeño en la clasificación de grandes conjuntos de datos [25]. Una SVM consiste en dividir un espacio de características en regiones, las cuales son llamadas de decisión y si uno de los vectores de características se encuentra en alguna de éstas le es asignada la clase correspondiente [9]. Para ilustrar la “funcionalidad” de este algoritmo, se considera el ejemplo sencillo de la clasificación binaria, donde dos clases son representadas como vectores de características bidimensionales (Figura 4.3) tal que $x_i = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ y que pueden ser separadas en dos regiones por medio de un vector llamado de “decisión lineal” que se define como

$$f(x) = w_1x_1 + w_2x_2 + b = 0, \tag{4.5}$$

en donde los coeficientes w_1 y w_2 determinan la pendiente de la línea y b es la contribución de peso parcial [25]. De manera general para un espacio de características p -dimensional $x_i = (x_1x_2 \cdots x_p)^T \in \mathbb{R}^p$ la función de decisión lineal cambia por una superficie o discriminante llamado hiperplano tal que

$$\begin{aligned} f(x) &= \sum_{i=1}^p w_i x_i + b \\ &= \langle \mathbf{w}, \mathbf{x} \rangle + b \\ &= \mathbf{w}^T \mathbf{x} + b = 0. \end{aligned} \tag{4.6}$$

El hiperplano debe ser seleccionado, de tal forma que la distancia mínima con respecto a los vectores de entrenamiento más cercanos o también llamados de soporte, sea la máxima, esto es posible cuando la distancia d es igual a $\frac{1}{\|\mathbf{w}\|}$ o también expresada como

$$\min_{x_i} \|\mathbf{w}^T \mathbf{x} + b\| = 1. \quad (4.7)$$

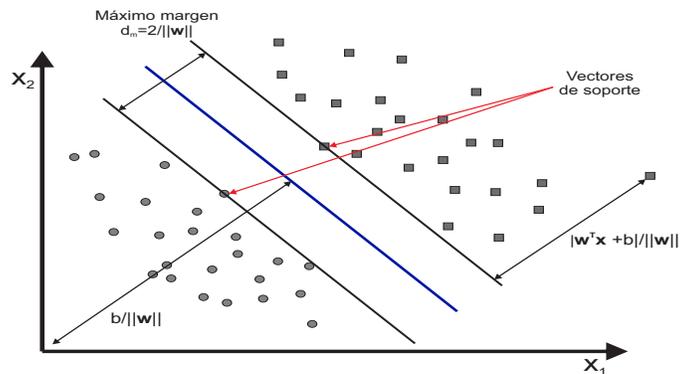


Figura 4.3: Ejemplo de una máquina de soporte vectorial de clasificación binaria.

Esta es la idea básica de las máquinas de soporte vectorial, sin embargo si se supone un conjunto de datos cuyo espacio de características no pueda ser separado eficientemente por los vectores de decisión lineal, es necesario usar más dimensiones para poder ser capaces de encontrar una decisión lineal que separe todas las clases. A esta modificación en los vectores se le conoce como kernel y existen tres diferentes tipos de funciones base que son comúnmente utilizadas; las polinomiales, sigmoides y radiales. La elección del kernel y sus parámetros puede suponer el rendimiento que tendrá el SVM ante el conjunto de datos, en nuestro caso fueron seleccionados los siguientes kernels, por ser los que mejores resultados han presentado en diversos estudios [25]:

1. SVM Lineal.
2. SVM Cuadrático.
3. SVM Cúbico.
4. SVM de Media Gaussiana.

4.1.4 K- Vecindarios Cercanos

Otro de los algoritmos con mayor tendencia de uso es el llamado K-Vecindarios Cercanos (KNN por sus siglas en inglés), el cual es un algoritmo de aprendizaje que dado un conjunto de entrenamiento y una muestra de datos que contengan un etiqueta con la clase conocida es posible calcular la distancia entre cada muestra de los datos de entrenamiento y un nueva muestra sin etiqueta de clase [26]. La forma de realizar la clasificación de datos con base a este método, consiste en generar una matriz X de $p \times r$ que contenga los datos de entrenamiento con las etiquetas de las clases conocidas; posteriormente se genera una matriz de validación Y de $q \times r$, en donde p y q son el número de muestras y r es el número de predictores. Una

vez generadas estas matrices el algoritmo de clasificación se divide en cinco importantes pasos descritos a continuación:

1. Calcular la distancia que existe entre una muestra de validación y_j y una muestra de entrenamiento x_i .
2. Ordenar todos los valores de las distancias de manera ascendente.
3. Seleccionar k puntos con la distancia más corta.
4. Calcular la frecuencia en las diferentes clases para k puntos.
5. Seleccionar la clase con la frecuencia más alta dados los k puntos para predecir la clase de cada muestra del conjunto de validación.

La distancia entre las muestras puede ser calculada de múltiples formas, no obstante, la distancia de Mahalanobis es la más utilizada debido a que puede eliminar los efectos causados por las diferentes dimensiones de las muestras y reduce la distorsión causada por la combinación lineal de las características. La distancia de Mahalanobis es dada por

$$distancia(x_i, y_j) = \sqrt{(x_i - y_j)V^{-1}(x_i - y_j)^T}, \quad (4.8)$$

donde V es la covarianza entre la matriz X y Y . El algoritmo de aprendizaje KNN posee diferentes variaciones en el proceso de clasificación, ya que el uso de diferentes esquemas de validación puede facilitar la interpretación de los datos y el poder disminuir el tiempo de ejecución. Dentro de las variantes que existen de los esquemas de validación y selección de vecindarios pueden ser mencionados los siguientes algoritmos

1. k-Vecindarios Cercanos Finos (KNN Fine),
2. k-Vecindarios Cercanos Medios (KNN Medium),
3. k-Vecindarios Cercanos Cosenoidales (KNN Cosine),
4. k-Vecindarios Cercanos Cúbicos (KNN Cubic) y
5. k-Vecindarios Cercanos Ponderados (KNN Weighted),

los cuales pueden ser usados para un análisis comparativo y así determinar que tipo de esquema se adapta mejor al conjunto de datos de entrada.

4.1.5 Clasificadores en Conjunto

La técnica de combinar las predicciones generadas por múltiples algoritmos de clasificación para producir uno solo, se le conoce como clasificación en conjunto (Figura 4.4) y generalmente se pueden obtener resultados más precisos; debido a que ha sido demostrado que al realizar una buena combinación, un clasificador individual y uno en conjunto crean errores en diferentes espacios de entrada pero ambos obtienen resultados precisos [27]. Lo anterior es de suma importancia debido a que al generar una combinación de clasificadores, los resultados de estudiar un mismo conjunto de datos se pueden ver beneficiados al obtener una mayor precisión. No obstante, el combinar la salida de los clasificadores solo resulta útil si es que existe una discrepancia entre ellos, por lo tanto combinar dos o mas clasificadores idénticos no produce ningún beneficio y por lo tanto el desempeño de un clasificador en conjunto depende de la discrepancia entre todos los clasificadores, entre mayor sea ésta, mejores serán los resultados [27, 28].

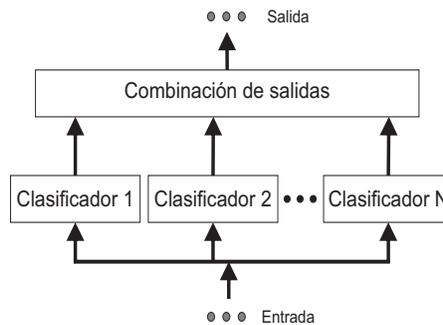


Figura 4.4: Diagrama a bloques de un sistema de clasificación en conjunto.

Debido a que la combinación de clasificadores puede ser muy diversa, han sido planteadas algunas estrategias para poder generar los conjuntos de clasificadores y así poder garantizar la máxima precisión. A continuación cuatro de los más utilizados algoritmos son descritos para su posterior implementación.

Empaquetamiento

La técnica de “empaquetamiento” es uno de los algoritmos más simples de implementar y además presenta buenos resultados. En este algoritmo el conjunto de datos de entrenamiento es utilizado para entrenar diferentes clasificadores del mismo tipo. El resultado de cada clasificador individual es combinado para poder tomar la mejor decisión basada en el mayor número de votos obtenidos por todo el conjunto. Con el fin de garantizar la diversidad en este escenario, se utiliza un modelo relativamente inestable para que se puedan obtener límites de decisión suficientemente diferentes para pequeñas perturbaciones en diferentes conjuntos de datos de entrenamiento [28].

Uno de los algoritmos con mayor uso es el clasificador de empaquetamiento de árboles de decisión, donde es combinado el algoritmo por empaquetamiento y el clasificador por árboles

de decisión [29]. En este algoritmo son tomados m muestras aleatorias del conjunto de datos originales y es entrenado N veces, entonces el resultado es obtenido por el voto de todas las iteraciones tal que

$$H(x) = c_{\arg \max_j \sum_{i=0}^N h_i^j(x)}, \quad (4.9)$$

donde h_i^j es la salida proveniente de los N clasificadores.

Impulsados

Un algoritmo de clasificación en conjunto de forma “impulsada” es aquel que combina máquinas de *aprendizaje débil* para formar una de *aprendizaje fuerte*. Similarmente a la técnica de empaquetamiento, la forma impulsada crea un conjunto de clasificadores que muestrea los datos, para obtener la mayoría de votos y poder tomar una decisión. Sin embargo, este algoritmo es orientado a proporcionar los datos de entrenamiento con mayo información para cada clasificador [28], esto es, en un primer clasificador son entrenados un subconjunto aleatorio del total de datos de entrenamiento disponibles y posteriormente el subconjunto de datos de entrenamiento para el segundo clasificador es elegido como el subconjunto más informativo del primero. Es decir, el segundo clasificador recibe los resultados del primero, de los cuales la mitad están clasificados correctamente y la otra mitad está mal clasificada. Lo cual puede ser replicado para N clasificadores más. Ejemplo de esto, son los siguientes algoritmos, los cuales son combinaciones entre la clasificación por árboles de decisión y la técnica de impulsado:

1. Árbol de decisión impulsado.
2. Sobre muestreo aleatorio (RUS por sus sigas en inglés) de árboles de decisión impulsados.

Sub-espacio LDA

Un sub-espacio de análisis discriminante lineal es una técnica en la que son combinadas dos técnicas previamente descritas en las secciones 4.1.1 y 4.1.2 las cuales son el análisis de componentes principales y el análisis de discriminación lineal, respectivamente. El secreto en esta combinación, radica en que la reducción dimensional en tareas de clasificación basadas en el PCA es más óptimo que los métodos basados en LDA, debido a que los basados en PCA buscan encontrar proyecciones con un error de construcción mínimo, mientras que los que se encuentran basados en LDA buscan proyecciones que conserven la información discriminante. Un sub-espacio LDA (SLDA) surge entonces, al realizar un LDA en un sub-espacio generado por el PCA de baja dimensión, y la transformación pueda descartar el espacio nulo de las matrices de dispersión con deficiencia de rango, para que las formulaciones de LDA permanezcan bien condicionadas; este proceso es ampliamente explicado en [30].

Sub-espacio KNN

Por último, la clasificación por el método de sub-espacio de k-vecindarios cercanos es un proceso estocástico, donde aleatoriamente se selecciona un número de componentes del vector de características dado en la construcción del clasificador. Para este caso en particular el clasificador utilizado es el de k-vecindarios cercanos mostrado en la sección 4.1.4 y del cual solo son seleccionadas las características que tengan contribuciones de distancia diferentes de cero. Esto quiere decir que geométricamente es equivalente a proyectar todos los puntos al sub-espacio seleccionado y los k-vecinos más cercanos se encuentran usando las distancias proyectadas [31], cada vez que se selecciona un nuevo sub-espacio aleatorio, se calcula un nuevo conjunto de k-vecindarios. Por lo tanto los k-vecindarios cercanos seleccionados en cada sub-espacio son tomados como el conjunto que por mayoría vota por la clase a la que pertenecerá cada elemento del conjunto de validación. El proceso puede ser expresado de manera formal como un conjunto de N puntos en un espacio de características n -dimensional, tal que

$$\{(x_1, x_2, \dots, x_n) | x_i \text{ es real para todo } 1 \leq i \leq n\}, \quad (4.10)$$

donde se pueden considerar un sub-espacio m -dimensional donde

$$\{(x_1, x_2, \dots, x_n) | x_i = 1 \text{ para } i \in I, x_i = 0 \text{ para } i \notin I\}, \quad (4.11)$$

donde I es un elemento m del sub-conjunto de $1, 2, \dots, n$ y $m < n$. En cada punto de la validación, los k-vecindarios cercanos entre los puntos proyectados del entrenamiento son usados para encontrar la distancia y así crear una serie de etiquetas para las clases $C = c_1, c_2, \dots, c_k$. Por último, después de p ($p \geq 1$) iteraciones pueden ser asignadas las clases con base a la frecuencia de aparición en la lista C . El algoritmo de clasificación por sub-espacio de KNN resulta ser una buena opción en la resolución de problemas en donde se cuente con un gran número de características, claro ejemplo de esto resulta, ser la clasificación de imágenes.

4.2 Evaluación de los clasificadores

Con la intención de comprobar el desempeño de los sistemas de clasificación enfocados en el reconocimiento de imágenes faciales y basados en la extracción de características por transformaciones wavelet, se consideraron cuatro diferentes pruebas. En la Tabla 4.1 se describen las principales características de cada prueba.

Para las pruebas 1 y 2, fue realizada una clasificación basada en el análisis de las energías calculadas en cada una de las sub-imágenes resultantes una vez aplicada la TWD2D en tres diferentes niveles de transformación (Sección 4.2.1 y 4.2.2) de las imágenes de prueba. Para las pruebas 3 y 4 se realizó una clasificación basada en el valor de los coeficientes extraídos por la TWD2D (Sección 4.2.3 y 4.2.4). Con base a los resultados de las operaciones de pre-procesamiento llevadas a cabo en el Capítulo 2, es evidente que para cada una de las pruebas descritas fueron clasificados los tres rasgos de mayor relevancia en un rostro (ojos, nariz y

Tabla 4.1: Descripción de las pruebas realizadas

Prueba	Descripción	Clases	Predictores			Observaciones
			1° Nivel	2° Nivel	3° Nivel	
1	Análisis por energía	22	4	7	10	208
2	Análisis por energía	22	4	7	10	88
3	Análisis por características extraídas	22	2016	504	126	208
4	Análisis por características extraídas	22	2016	504	126	88

boca) tomando en cuenta también, el área del rostro, debido a que engloba todos los rasgos característicos. Los clasificadores utilizados para todas las pruebas se describieron en la Sección 4.1 y fueron utilizados para llevar a cabo un análisis comparativo entre todos los resultados, para así poder determinar que características y que sistema de clasificación es el mas óptimo en el reconocimiento de imágenes faciales.

4.2.1 Prueba 1: Análisis por energía

De acuerdo a la Sección 3.1.3 la energía contenida en una señal x_o es conservada a través de los distintos niveles de la TWD, tal que, para la TWD2D esta propiedad puede ser expresada por

$$E = \sum_{i=0}^M \sum_{j=1}^N (x_{o,i,j})^2 = \sum_{i=1}^M \sum_{j=0}^N (W_{i,j})^2, \tag{4.12}$$

donde $x_{o,i,j}$ y $W_{i,j}$ son respectivamente cada uno de los elementos de la matriz original y los elementos de la matriz de coeficientes resultantes de la descomposición wavelet [1] de dimensiones $M \times N$. Como ha sido explicado en capítulos anteriores la energía se conserva pero sufre de una redistribución en cada uno de las sub-conjuntos resultantes, por lo que la primer prueba consistió en la extracción de la energía en cada una de las imágenes obtenidas después de ser descompuestas por las diferentes TWD2D en uno, dos y tres niveles de transformación. Este proceso determina cual de las sub-imágenes generadas aporta una mayor contribución de energía. Un punto importante es que para el caso especial de las TWP las sub-imágenes seleccionadas para su estudio fueron las que comprenden el grupo de las aproximaciones o promedios distribuidas por la matriz resultante (Sección 3.4 y 3.6).

En esta pruebas se consideraron las 208 imágenes seleccionadas de la base de datos **BioID** [20] y posteriormente fueron analizadas, para así determinar la existencia de rostros humanos dentro de ellas y segmentar las áreas que comprenden los principales rasgos en una cara humana (Sección 2.1), generando así 4 conjuntos de imágenes a los cuales se les aplicó una etapa de pre-procesamiento (Sección 2.2), dando como resultado tres conjuntos de 208 imágenes cada uno del área de los ojos, nariz y boca con dimensiones de 48×168 pixeles y

un conjunto de 208 imágenes del área del rostro con un tamaño de 168×168 píxeles cada una. Para un primer nivel de transformación el modelo generado es conformado por 22 clases y 208 observaciones en donde las energías de la imagen promedio, de detalles horizontales, verticales y diagonales son usadas como predictores para los 4 rasgos faciales. El modelo para un segundo nivel de transformación comprende los mismos elementos que en el primero, no obstante, los predictores usados comprenden las energías calculadas en las 7 sub-imágenes resultantes. Finalmente el modelo utilizado para un tercer nivel de transformación comprende el uso de las energías calculadas en las 10 sub-imágenes resultantes y usadas nuevamente como predictores. Los resultados obtenidos son divididos en tres partes; (1) análisis de los coeficientes por medio del método de PCA, (2) el análisis comparativo de la exactitud de las máquinas de aprendizaje restantes y (3) comparación en los tiempos de ejecución en los algoritmos de clasificación.

En primer lugar, los resultados obtenidos de aplicar un PCA fueron utilizados como modelos de predicción, es decir que las gráficas obtenidas no representan una clasificación de los datos sino una forma de reducir el espacio de características y así seleccionar que espacio podrá obtener una alto porcentaje de éxito al ser ingresado como entrada en las máquinas de aprendizaje. Debido a la gran cantidad de gráficas obtenidas una vez completado el proceso, en este punto son mostrados solamente los resultados de mayor relevancia y que representan las diferentes variaciones planteadas para esta prueba.

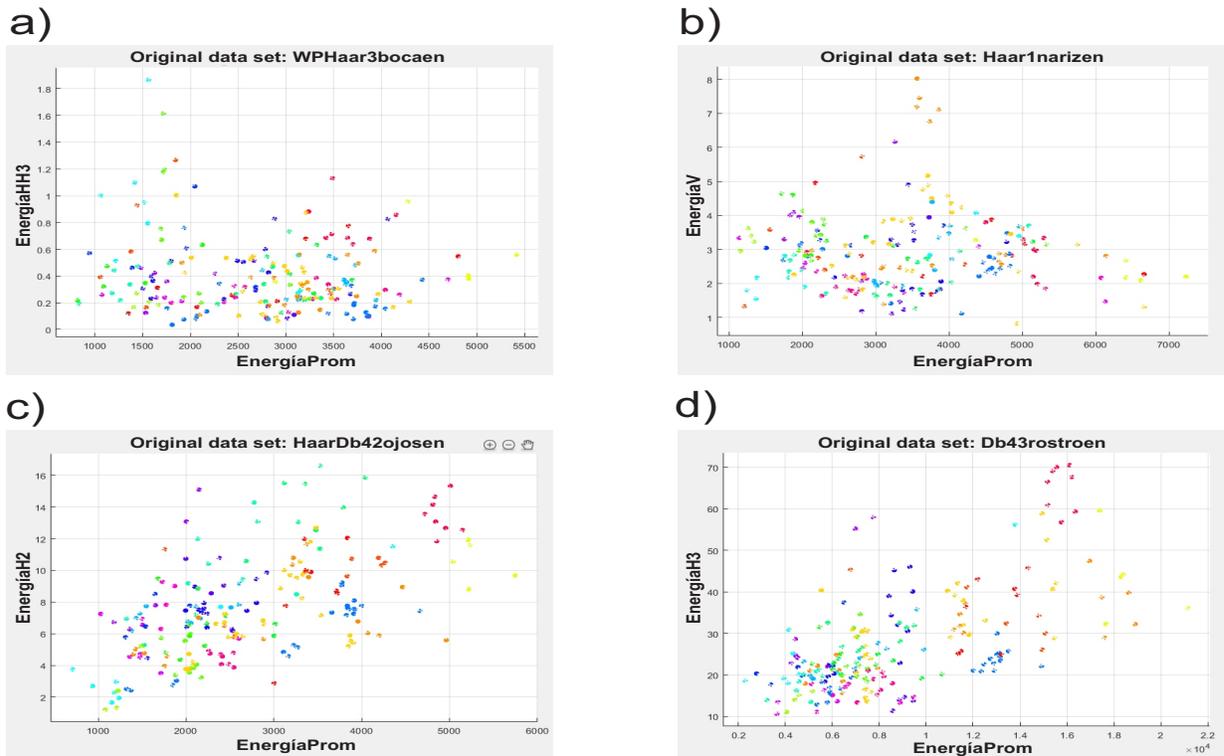


Figura 4.5: Gráficas de dispersión usando PCA del espacio de características generado por a) TW Packet de 3 niveles en el área de la boca, b) TW Haar de un nivel en el área de la nariz, c) TW Haar-DB4 de dos niveles en el área de los ojos y d) TW DB4 de 3 niveles en el área del rostro.

La Figura 4.5 representa los patrones de dispersión de los coeficientes en algunos ejemplos, los cuales ilustran a la perfección los diferentes patrones que pudieron ser encontrados a través de las distintas combinaciones de parámetros en la base de datos. Se puede observar que los datos que se encuentran mezclados y aglomerados en un mismo espacio, por lo que son bastante difíciles de agrupar por clases y por lo tanto es propuesta la idea de que en un espacio de características entre mayor sea la separación entre cada componente, mejor será llevada a cabo la clasificación de los datos. Al observar los valores de las energías en cada una de las imágenes, las versiones promedio son las que mayor energía aportan y por lo tanto es más fácil observar una distinción entre cada clase; las gráficas de dispersión para los detalles son las que menor energía contienen y los valores son más parecidos entre si por lo que se puede decir que son los que pueden presentar mayores errores en los algoritmos de clasificación. No obstante, para el caso de las energías en los detalles horizontales del área de los ojos y boca, resultaron tener una distribución de datos a través del plano de manera dispersa, al igual que para los detalles verticales en el área de la boca y por lo tanto para la experimentación no fue excluido ninguno de los valores de energía debido a que en este punto, no pudo ser comprobado que conjunto de características brinda una aportación positiva o negativa en el sistema de clasificación.

Por otra parte, la exactitud obtenida por las 15 máquinas de aprendizaje descritas en la Sección 4.1, se encuentra en un rango del 0% al 100%. Esto quiere decir que entre mayor sea la exactitud más elementos dentro del conjunto de datos de validación fueron correctamente clasificados. En la Tabla 4.2 son mostrados los parámetros con los que se obtuvieron los mejores resultados y la tasa de éxito en cada uno. Además en las Figuras 4.6 a 4.11 se encuentran ordenados por el tipo de TW utilizada los resultados de todos los clasificadores implementados en las cuatro áreas del rostro analizadas.

Tabla 4.2: Resultados de la Prueba 1

Rasgo	TWD2D	Nivel	Clasificador	Exactitud	Tiempo/segundos
Boca	Daubechies 4	1	KNN Finos	50.0 %	1.2441
	Daubechies 4-Haar	3	Árboles Empaquetados	50.0 %	2.5219
	Haar-Daubechies 4	2	Árboles Empaquetados	50.0 %	2.7234
	Daubechies 4	3	SVM Cúbico	48.1 %	10.333
Nariz	Daubechies 4-Haar	2	KNN Medios	53.8 %	0.27045
	Haar	1	KNN Finos	51.9 %	0.24848
	Haar	1	KNN Ponderados	51.9 %	0.24659
	Haar	3	Árboles Empaquetados	51.9 %	2.5086
Ojos	Haar-Daubechies 4	2	SVM Cúbico	80.8 %	10.399
	Haar-Daubechies 4	3	SVM Cuadrático	76.9 %	10.424
	Haar-Daubechies 4	3	KNN Finos	76.9 %	0.27023
	Haar-Daubechies 4	3	Árboles Empaquetados	76.9 %	2.6247
Rostro	Daubechies 4	3	Árboles Empaquetados	78.8 %	2.4641
	Haar	1	KNN Finos	76.9 %	0.26812
	Daubechies 4-Haar	2	Árboles Empaquetados	73.1 %	2.3796
	Haar-Daubechies 4	2	KNN Finos	71.2 %	0.25864
	Haar-Daubechies 4	3	Árboles Empaquetados	71.2 %	2.5306

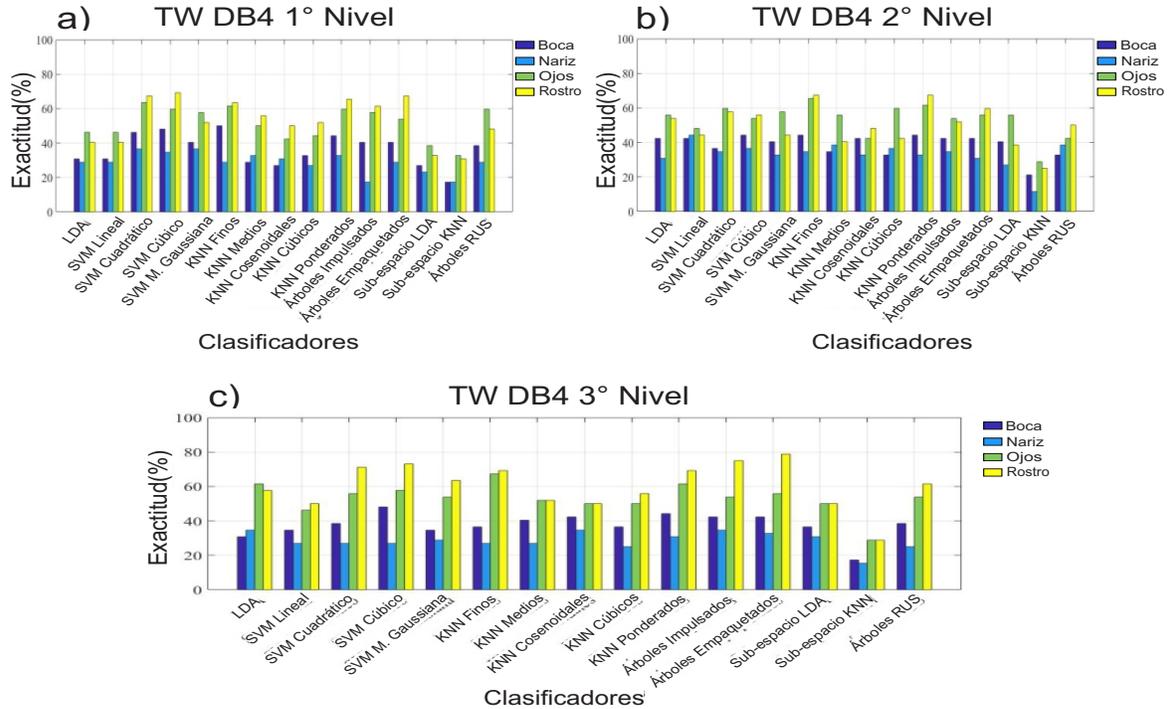


Figura 4.6: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4.

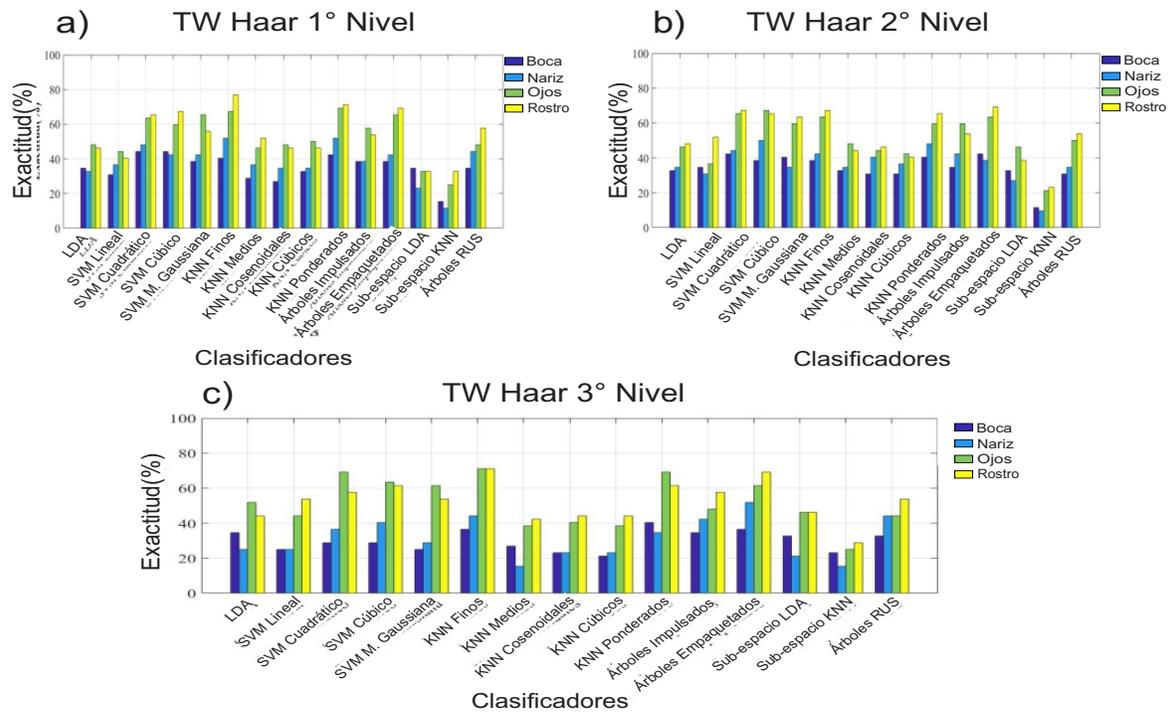


Figura 4.7: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar.

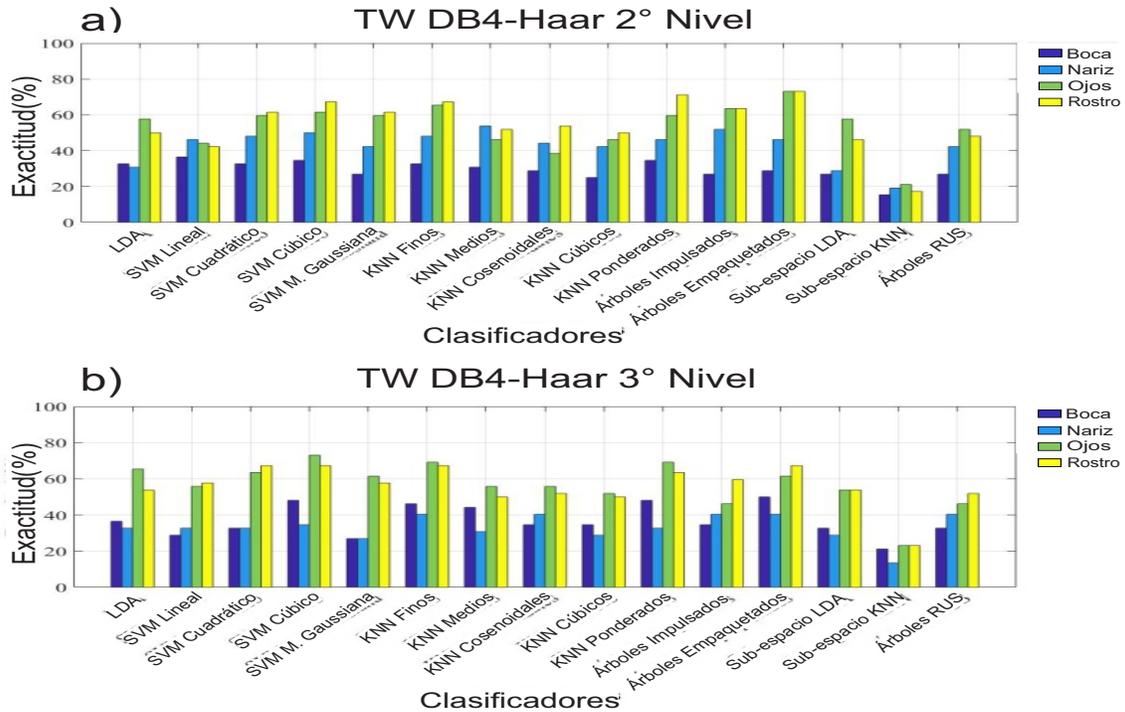


Figura 4.8: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4-Haar.

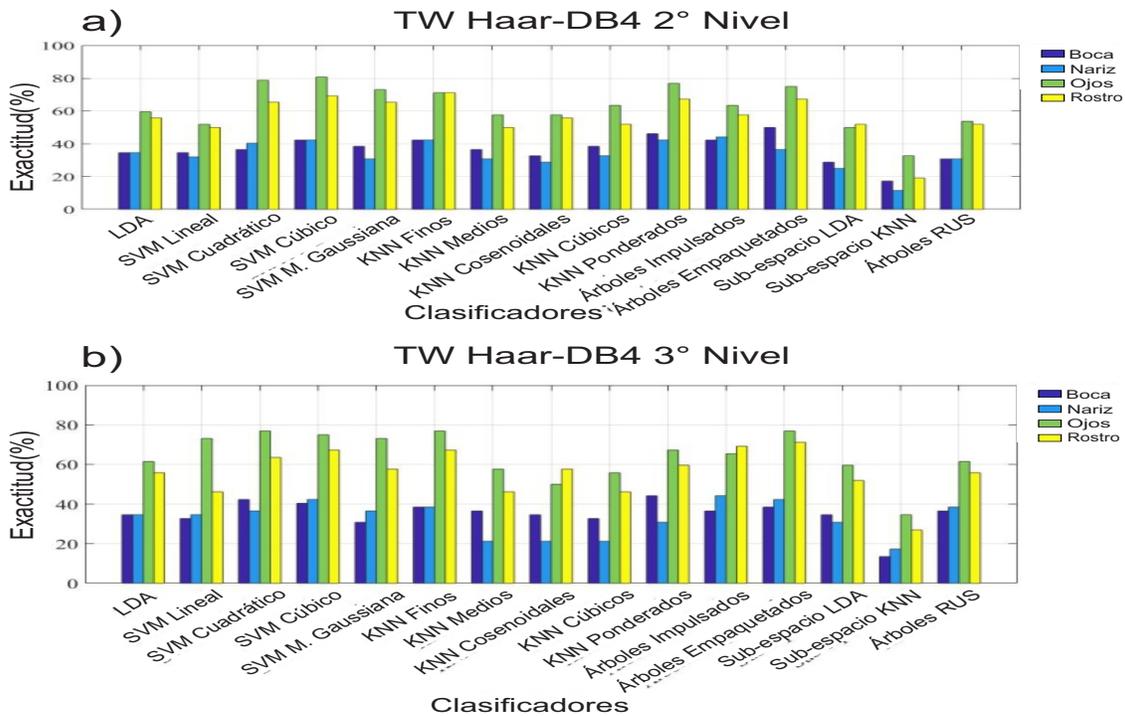


Figura 4.9: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar-DB4.

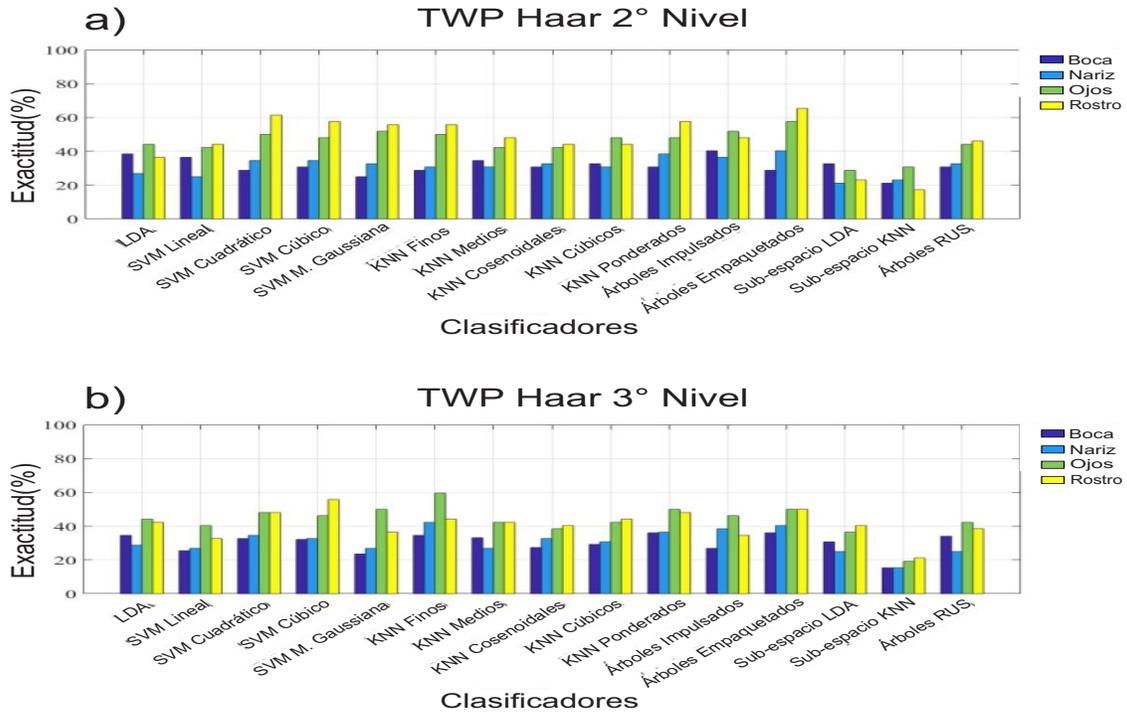


Figura 4.10: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de Haar.

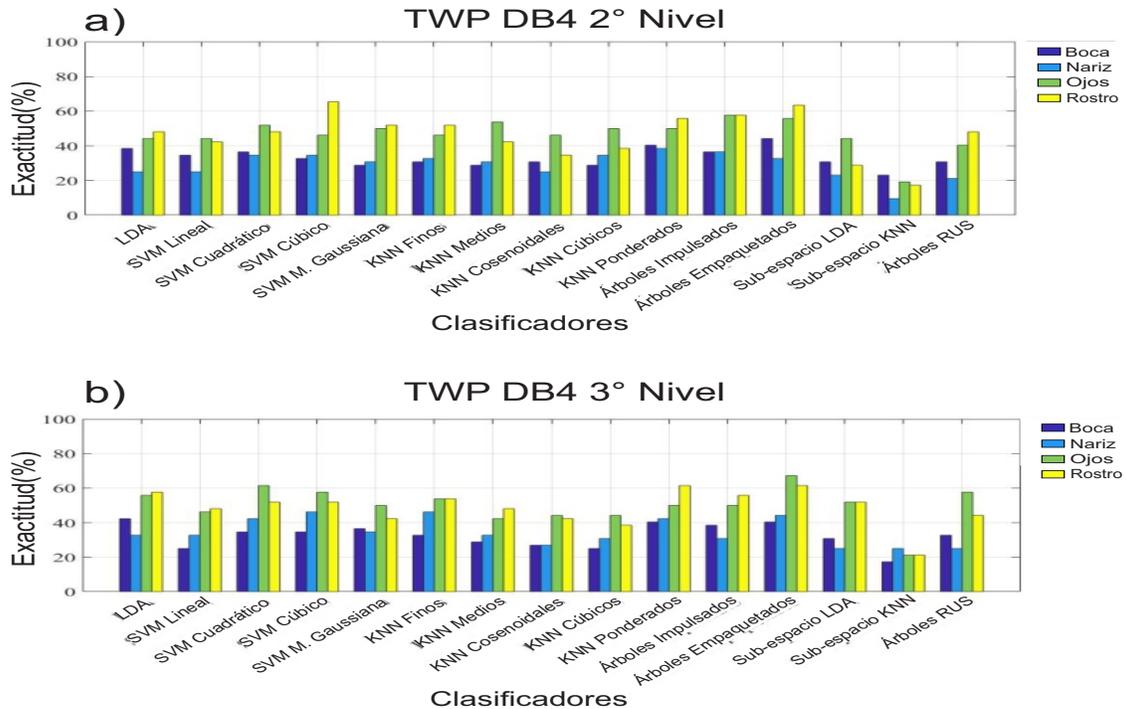


Figura 4.11: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de DB4.

Asimismo, el tiempo de ejecución para cada clasificador es mostrado en la Tabla 4.2. Los resultados obtenidos en esta parte son de gran importancia debido a que gracias a la extracción de características a través de la TWD2D, los tiempos pueden ser disminuidos de acuerdo al nivel de transformación que le sean aplicados a las imágenes, debido a que la cantidad de datos procesados son reducidos en un factor de 2 y por lo tanto se genera un mayor rendimiento en cuanto al costo computacional al efectuar las operaciones correspondientes.

A través de la TWD2D fueron extraídos los valores pertenecientes a la energía contenida en un conjunto de imágenes faciales, posteriormente estos valores fueron utilizados como predictores en diferentes espacios de características y su rendimiento fue medido a través de múltiples sistemas de clasificación. Las gráficas de dispersión generadas por el PCA, resultaron ser un excelente método para predecir el rendimiento que presentara un conjunto de datos antes de ser clasificado, esto fue comprobado gracias a que las gráficas que presentaron una aglomeración de datos en una misma región del plano fueron las que obtuvieron peores resultados y las que generaron patrones dispersos sobresalieron del resto de resultados. Una vez analizadas las gráficas de exactitud para las diferentes combinaciones de parámetros pueden ser destacados dos aspectos importantes; el primero de ellos es que de los cuatro rasgos analizados el área de los ojos y rostro tuvieron un mayor porcentaje de exactitud en comparación con las otras dos (boca y nariz), por lo que estas dos regiones contienen componentes con mayor poder de discriminación y el segundo aspecto es el más importante ya que de acuerdo a los resultados mostrados en la Tabla 4.2 el método propuesto para la extracción de características utilizando una TW Híbrida entre las funciones de Haar y Daubechies 4 en dos y tres niveles fueron las que sin duda obtuvieron los mejores resultados, al obtener un porcentaje de éxito alto para la mayoría de las máquinas de aprendizaje y también de todos los rasgos. Por último, los tiempos de ejecución en todas y cada una de las variaciones de parámetros ejecutados en esta prueba fueron de una proporción mucho menor a la que podría ser esperada si son procesadas las imágenes de la base de datos en su versión original, ya que las dimensiones en cada categoría analizada representan un reducción en un factor de 2 debido a las propiedades de la TW (Capítulo 3), disminuyendo así el costo computacional en los procesos.

4.2.2 Prueba 2: Análisis por energía

La variación en los factores intrínsecos y extrínsecos entre imágenes de rostros humanos puede representar un gran problema, debido a que a pesar de adquirir imágenes de un mismo rostro procurando mantener las condiciones de captura, se encuentran presentes variaciones que pueden no ser observadas a simple vista. Lo anterior quiere decir que algunos de estos factores obstaculizan el proceso de clasificación debido a que los elementos de una misma clase pueden tener componentes demasiado diferentes y con esto aumentar el error en la etapa de validación.

En esta prueba vuelve a ser utilizado un análisis por energía de los coeficientes extraídos en imágenes a través de la TWD2D (Ecuación 4.12), pero con el objetivo de corregir los errores producidos por las variaciones entre imágenes de una misma clase fueron seleccionadas cuatro muestras de cada clase de entre las 208 contenidas en la base **BioID** una vez que fue pre-procesada. Dando como resultado un conjunto de 88 muestras por cada rasgo adquirido, teniendo en cuenta las 22 clases existentes; la selección de muestras fue llevada a cabo

seleccionando imágenes que contuvieran rostros con dirección frontal y centrada así como una iluminación adecuada con la que se pudieran visualizar a la perfección cada uno de los rasgos, además se restringieron las imágenes que presentaran los ojos o boca de manera cerrada. Una vez llevada a cabo la selección de las 88 muestras pudo generarse el espacio de características necesario para la clasificación y poder realizar el análisis comparativo de los resultados.

Las gráficas de dispersión resultantes de aplicar PCA ha demostrado ser un buen auxiliar en el estudio de los espacios de características. En la Figura 4.12 se muestra la comparación entre los resultados obtenidos al ejecutar la Prueba 1 (ver Figura 4.5a) utilizando la TW de Haar en un primer nivel de transformación para el área de la nariz, y los obtenidos al ejecutar la Prueba 2 manteniendo los mismos parámetros del primer caso. Se puede observar una mejor distribución de los datos en los resultados de esta prueba, ver Figura 4.12b. Por tanto es posible esperar un aumento en el porcentaje de éxito al aplicar los procesos de clasificación.

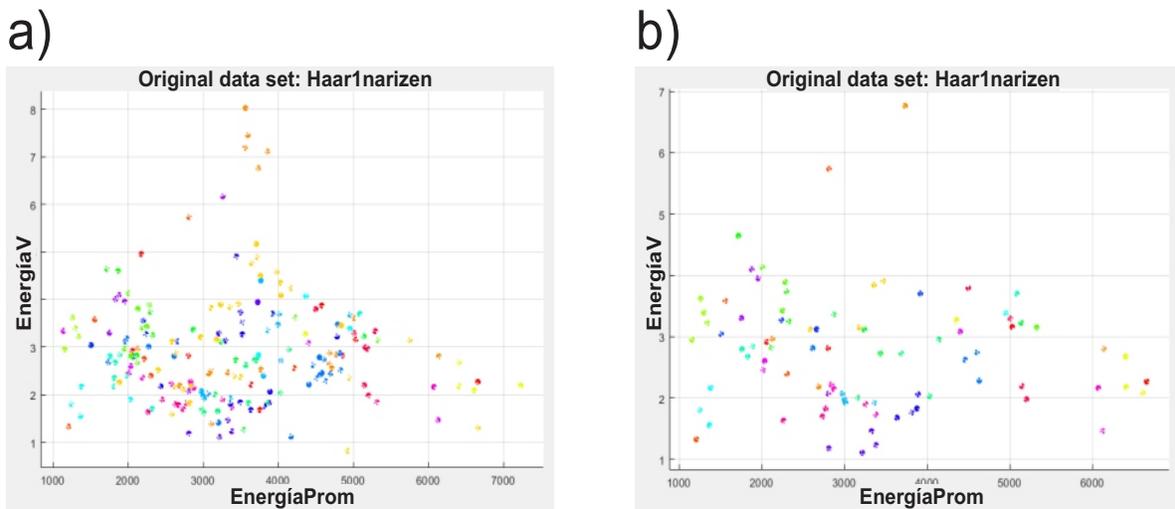


Figura 4.12: Comparación entre las gráficas de dispersión generadas por el PCA de la TW de Haar en un nivel del área de los ojos de a) la Prueba 1 de experimentación y b) de la Prueba 2.

Una vez que fueron aplicados los diferentes sistemas de clasificación para los distintos conjuntos de datos, se llevó a cabo un análisis comparativo entre todos los resultados y los mejores obtenidos son mostrados en la Tabla 4.3, donde pudo ser observado un aumento en la exactitud con la que fueron clasificadas las imágenes. En las Figuras 4.13 a 4.18 se encuentran mostradas las gráficas con los porcentajes de éxito para las experimentaciones de todos los parámetros planteados.

Además, en la Tabla 4.3 se incluyen los tiempos de ejecución para las combinaciones que generaron un mejor desempeño de exactitud. Se pudo observar que los tiempos medidos en todos los casos disminuyó entre cada nivel de transformación aplicado, lo cual se puede atribuir a la menor cantidad de imágenes en total que fueron analizadas.

El tiempo de ejecución calculado en los distintos experimentos, presentó una disminución en comparación a la Prueba 1. No obstante, es evidente que la cantidad de datos procesados

Tabla 4.3: Resultados de la Prueba 2

Rasgo	TWD2D	Nivel	Clasificador	Exactitud	Tiempo/segundos
Boca	Haar-Daubechies 4	2	LDA	63.6 %	0.26038
	Haar-Daubechies 4	2	Sub-espacio LDA	63.6 %	2.234
	Daubechies 4-Haar	3	LDA	59.1 %	0.25932
	Daubechies 4-Haar	3	KNN Ponderados	59.1 %	0.25402
Nariz	Daubechies 4	3	Árboles Impulsados	63.6 %	2.6245
	Haar	3	KNN Finos	63.6 %	1.425
	Daubechies 4-Haar	2	SVM M. Gaussiana	63.6 %	9.8594
	Haar-Daubechies 4	2	SVM M. Gaussiana	63.6 %	9.8439
Ojos	Daubechies 4	2	Sub-espacio LDA	86.4 %	2.2847
	Haar	3	KNN Ponderados	86.4 %	0.24777
	Haar-Daubechies 4	3	Árboles Empaquetados	86.4 %	2.4048
	Haar	2	KNN Finos	81.8 %	0.25112
	Haar-Daubechies 4	2	KNN Ponderados	81.8 %	0.24932
Rostro	Daubechies 4-Haar	3	SVM Cuadrático	90.9 %	10.34
	Packet Daubechies 4	2	SVM M. Gaussiana	90.9 %	10.21
	Haar	2	LDA	86.4 %	0.27223
	Haar-Daubechies 4	2	LDA	86.4 %	0.26291
	Haar-Daubechies 4	3	LDA	86.4 %	0.2609

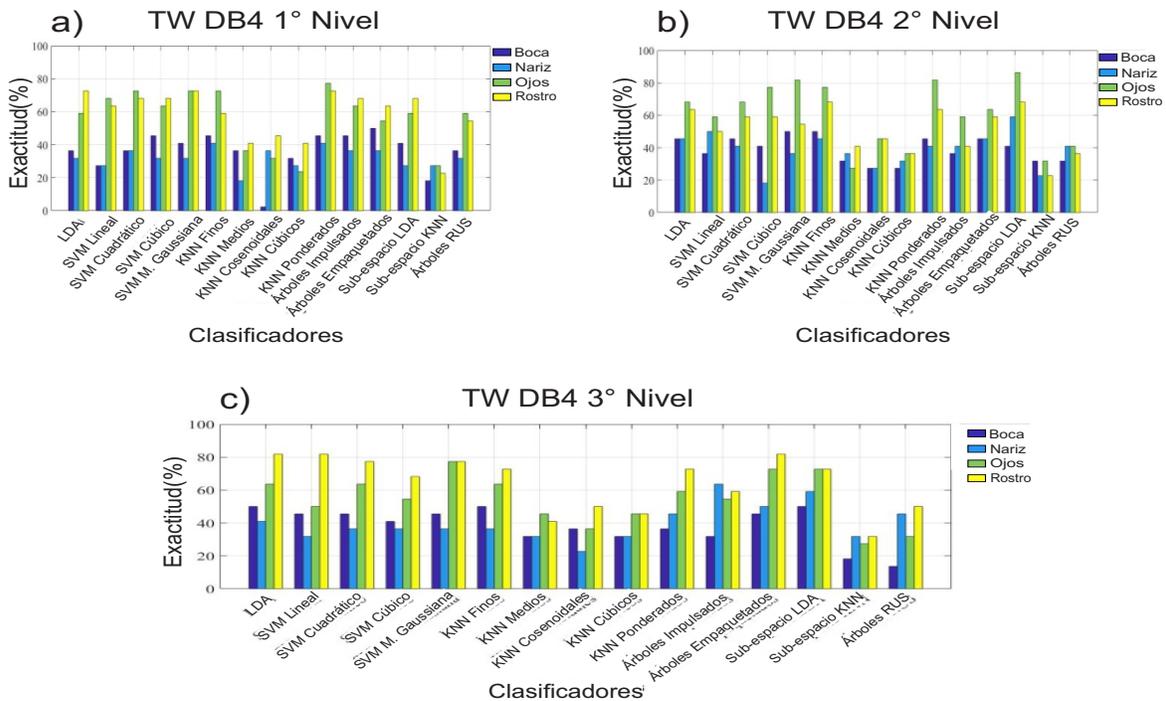


Figura 4.13: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4.

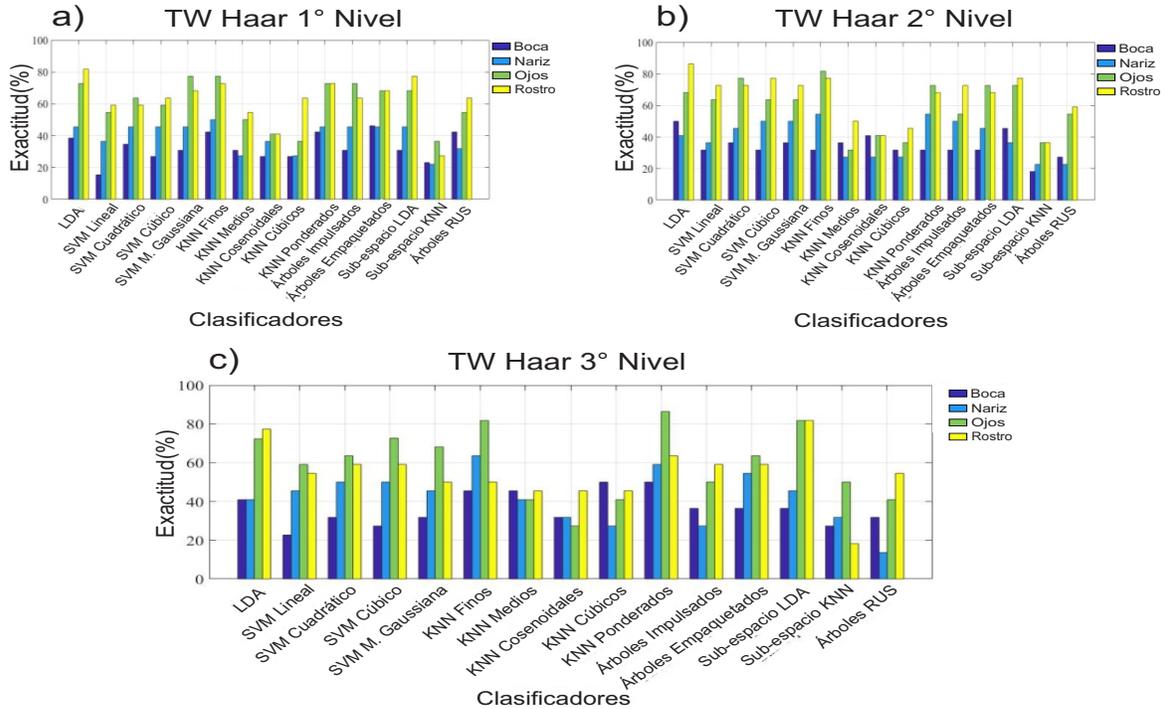


Figura 4.14: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar.

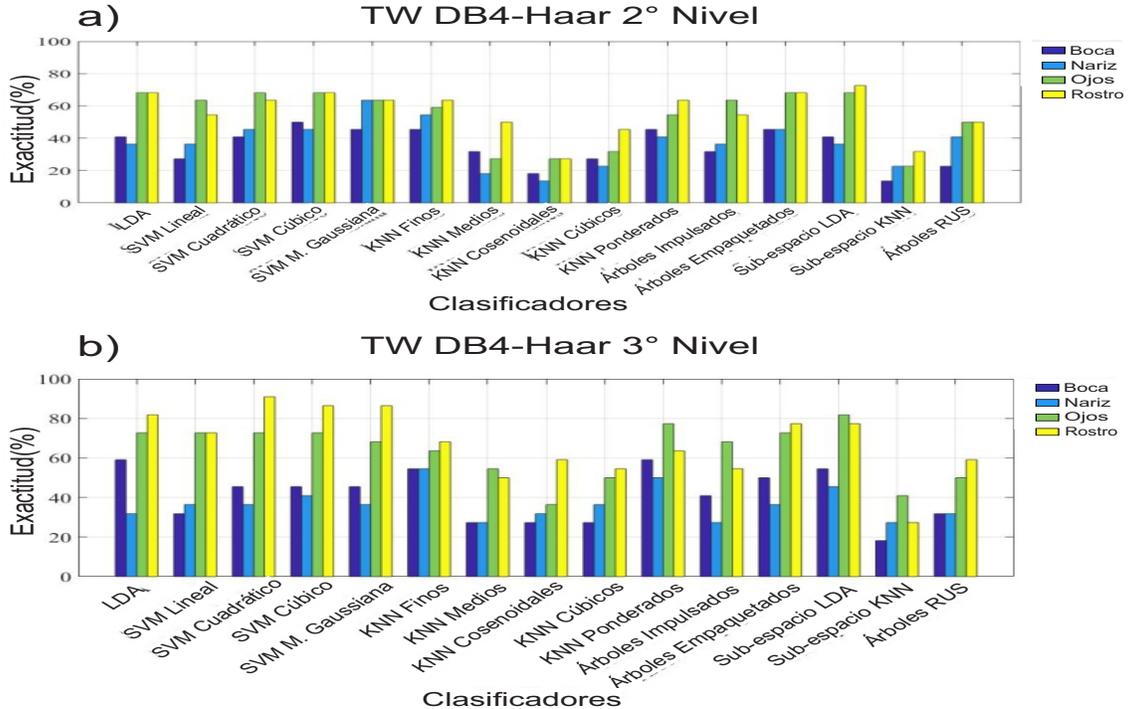


Figura 4.15: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4-Haar.

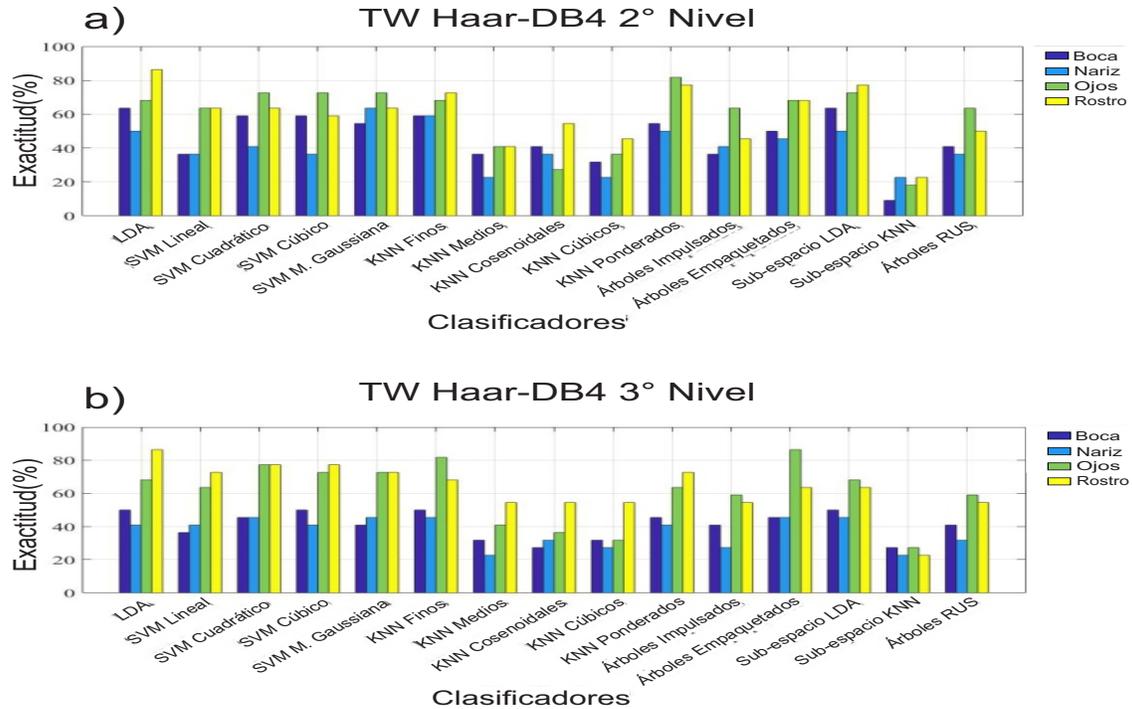


Figura 4.16: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar-DB4.

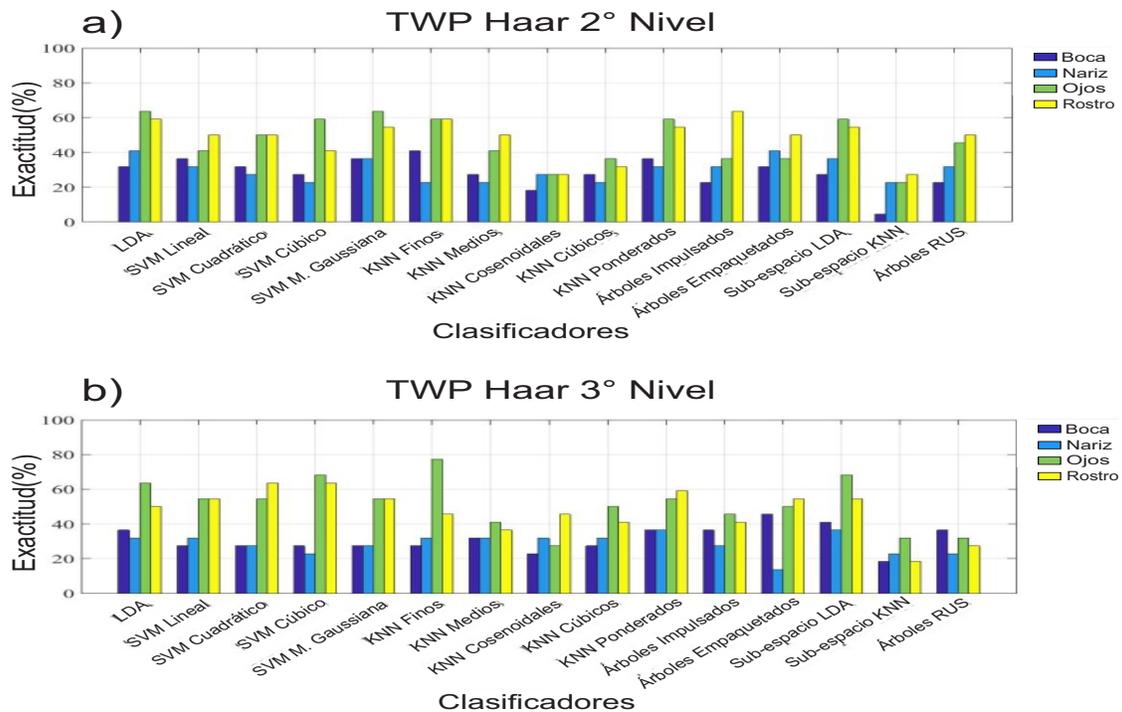


Figura 4.17: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de Haar.

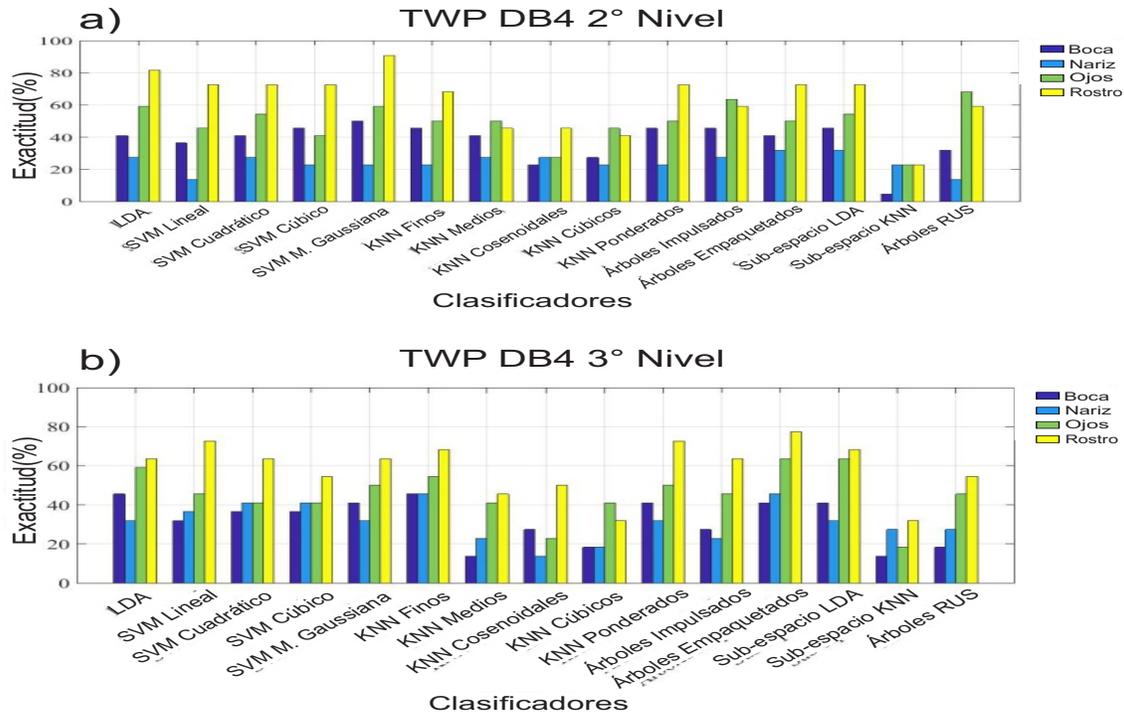


Figura 4.18: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de DB4.

fue mucho menor, por lo que se puede decir que para esta prueba el costo computacional fue disminuido, pero que al igual que en la Prueba 1 puede incrementarse entre mayor sea la cantidad de datos ingresados. Sin embargo, al analizar los resultados de los sistemas de clasificación fue observado un aumento considerable en el porcentaje de exactitud, por lo tanto pueden ser concluidos dos puntos importantes; (1) los factores intrínsecos y extrínsecos en las etapas de adquisición de imágenes son aspectos de suma importancia y que deben ser tomados en cuenta si es que se desea aplicar algún tipo de sistema de clasificación de imágenes faciales ya que ha sido demostrado que los porcentajes de éxito aumentan al tomar esto en cuenta y (2) a pesar de que todos los clasificadores aumentaron su porcentaje de éxito utilizando cualquiera de las combinaciones para la extracción de características, una vez más para el caso de las transformaciones wavelet híbridas los resultados de estas fueron las que destacaron del resto en la mayoría de los clasificadores utilizados demostrando así la efectividad de este nuevo método.

4.2.3 Prueba 3: Análisis de coeficientes extraídos

Los coeficientes extraídos mediante los métodos de TWD2D permiten representar las imágenes de las áreas faciales como un conjunto de sub-imágenes que contienen las versiones promedio y de detalles, dichas sub-imágenes o matrices contienen los coeficientes necesarios para modificar la resolución y resaltar los detalles más significantes de las señales. Para cada nivel de transformación el número de sub-imágenes aumenta de acuerdo a los algoritmos piramidales de descomposición (Sección 3.1.4) de las diferentes TW y el número de coeficientes en ellas disminuye, gracias a esta propiedad es posible analizar no solo los componentes más

relevantes de cada señal sino además disminuir la cantidad de datos en cada una de estas.

Para esta prueba los coeficientes de cada una de las sub-imágenes correspondientes a la versión promedio y de detalles en cada nivel de transformación, fueron utilizados como los predictores necesarios para conformar un espacio de características y con esto poder llevar a cabo la clasificación. Por ejemplo, para un primer nivel de transformación las sub-imágenes corresponden a una matriz de aproximación cuyos coeficientes son re-ordenados en forma de un vector renglón, de igual manera las sub-imágenes de detalles horizontales, verticales y diagonales son ordenadas como vectores renglón y finalmente es generado un espacio de características para cada tipo de sub-imagen resultante. Iterativamente este mismo proceso puede ser aplicado para todos los niveles de transformación posibles en las imágenes de los rasgos faciales y el resultado en el número de predictores es el mostrado en la Tabla 4.1.

Debido a la enorme cantidad de posibles combinaciones que pueden producirse por las diferentes TW en cada uno de sus niveles de descomposición y los cuatro rasgos seleccionados del rostro humano, fueron seleccionadas solo las combinaciones que pudieran producir resultados relevantes y así disminuir la cantidad de experimentaciones. La selección de las muestras fue llevada a cabo de acuerdo a los estudios realizados en las Secciones 4.2.1 y 4.2.2 en donde fue calculada la energía de todas las sub-imágenes resultando ser las versiones promedio las que contuvieron la mayor cantidad de energía para todos los niveles de transformación por lo que representaron ser el conjunto de muestras a ser estudiadas en esta prueba. Sin embargo, a pesar de que el resto de la energía se encuentra distribuida en las versiones de los detalles, para el caso de los detalles horizontales en donde visualmente se resaltan aspectos importantes en el área de los ojos y boca también fueron incluidos en el conjunto de muestras a analizar, al igual que las versiones de detalles verticales para el área de la nariz y las versiones de detalles horizontales y verticales para el área del rostro. El resto de combinaciones entre las versiones de detalles y los rasgos extraídos fueron excluidos debido al poco contenido de energía y a la pobre información visual que brindaron a excepción del caso especial de las TWP, ya que como fue descrito en la Sección 3.4 su principal característica es seguir transformando las versiones de detalles, por lo que para estas transformaciones los detalles verticales y horizontales de cada rasgo forman parte del conjunto de datos utilizados.

El análisis de componentes principales ha resultado ser un algoritmo con el que de acuerdo a las gráficas de dispersión que genera, nos es posible determinar el espacio de características con el que se obtendrán mejores resultados y poder descartar el resto. En la Figura 4.19 son mostradas las gráficas de dispersión generadas por el PCA considerando la TW de Haar y Daubechies 4 aplicadas al área de los ojos en un primer nivel de transformación. La Figura 4.19a muestra una clara distribución de los datos en el plano de de la versión promedio de la TW Db4. Para el conjunto conformado por los coeficientes de detalles horizontales, Figura 4.19b, la mayoría de los datos se agrupa en un mismo espacio y es difícil observar la distinción entre las clases. El mismo problema es observado para los resultados de la TW Haar de las Figuras 4.19c y 4.19d, correspondientes a la versión promedio y de detalles horizontales, respectivamente. Este mismo comportamiento fue comprobado para el resto de las versiones calculadas por todas las TW.

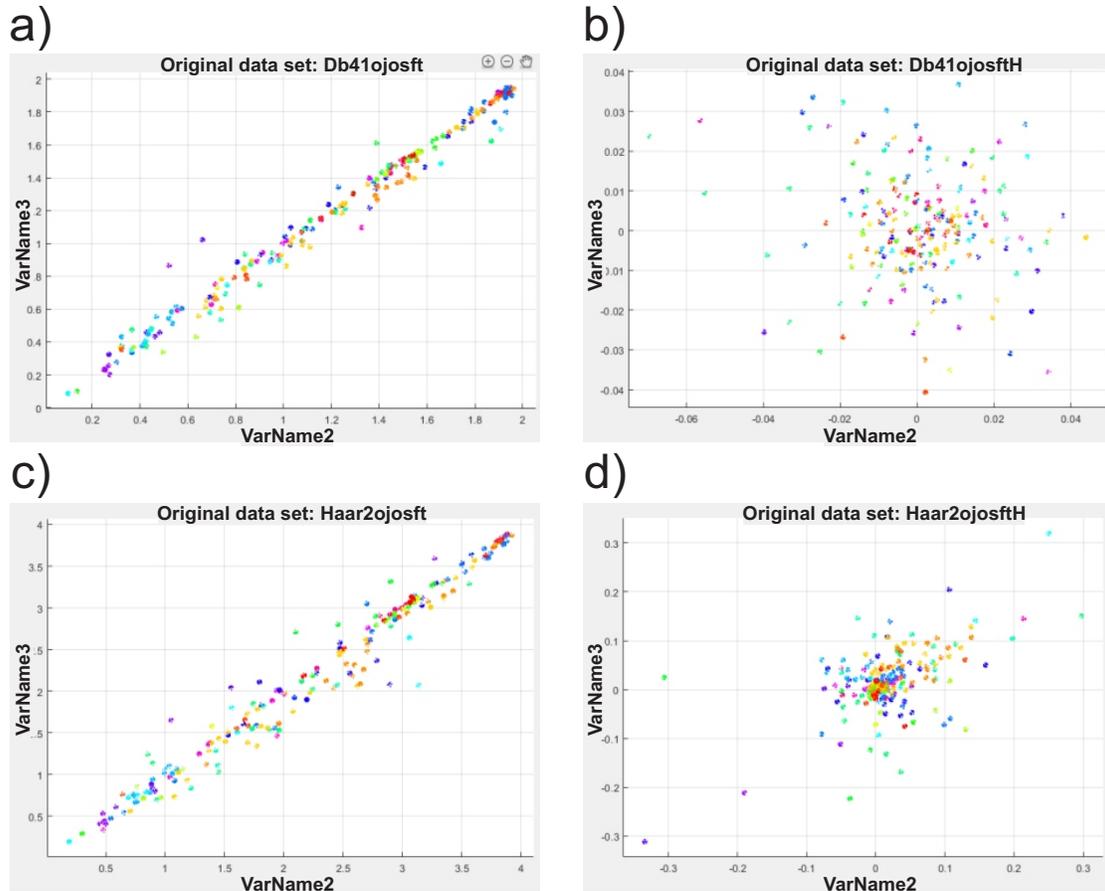


Figura 4.19: Gráficas de dispersión generadas por el PCA de a) la versión promedio de la TW DB4, b) la versión de detalles horizontales de la TW DB4 , c) la versión promedio de la TW Haar y d) la versión de detalles horizontales de la TW Haar, todas en un primer nivel de transformación del área de los ojos.

Una vez llevado a cabo el entrenamiento y validación de los datos por todos los clasificadores, fue realizado un análisis comparativo del porcentaje de éxito para las distintas combinaciones de parámetros en la extracción de características; en la Tabla 4.4 son mostrados los resultados que obtuvieron la exactitud más alta así como su combinación de parámetros. Las Figuras 4.20 a 4.25 contiene las gráficas de exactitud medida para todas las experimentaciones utilizando las versiones promedio de las imágenes.

Los resultados de las experimentaciones donde se utilizaron las versiones de detalles, no forman parte de esta sección debido a que al comparar los resultados obtenidos la tasa de éxito, fue muy por debajo a las obtenidas por las versiones promedio y por lo tanto no son incluidas. Es importante recordar que para el caso de las TWP, las versiones de detalles son las únicas utilizadas, debido a que cualquier versión promedio de estas resulta ser la misma que para una TW que utilice la misma función wavelet y es por eso que los resultados mostrados al utilizar estas transformaciones son incluidos dentro de las mismas gráficas de las Figuras 4.20 a 4.25.

Tabla 4.4: Resultados Prueba 3

Rasgo	TWD2D	Nivel	Clasificador	Exactitud	Tiempo/segundos
Boca	Haar	2	KNN Ponderados	65.4 %	0.15401
	Haar-Daubechies 4	2	Árboles Empaquetados	65.4 %	0.84659
	Haar-Daubechies 4	3	SVM Cuadrático	65.4 %	11.398
	Daubechies 4	1	Árboles Empaquetados	63.5 %	7.4489
	Daubechies 4	3	Sub-espacio LDA	63.5 %	5.169
	Haar	3	SVM Cuadrático	63.5 %	12.093
	Haar	3	KNN Finos	63.5 %	0.32579
	Haar	3	KNN Ponderados	63.5 %	0.33008
Nariz	Daubechies 4	3	Sub-espacio LDA	78.8 %	3.2519
	Daubechies 4	2	Sub-espacio LDA	73.1 %	1.4639
	Daubechies 4-Haar	2	Árboles Empaquetados	73.1 %	0.87391
	Haar	3	Sub-espacio LDA	71.2 %	3.2734
Ojos	Daubechies 4-Haar	2	LDA	98.1 %	0.16504
	Daubechies 4	3	KNN Finos	98.1 %	0.3225
	Daubechies 4	1	LDA	94.2 %	4.7165
	Daubechies 4	1	KNN Finos	94.2 %	3.3501
	Daubechies 4-Haar	3	KNN Finos	94.2 %	0.32148
	Daubechies 4-Haar	3	Sub-espacio LDA	94.2 %	3.339
	Daubechies 4-Haar	3	Sub-espacio KNN	94.2 %	2.1285
	Haar-Daubechies 4	3	Árboles Empaquetados	94.2 %	3.276
	Haar-Daubechies 4	3	Sub-espacio LDA	94.2 %	3.2625
	Packet Haar	3	KNN Finos	94.2 %	0.36254
Rostro	Daubechies 4-Haar	2	KNN Finos	100.0 %	0.46347
	Haar-Daubechies 4	3	Sub-espacio LDA	100.0 %	8.1994
	Haar	2	LDA	98.1 %	0.84073
	Haar	2	Sub-espacio LDA	98.1 %	2.876
	Daubechies 4-Haar	3	LDA	98.1 %	1.0165
	Daubechies 4-Haar	3	Sub-espacio LDA	98.1 %	7.6657
	Haar-Daubechies 4	2	LDA	98.1 %	0.52839
	Haar-Daubechies 4	2	KNN Finos	98.1 %	0.52664
	Haar-Daubechies 4	2	Sub-espacio KNN	98.1 %	1.0561

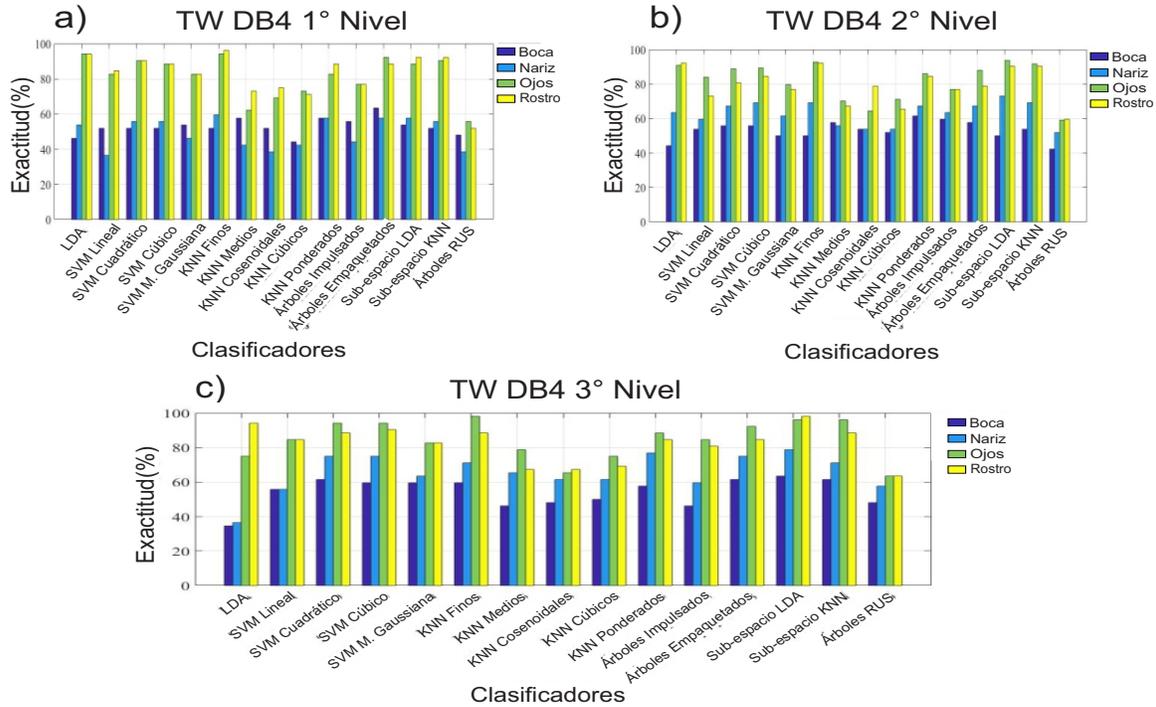


Figura 4.20: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4.

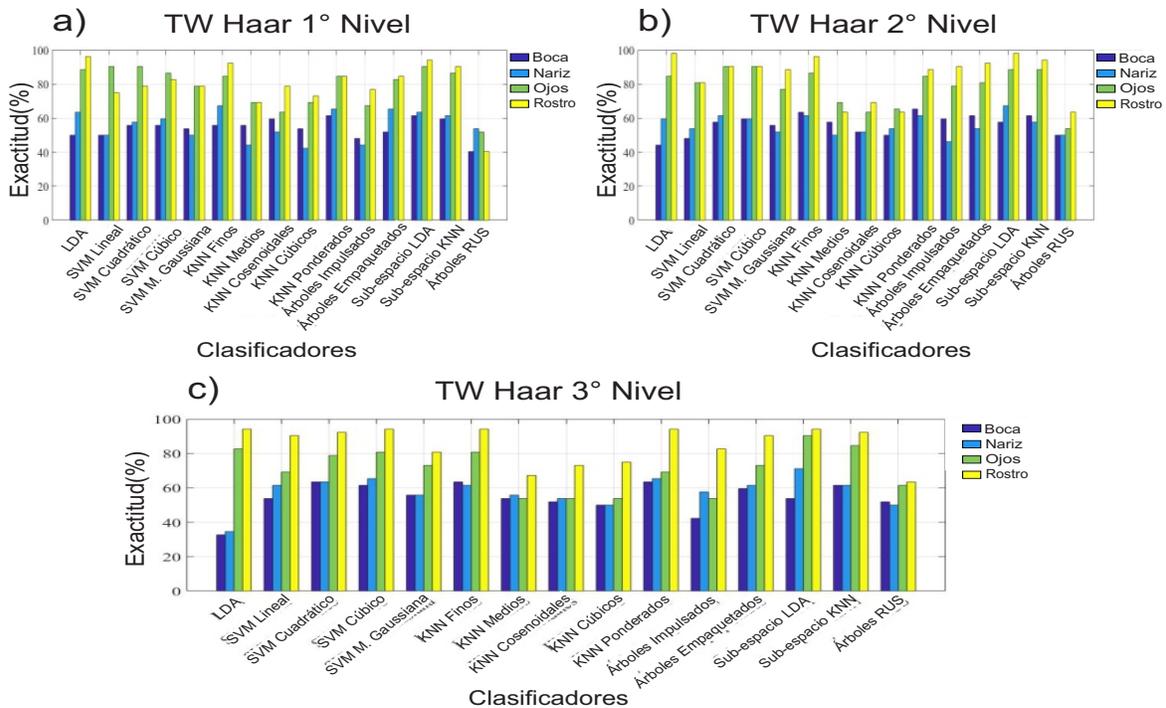


Figura 4.21: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar.

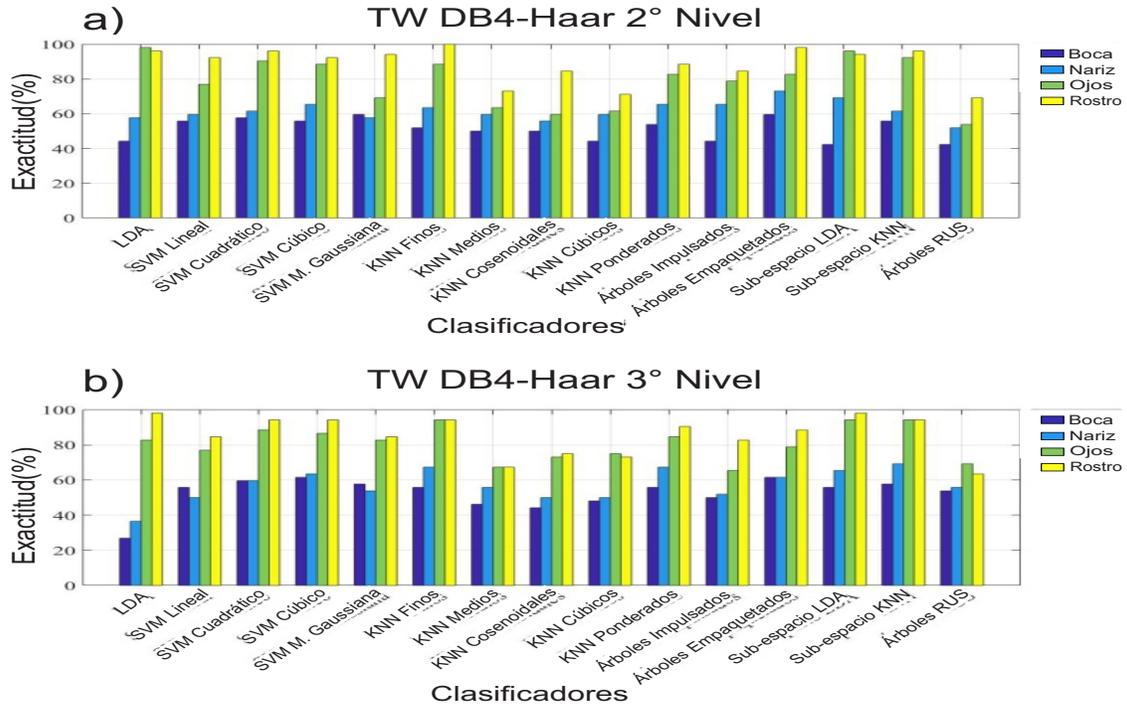


Figura 4.22: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4-Haar.

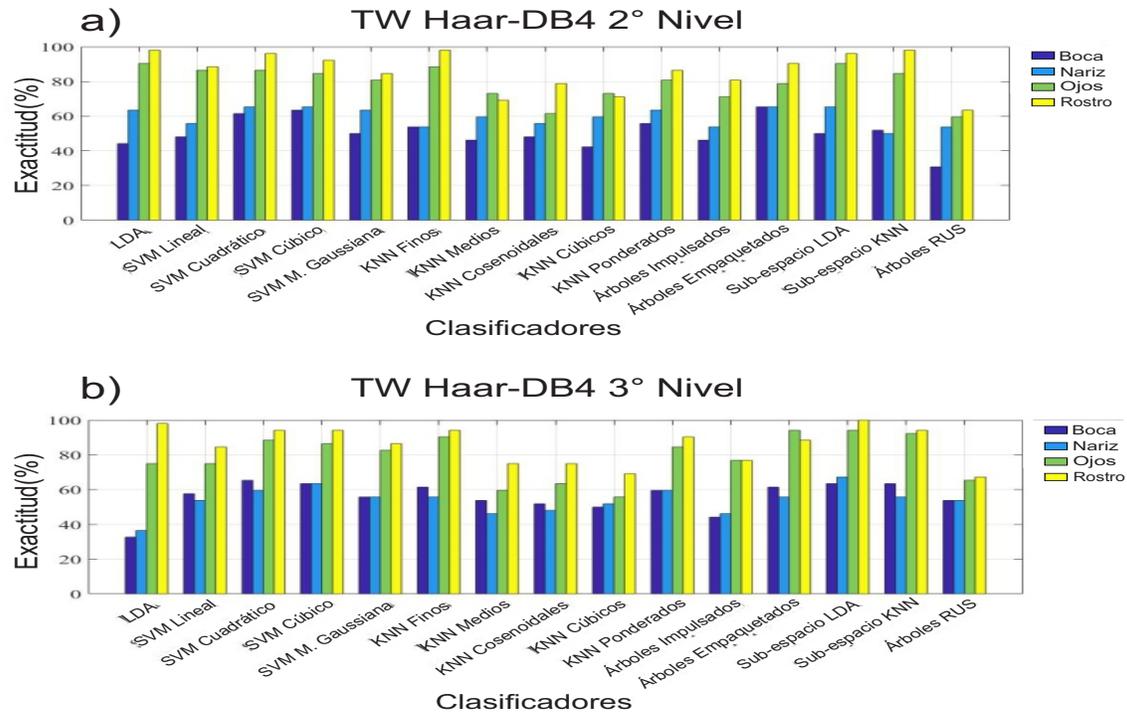


Figura 4.23: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar-DB4.

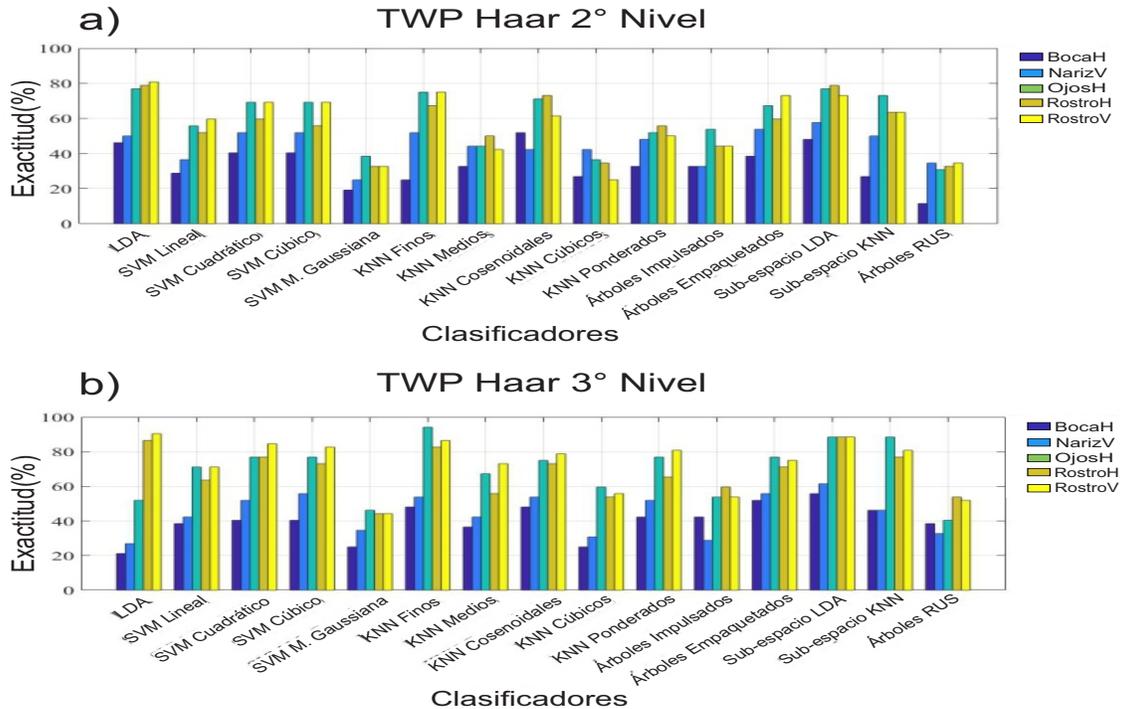


Figura 4.24: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de Haar.

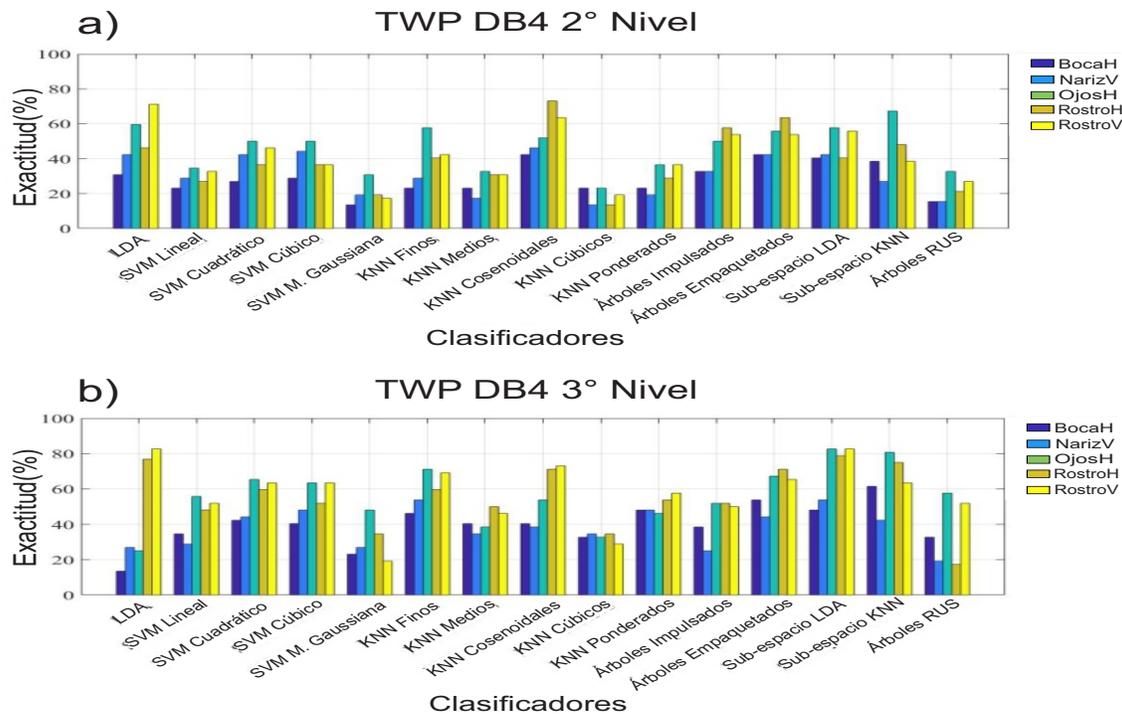


Figura 4.25: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de DB4.

Los resultados obtenidos para los tiempos de ejecución, presentaron una tendencia a disminuir conforme más niveles de transformación le fueron aplicados a las imágenes, los cuales se incluyeron en la Tabla 4.4, en este caso el espacio de características contuvo predictores con un mayor número de valores en comparación a las pruebas 1 y 2, no obstante, los clasificadores implementados son capaces de ejecutar una gran cantidad de operaciones por lo que no se presentó un aumento significativo en los tiempos de ejecución.

Nuevamente puede decirse que el análisis por componentes principales demostró ser un método que no puede dejarse de lado cuando de clasificación de imágenes se trata, este algoritmo permitió una visualización de los componentes y sus distintas distribuciones ayudando de esta manera en la selección de los parámetros necesarios para la clasificación. Lo anterior quedó demostrado gracias a que el porcentaje de éxito medido en los conjuntos que presentaron la distribución de sus datos de manera aglomerada en las gráficas de dispersión resultó tener los resultados más bajos y debido a esto se puede decir que cualquier espacio de características que presente este tipo de distribución no podrá ser clasificado de forma eficiente, dejando un precedente para posteriores experimentaciones.

La tasa de éxito obtenida por todos los clasificadores fue evidentemente superior a la registrada por las pruebas 1 y 2, en el área de los ojos y rostro los resultados fueron los mejores mientras que para el área de la nariz y boca el porcentaje fue el más bajo, no obstante para todos los casos se incrementó la exactitud calculada en comparación a las anteriores experimentaciones. Se puede concluir que el uso de los coeficientes generados por las distintas transformaciones wavelet como predictores en la mayoría de los algoritmos de clasificación brindan una excelente tasa de éxito; y que para el caso específico de imágenes faciales la extracción de características por medio de las TWD2D de la clase híbrida en el área de los ojos y rostro obtienen los mejores resultados para la mayoría de los clasificadores en un tiempo menor al que se podría tener si es utilizado un conjunto de imágenes sin serles aplicado ningún tipo de procesamiento de acuerdo a que el costo computacional en cuestión del tiempo de ejecución de todo el sistema fue disminuido por cada nivel de transformación que le fueron aplicados a las imágenes, ya que los coeficientes resultantes fueron reducidos en un factor de 2 y permitieron tener una menor cantidad de operaciones.

4.2.4 Prueba 4: Análisis de coeficientes extraídos

El porcentaje de éxito obtenido en los clasificadores de imágenes, pudo ser incrementado al utilizar una base de imágenes que cumpliera con los factores intrínsecos y extrínsecos planteados en la Sección 4.2.2. Es por eso, que para esta prueba fue implementado un análisis comparativo de los resultados brindados por los distintos clasificadores, en donde la entrada de estos es conformada por una serie de imágenes, que cumplen los parámetros de adquisición necesarios para reducir el error causado por la diferencia de valores entre cada clase. Una vez que fue seleccionado el conjunto de imágenes, fueron extraídos los coeficientes generados a través de las diferentes TWD2D. Por lo tanto para un total de 88 imágenes de cada rasgo facial el espacio de características fue conformado por 22 clases distintas en donde los predictores para cada clase fueron los coeficientes de las sub-imágenes seleccionadas en cada nivel de transformación utilizando el mismo criterio que en la Sección 4.2.3. En la Figura 4.26 se

encuentran las gráficas de dispersión de las sub-imágenes promedio y de detalles de algunas de las TW de Haar y Daubechies 4, en donde fueron observados los patrones de distribución de los datos y se encontró que para las versiones promedio de cualquiera de los rasgos la distribución fue mas favorable que para las versiones de detalles. Por ejemplo, la versión promedio de los ojos en un primer nivel de transformación utilizando la TW Db4 obtuvo la gráfica de dispersión de la Figura 4.26a. En este caso los datos pueden ser mejor segmentados de acuerdo al análisis llevado a cabo en la Prueba 3, donde la aglomeración de los datos en una misma área del plano presenta mayor porcentaje de error en los algoritmos de clasificación. Caso contrario a los resultados de los detalles horizontales (Figura 4.26b). Una situación similar se puede observar en los resultados ilustrados en las Figuras 4.26c y 4.26d, los cuales corresponden a la versión promedio y de detalles horizontales con la TW de Haar, respectivamente. En general, para toda la base de imágenes se presentó el mismo comportamiento.

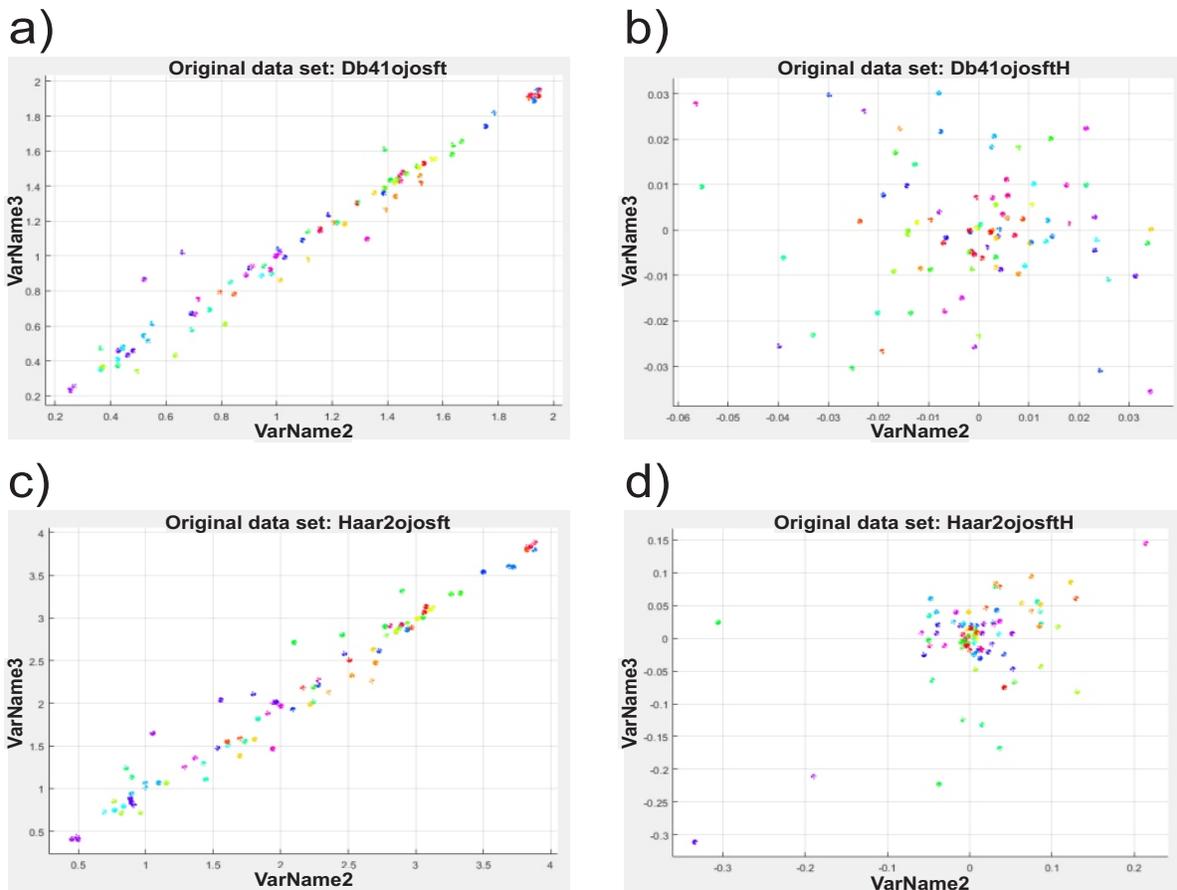


Figura 4.26: Gráficas de dispersión generadas por el PCA de a) la versión promedio de la TW DB4, b) la versión de detalles horizontales de la TW DB4 , c) la versión promedio de la TW Haar y d) la versión de detalles horizontales de la TW Haar, todas en un primer nivel de transformación del área de los ojos.

Los porcentajes de exactitud calculados para cada uno clasificadores, de acuerdo a los parámetros utilizados en cada conjunto de imágenes, fueron analizados y comparados para determinar los clasificadores que mayor tasa de éxito obtuvieron. La Tabla 4.5 incluye

detalladamente los parámetros con los que se obtuvieron los resultados así como el porcentaje de exactitud que se obtuvo. Las gráficas de comparación entre los valores de exactitud y los diferentes clasificadores para cada TW son mostradas en las Figuras 4.27 a 4.32.

Tabla 4.5: Resultados Prueba 4

Rasgo	TWD2D	Nivel	Clasificador	Exactitud	Tiempo/segundos
Boca	Daubechies 4	3	SVM M. Gaussiana	63.6 %	10.514
	Daubechies 4	3	KNN Finos	63.6 %	0.33465
	Haar-Daubechies 4	2	Sub-espacio LDA	59.1 %	0.79155
	Haar-Daubechies 4	2	Sub-espacio KNN	59.1 %	0.53798
Nariz	Daubechies 4	3	Sub-espacio LDA	77.3 %	2.8737
	Haar	1	SVM Cuadrático	68.2 %	4.9999
	Haar	1	KNN Finos	68.2 %	0.56169
	Haar-Daubechies 4	3	KNN Finos	68.2 %	0.075088
	Haar-Daubechies 4	3	Sub-espacio LDA	68.2 %	0.55763
	Haar-Daubechies 4	3	Sub-espacio KNN	68.2 %	0.44465
Ojos	Daubechies 4	2	LDA	100.0 %	0.17354
	Daubechies 4	2	KNN Finos	100.0 %	0.15925
	Daubechies 4	2	Sub-espacio KNN	100.0 %	0.62138
	Haar	2	KNN Finos	100.0 %	0.15024
	Haar	2	KNN Ponderados	100.0 %	0.15053
	Haar	2	Sub-espacio KNN	100.0 %	0.61231
	Haar	3	LDA	100.0 %	0.081319
	Haar-Daubechies 4	3	KNN Finos	100.0 %	0.073786
	Haar-Daubechies 4	3	KNN Ponderados	100.0 %	0.070665
	Haar-Daubechies 4	3	Sub-espacio KNN	100.0 %	0.46259
Rostro	Daubechies 4	2	LDA	100.0 %	0.56182
	Daubechies 4	2	Sub-espacio LDA	100.0 %	1.6276
	Haar	1	LDA	100.0 %	0.63741
	Haar	1	Sub-espacio KNN	100.0 %	4.6088
	Daubechies 4-Haar	2	LDA	100.0 %	0.49593
	Daubechies 4-Haar	2	Sub-espacio LDA	100.0 %	1.5348
	Daubechies 4-Haar	3	LDA	100.0 %	0.13462
	Daubechies 4-Haar	3	KNN Finos	100.0 %	0.12309
	Daubechies 4-Haar	3	Sub-espacio LDA	100.0 %	0.88595
	Haar-Daubechies 4	3	KNN Finos	100.0 %	0.12428
	Haar-Daubechies 4	3	Sub-espacio LDA	100.0 %	0.75068

En la Tabla 4.5 fueron adjuntos los tiempos de ejecución para los clasificadores que presentaron la mayor exactitud y que en general por cada nivel de transformación pudieron ser disminuidos, al reducir el número de coeficientes utilizados en los predictores de los distintos espacios de características.

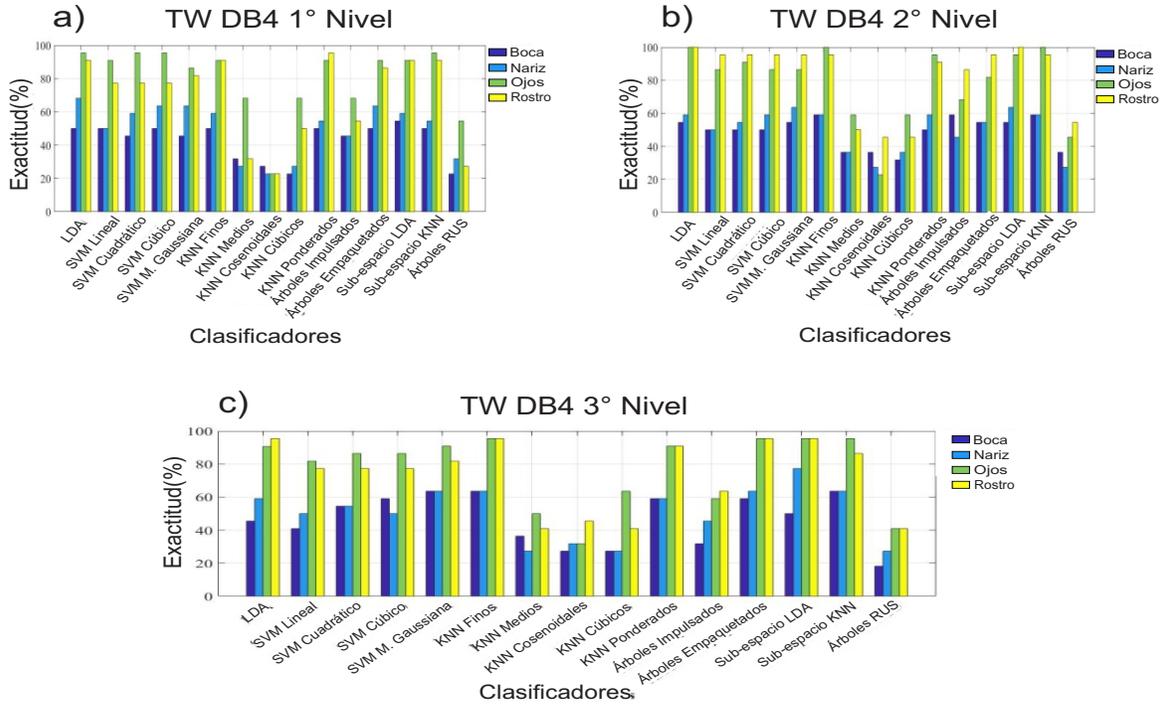


Figura 4.27: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4.

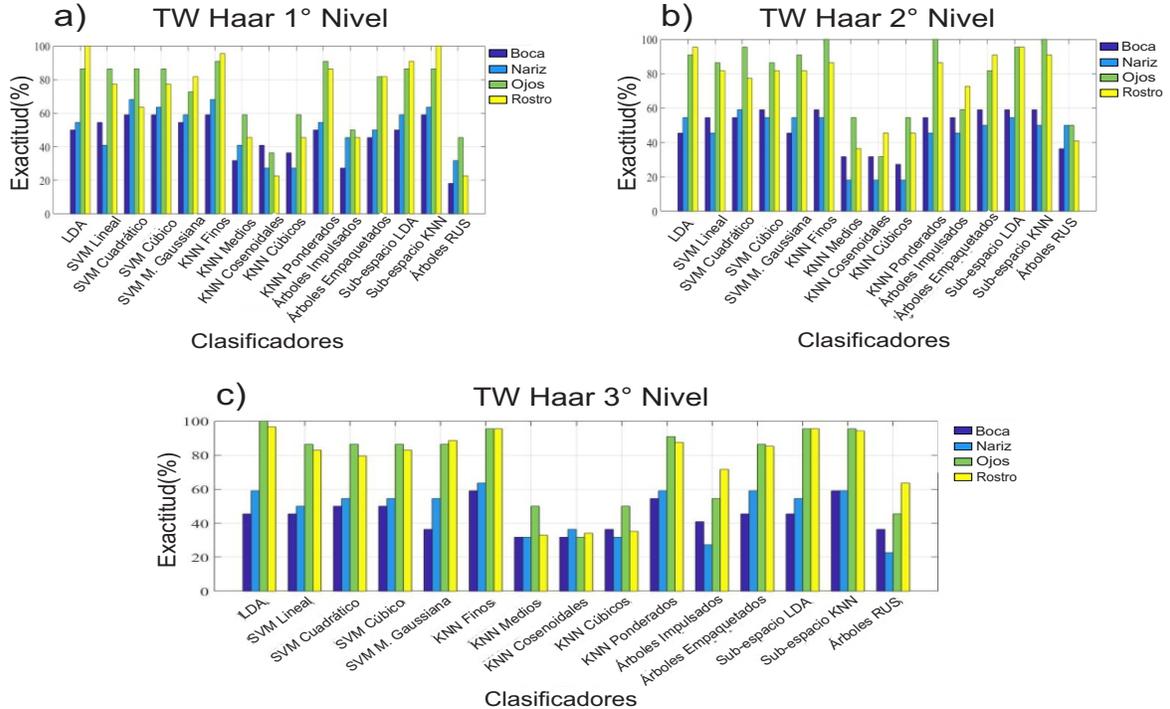


Figura 4.28: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar.

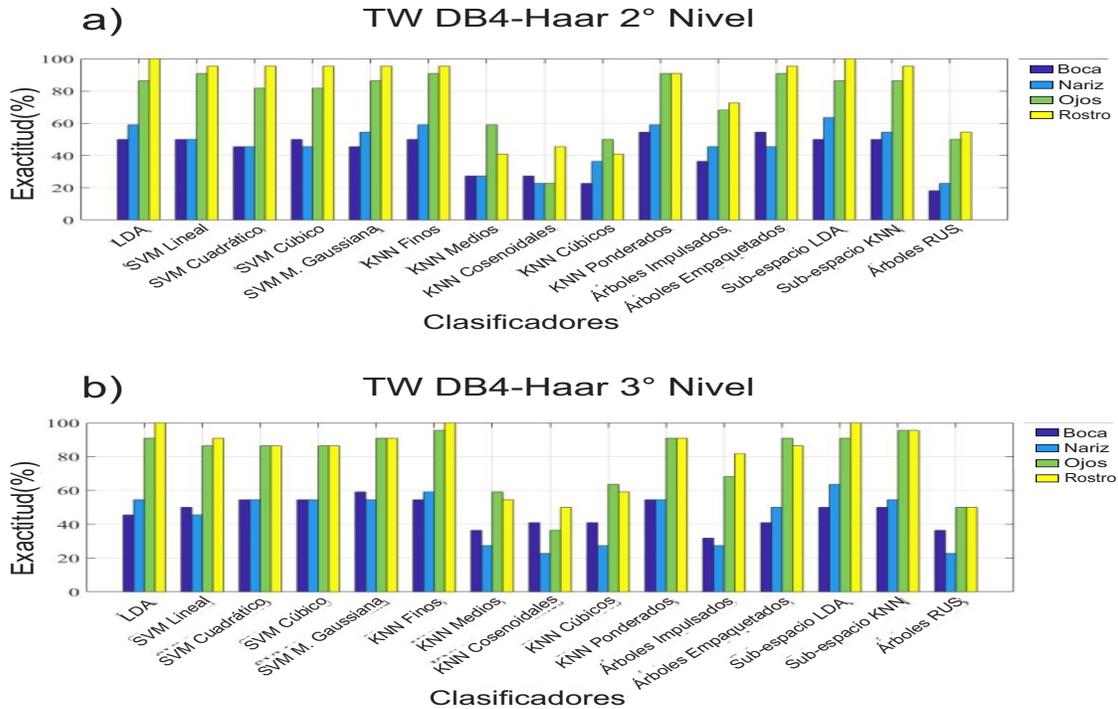


Figura 4.29: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW DB4-Haar.

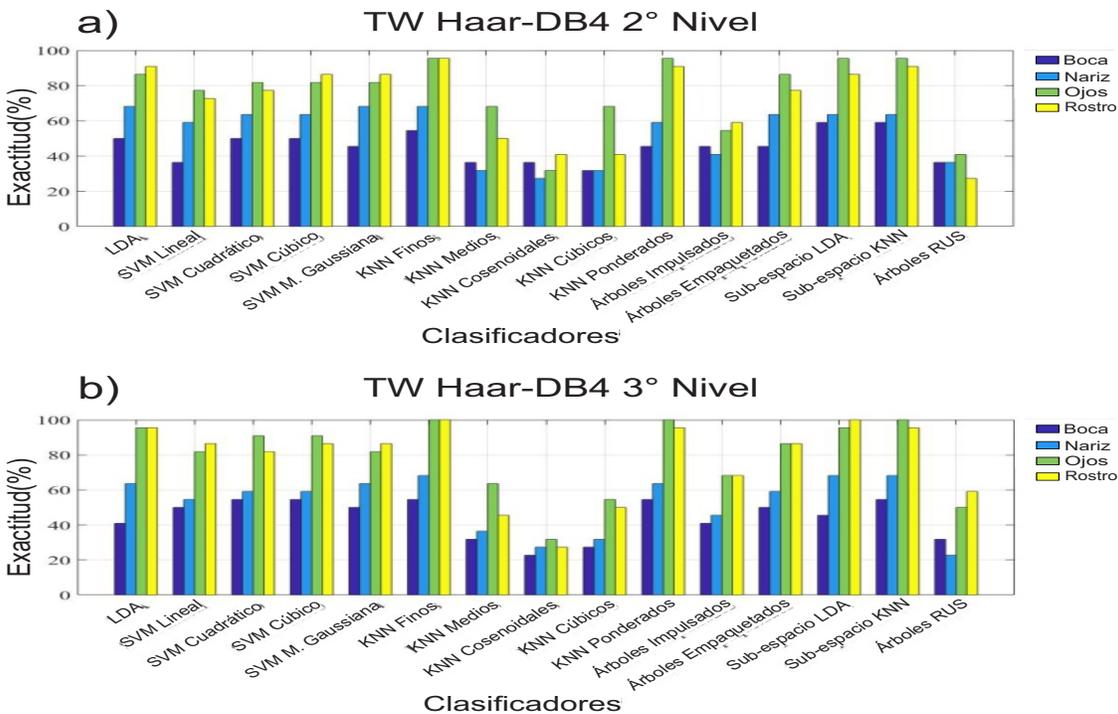


Figura 4.30: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Haar-DB4.

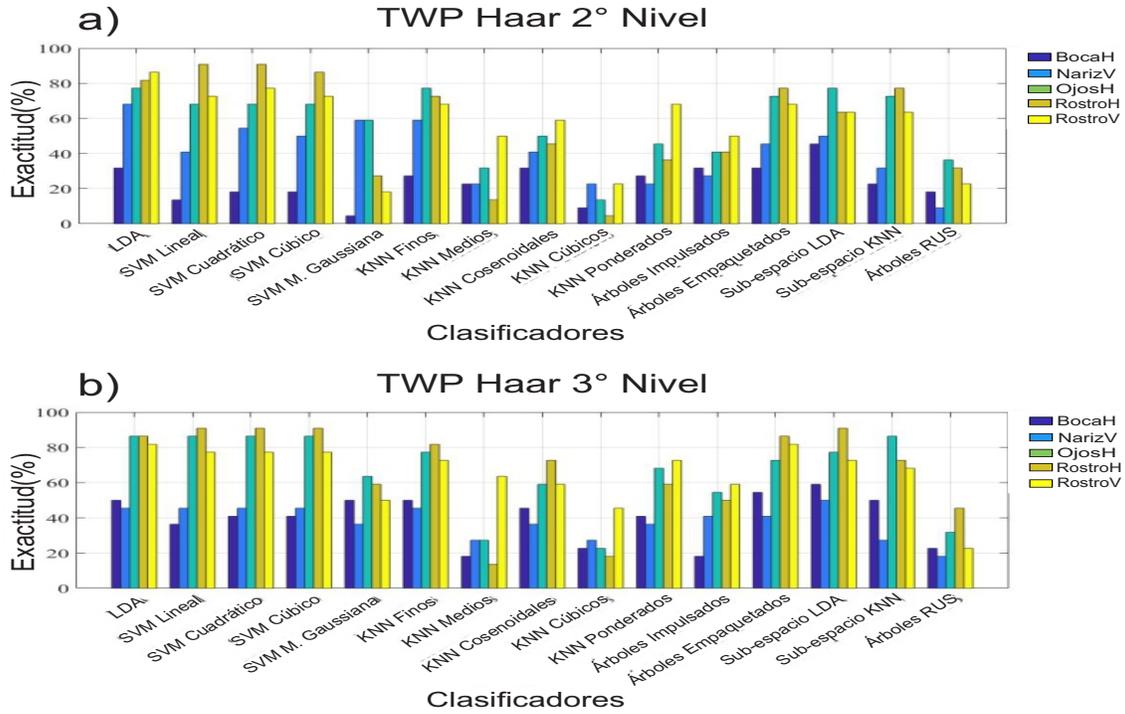


Figura 4.31: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de Haar.

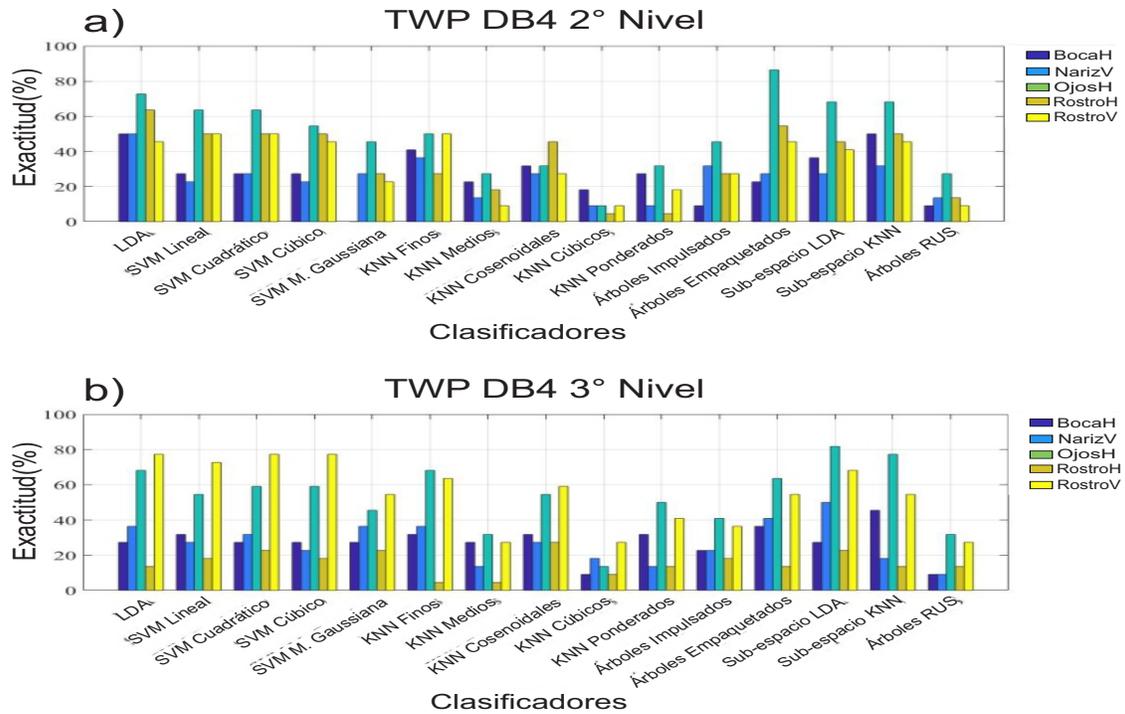


Figura 4.32: Porcentaje de éxito para los diferentes clasificadores en el espacio de características de la TW Packet de DB4.

Uno de los puntos más importantes dentro de la experimentación llevada a cabo, tiene que ver con la gran importancia que la etapa de adquisición puede tomar en los resultados de los algoritmos de clasificación, ya que en comparación con todos y cada uno de los resultados en las pruebas implementadas, este método generó los mayores porcentajes de exactitud en la mayoría de los clasificadores. Por lo tanto, en un sistema de clasificación de imágenes faciales resulta de suma importancia la selección de los datos con base a los factores intrínsecos y extrínsecos así como el método de extracción de características.

La extracción de características por medio de transformaciones wavelet, además de poder disminuir el costo computacional en cualquier tipo de sistema de clasificación, ha proporcionado un conjunto de coeficientes con las características necesarias para poder implementar algoritmos de reconocimiento de alto desempeño, donde el uso de transformaciones wavelet híbridas para la extracción de características demostraron no ser solo un novedoso método si no además de encontrarse presente en la mayoría de los resultados con buenos porcentajes de exactitud en el proceso de clasificación.

CAPÍTULO 5

CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se implementó un sistema de extracción de características en imágenes faciales basado en la transformada wavelet en 2D y algunas de sus variantes. Tal sistema podría ser importante en los sistemas de reconocimiento de personas, además de auxiliar en disminuir el costo computacional de cualquier sistema de clasificación. En este sistema se propuso una Transformación Wavelet Híbrida, la cual fue puesta a prueba bajo un análisis comparativo con algunas de las más conocidas variante de TW. Para esto se implementó un sistema de reconocimiento de personas dividido en tres etapas: (1) etapa de detección de rostros y pre-procesado, (2) extracción de características y (3) clasificación de imágenes. Los resultados obtenidos brindaron la información necesaria para conformar una serie de contribuciones y ventajas de incluir las TWH en la etapa de extracción de características. A continuación se describen tales contribuciones.

1. El uso de las TWD2D en la etapa de extracción de características en los sistemas de reconocimiento. Con dicha herramienta se obtuvieron buenos resultados que nos permitieron reducir el tamaño de las imágenes conservando sus principales componentes. No obstante, la selección de la función wavelet con la que se lleve a cabo la extracción debe garantizar una buena tasa de reconocimiento por lo que la efectividad depende del comportamiento que ofrezca la TW con las propiedades de la imagen. Los resultados obtenidos mostraron claramente que en la mayoría de los algoritmos de clasificación, las TWH sobresalieron del resto, por lo tanto en temas de identificación de personas el método propuesto resultó ser la opción que mayores ventajas aporta.
2. Dentro de este trabajo fueron estudiadas diversas técnicas de procesamiento de imágenes faciales que se pusieron a prueba, nos auxiliaron en la mejora de los resultados al disminuir alteraciones provocadas por las diferentes condiciones en la adquisición de los datos. Los factores intrínsecos y extrínsecos en la etapa de adquisición, demostraron ser de gran importancia en los algoritmos de clasificación debido a que los resultados pueden verse seriamente afectados por las variaciones de estos, por lo que la selección de las imágenes de prueba es un punto que debe tomarse en cuenta si se quiere garantizar una buena tasa de reconocimiento.

3. Con base a la experimentación y resultados presentados, es importante recalcar que para los cuatro rasgos analizados del área de los rostros humanos, la identificación tomando el área de la boca y nariz fueron las que presentaron los más bajos niveles de exactitud al ser clasificadas, mientras que para la de los ojos y el rostro completo la tasa de éxito fue muy buena. Esto lleva a la conclusión que las características brindadas por los ojos y rostro, son las que otorgan la mayor información y por lo tanto las que deben ser seleccionadas para este tipo de tareas. Sin embargo es importante recalcar que los coeficientes totales contenidos en las áreas de los ojos, son mucho menores a los contenidos en los rostros completos por lo que las imágenes del área de los ojos necesitan una menor cantidad de operaciones para ser clasificadas.
4. Algunas de las ventajas encontradas en la implementación de los sistemas de reconocimiento de personas basados en la extracción de características por medio de TWD2D, son en primer lugar los bajos tiempos de ejecución de todo el sistema ya que para cualquier nivel de transformación los datos contenidos en las imágenes de prueba son disminuidos en un factor de 2, lo que quiere decir que el rendimiento del costo computacional puede ser mejorado al minimizar las operaciones necesarias para poder llevar a cabo la identificación; otra ventaja es que el diseño de todo el sistema permite ingresar nuevos elementos y nuevas clases a la base de datos de manera simple ya que los pasos de pre-procesamiento y extracción solo son llevadas a cabo una sola vez para cada imagen por lo que no se requiere volver a repetir el proceso para toda la base de datos.

En las secciones [4.2.1](#) y [4.2.2](#) los resultados obtenidos al clasificar las imágenes con base a la energía calculada en cada una de las sub-imágenes resultantes de aplicar las TWD2D no presentaron un alto porcentaje de exactitud, por lo que para investigaciones futuras se planea efectuar algunas mejoras para este conjunto de pruebas debido a que la información extraída de las versiones de detalles fue muy escasa. Debido a lo anterior se buscará implementar algún algoritmo que distribuya la energía contenida en las versiones promedio a través de todas las demás versiones de detalles y así realzar las características para producir predictores de mejor calidad. Por último como trabajo futuro se planteó la posibilidad de diseñar una máquina de aprendizaje, en la que se puedan ajustar sus parámetros con base a los resultados obtenidos en este trabajo con la intención de optimizar al máximo la tasa de éxito en la identificación de imágenes faciales.

Bibliografía

- [1] Paul S. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, CRC Press, Boca Raton, FL, **2017**.
- [2] Janusz Bobulski, *The Method of User's Identification Using the Fusion of Wavelet Transform and Hidden Markov Models*, Computing, Multimedia and Intelligent Techniques, **2005**, V. 1, pp. 173-180.
- [3] Basma Ammour, Toufik Bouden, Larbi Boubchir, Souad Biad, *Face Identification Using Local and Global Features*, 40th International Conference on Telecommunications and Signal Processing (TSP), **2017**, pp. 784-788.
- [4] Chien J.T. Wu Ch. Ch, *Discriminant Waveletface and Nearest Feature Classifiers for Face Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **2002**, 24(12), pp. 1644-1649.
- [5] Gilbert Strang and Troung Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, **1996**.
- [6] Vinay K. Ingle, John G. Proakis, *Digital Signal Processing using MATLAB*, Cengage Learning, Stamford, CT, **2012**.
- [7] José Salomé Murguía Ibarra. (**1999**). Tratamiento Multiresolución de Señales e Imágenes con Ondeletas de Haar (Tesis de maestría). Universidad Autónoma de San Luis Potosí, S.L.P, México.
- [8] Ingrid Daubechies, *Ten lectures on Wavelets*, SIAM, Philadelphia, PA, **1992**.
- [9] Cecilia Vargas Olmos. (**2010**). Procesamiento de Imágenes con Métodos de Ondeleta (Tesis de maestría). Universidad Autónoma de San Luis Potosí, S.L.P, México.
- [10] Martin Vetterli and Jelena Kovacevic. *Wavelets and subband coding*. Prentice Hall, Englewood Cliffs NJ, **1995**.
- [11] Stéphane Mallat. *A Wavelet tour of signal processing*. Academic Press, San Diego, CA, **1999**.
- [12] James S. Walker. *A Primer on Wavelets and their Scientific Applications*. Chapman and Hall/CRC, Boca Raton, FL, **2008**.

- [13] John Lygeros, Claire Tomlin, Shankar Sastry. *Hybrid Systems: Modelling, Analysis and Control*. <http://www-inst.cs.berkeley.edu/ee291e/sp09/handouts/book.pdf>, **2008**.
- [14] Bernd Jähne, Horst Haußecker, *Computer vision and applications: a guide for students and practitioners*, Academic Press, San Diego, CA, **2000**.
- [15] Janusz Bobulski, *Wavelet Transform in Face Recognition*, Biometrics, Computer Security Systems and Artificial Intelligence Applications, **2006**, pp. 23-29.
- [16] S. Gong, S. J. McKenna, and A. Psarrou, *Dynamic Vision: From Images to Face Recognition*, Imperial College Press (World Scientific Publishing Company), **2000**.
- [17] John C. Russ, J. Christian Russ, *Introduction to image processing and analysis*, CRC press, Boca Raton, FL, **2008**.
- [18] T. Jebara, **May, 1996**, 3D Pose Estimation and Normalization for Face Recognition (Undergraduate Thesis), Center for Intelligent Machines, McGill University.
- [19] Rabia Jafri, Hamid R. Arabnia, *A Survey of Face Recognition Techniques*, Journal of Information Processing Systems, Vol. 5, No.2, **2009**, pp 41-68.
- [20] Base de imágenes BioID, <https://www.bioid.com/facedb/>
- [21] Paul Viola, Michael Jones, *Rapid Object Detection using a Boosted Cascade of Simple Features*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, HI, USA: IEEE Comput. Soc., **2001**, V. 1, pp. 511-518.
- [22] Fatemeh Shahali, Azadeh Nazemi, Zohreh Azimifar, *Single Sample Face Identification Utilizing Sparse Discriminative Multi Manifold Embedding*, Artificial Intelligence and Signal Processing Conference (AISP), **2017**, IEEE, pp. 129-133.
- [23] C. Papageorgiou, M. Oren, T. Poggio, *A general framework for object detection*, International Conference on Computer Vision, **1998**.
- [24] Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, **2012**.
- [25] Stephen Marsland, *Machine Learning an Algorithmic Perspective*, CRC Pres, Boca Raton, FL, **2015**.
- [26] Yang Yinghua, Shi Guoqiang, Shi Xiang, *Fault Monitoring and Classification of Rotating Machine Based on PCA and KNN*, Chinese Control And Decision Conference (CCDC), **2018**.
- [27] D. Opitz, R. Maclin, *Popular ensemble methods: An empirical study*, Journal of Artificial Intelligence Research, HI, **1999**, V. 11, pp. 169-198.
- [28] R. Polikar, *Ensemble based systems in decision making*, IEEE Circuits and Systems Magazine, **2006**, V. 6(3), pp. 21-45.

-
- [29] Li Zhang, Junjun Xiong, Heng Zhao, Xiaohua Zhu, Changzhi Li, *Sleep Stages Classification by CW Doppler Radar Using Bagged Trees Algorithm*, IEEE Radar Conference (Radar-Conf), **2017**, pp. 788-791.
- [30] Wei Li, *Noise-Adjusted Subspace Discriminant Analysis for Hyperspectral Imagery Classification*, IEEE Geoscience and Remote Sensing Letters, **2013**, V. 10, pp. 1374-1378.
- [31] Tin Kam Ho, *Nearest Neighbors in Random Subspaces*, Advances in Pattern Recognition, **1998**, V. 1451, pp. 640-648.
- [32] Klaus D. Toennies, *Guide to Medical Image Analysis*, Springer-Verlag London, **2012**.
- [33] <https://la.mathworks.com/help/stats/choose-a-classifier.html#bunt0ky>