



UNIVERSIDAD AUTÓNOMA METROPOLITANA - IZTAPALAPA
DIRECCIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

SISTEMA NO-SUPERVISADO
DE CLASIFICACIÓN DE ROSTROS
CON TECNICAS BASADAS EN CBIR

Tesis que presenta
Cesar Benavides Alvarez
Para obtener el grado de
Maestro en ciencias y tecnologías de la información

Asesores:

Dra. Graciela Román Alonso
Dr. Juan Villegas Cortez

Jurado calificador

Presidente: Dr. Juan Humberto Sossa Azuela

Secretario: Dr. Miguel Alfonso Castro Gracia

Vocal: Dr. Sergio de los Cobos Silva

Vocal: Dr. Juan Villegas Cortez

México D.F. — 14 de Octubre de 2015

RESUMEN

En este trabajo se propone una metodología para la clasificación de imágenes de rostros. Para esto se usa un conjunto de rasgos descriptivos obtenidos a partir de un conjunto de puntos de interés seleccionados dentro del rostro, usando el clasificador k -NN; se pretende robustecer el análisis de rasgos geométricos por medio de los puntos de interés SURF, con el análisis de características estadísticas. El propósito de este trabajo es verificar qué tan discriminantes son las características de la media, la desviación estándar y la homogeneidad, provenientes de la matriz de coocurrencia, para describir las diferentes características de textura presentes en el rostro. Se propone el uso del algoritmo k -medias con el fin de encontrar racimos (clusters) no correlacionados de las imágenes de los rostros. El trabajo de extracción de características es sumamente costoso en cómputo y tardado, por lo que se propone de la misma manera utilizar el cómputo paralelo, para acelerar la etapa de procesamiento de las imágenes, la extracción de características y obtener resultados en un tiempo menor, ya que se plantea implementar el uso de esta metodología en bases de datos masivas (big data). Finalmente se logró una clasificación de rostros al 100% bajo condiciones controladas y se propuso una arquitectura en paralelo de cómputo escalable con resultados muy satisfactorios.

Descriptores: Clasificación de rostros, puntos de interés, algoritmo k -medias, clasificador k -NN, cómputo paralelo.

Agradecimientos

A mis padres **Alfredo** y **Francisca**, por apoyarme incondicionalmente, por su AMOR, aliento, esperanza y paciencia. Siempre pendientes de mí en todos los aspectos y sobre todo siempre creyeron en mí.

A mis hermanos **Benjamin** y **Alfredo**, que aunque no entienden mucho lo que hago, siempre preguntaban que estaba haciendo y si me faltaba mucho por terminar, por su tolerancia, por no estar con ellos como hubiera querido.

A **Miriam Reyes**, por todo tu apoyo incondicional, sin tus consejos y regaños no lo hubiera logrado, por estar siempre a mi lado durante todo este periodo, no se como agradecer todo lo que me enseñaste para ser lo que soy ahora.

A mis sobrinos **Magdiel** y **Rud**, por interrumpir mis periodos de frustración y distraerme cuando no me salían las cosas.

A mis asesores la **Dra. Graciela Román Alonso (UAM-I)** y el **Dr. Juan Villegas Cortez (UAM-A)**, por haberme transmitido sus conocimientos, por su paciencia, por tomarse un momento de su tiempo para aclarar todas mis dudas, por la bonita amistad que surgió entre nosotros, por que siempre creyeron en mi desde el principio y confiaron en mi para enviarme de estancia corta a España sin pensarlo. De verdad que no tengo palabras para agradecer todo lo que hicieron por mi.

Al **Dr. Carlos Avilés Cruz (UAM-A)**, por las facilidades que me dio para trabajar dentro del laboratorio de Multimedia, de la UAM-Azcapotzalco y el apoyo con este trabajo de tesis, por su amistad y buenos consejos.

A los miembros de comité revisor:

- **Dr. Humberto Sossa Azuela**, por las múltiples revisiones de la tesis y pronta atención en mis errores, por señalar mis debilidades en la escritura y hacer de mi trabajo algo mejor.
- **Dr. Sergio Gerardo de los Cobos Silva**, por aceptar revisar el trabajo y ser uno de mis sinodales.

- **Dr. Miguel Alfonso Castro García**, por sus comentarios, sugerencias y consejos al final de las revisiones, por las enseñanzas durante los cursos de este posgrado.

Al **Dr. Francisco Fernandez de Vega** y al **Dr. Francisco Cháves de la O**, por recibirme en el Centro Universitario de Mérida, Universidad de Extremadura, y darme la oportunidad de trabajar dentro de su laboratorio de investigación, por la confianza que me dieron para apoyarlos dentro de la organización del congreso MAEB 2015, por todos los consejos y la motivación que me dieron para continuar con mis estudios para un doctorado.

A todos mis compañeros de generación, donde surgió una gran amistad.

A mis compañeras de carrera **Areli Rojo y Diana Román**, que siempre están pendientes de mi desarrollo profesional, por su amistad incondicional.

A los compañeros del laboratorio de evolución artificial, **Daniel Lanza y Angeles Vasquez**, por el apoyo que me dieron durante mi estancia en España dentro del CUM y la buena amistad que surgió entre nosotros.

Al **Consejo Nacional de Ciencia y Tecnología (CONACyT)** por la beca otorgada durante mis estudios de maestría.

Finalmente a todas aquellas personas que de alguna forma u otra contribuyeron conmigo e hicieron posible la realización de este trabajo...

“...MUCHAS GRACIAS A TODOS Y CADA UNO DE USTEDES.....”

Índice general

1. Introducción	1
1.1. Objetivos	8
1.1.1. Objetivo general	8
1.1.2. Objetivos específicos	8
2. Estado del arte	9
2.1. El rostro visual	9
2.1.1. Reconocimiento de rostros	10
2.1.2. Automatización del reconocimiento de rostros	14
2.2. CBIR	18
3. Marco teórico	21
3.1. Procesamiento digital de imágenes	21
3.1.1. Definiciones previas	21
3.1.2. Paradigma de la imagen digital	22
3.2. Reconocimiento de patrones	29
3.2.1. Patrones	29
3.2.2. Elementos de un sistema de reconocimiento de patrones	30
4. Metodología propuesta	33
4.1. Arquitectura propuesta	34
4.1.1. Entrenamiento CBIR	35
4.1.2. Entrenamiento SURF	39
4.1.3. Recuperación CBIR	41
4.1.4. Recuperación SURF	43
4.2. Bases de datos	45
5. Arquitectura paralela del clasificador	47
5.1. Paralelización del la versión secuencial	47
6. Resultados experimentales	51
6.1. Pruebas CBIR	54
6.2. Pruebas SURF	63
6.3. Rendimiento del sistema paralelo	69

7. Conclusiones y trabajo futuro	71
7.1. Conclusiones	71
7.2. Trabajo futuro	72
Bibliografía	75
1. Glosario de términos	81
1.0.1. Información estadística de una imagen.	82
1.0.2. Acondicionado de la imagen.	84
1.0.3. Modelos de color	97
2. Cómputo paralelo	101
2.1. Introducción	101
2.1.1. SISD(Single instruction-Single data)	103
2.1.2. Multiprocesadores	103
2.1.3. Multicomputadoras	104

Índice de figuras

1.1. Pasos del sistema CBIR aplicado en escenarios naturales	5
1.2. Metodología tradicional de CBIR en escenarios naturales	6
1.3. Ejemplo de los rostros	7
2.1. Características geométricas	10
2.2. Sistema Bertillon	12
2.3. Cambios del rostro	15
2.4. Diagrama general reconocimiento de rostros	18
3.1. Representación de una imagen digital	22
3.2. Resolución espacial	23
3.3. Resolución	24
3.4. Representación discretización	25
3.5. Sistema DIP	26
3.6. Espectro electromagnético	28
3.7. Elementos de un sistema de reconocimiento de patrones.	30
3.8. Diseño de un sistema de reconocimiento de patrones.	32
4.1. Metodología clásica	34
4.2. Imagen sistema CBIR	35
4.3. Arquitectura de entrenamiento CBIR	36
4.4. Puntos de interés	37
4.5. Generación de ventanas para extracción de rasgos estadísticos	38
4.6. Ejemplo de algoritmo k-Means	39
4.7. Arquitectura de entrenamiento SURF	40
4.8. Imagen dividida en rejillas	41
4.9. Extracción de puntos interés SURF	41
4.10. Arquitectura de recuperación	42
4.11. Ejemplo de algoritmo KNN	44
4.12. Arquitectura de recuperación SURF	45
5.1. Arquitectura paralela del sistema	49
6.1. Metodología de prueba	52

6.2. Imágenes con ruido	53
6.3. Gráficas con diferente tamaño de k y p	55
6.4. Gráfica resultado AR	56
6.5. Gráfica resultado MUCT	57
6.6. Gráfica resultado Yale	58
6.7. Gráfica resultado Yale	59
6.8. Gráfica resultado BioID	60
6.9. Gráficas desempeño CBIR	61
6.10. Gráficas desempeño CBIR	62
6.11. Gráfica resultado AR técnica SURF	63
6.12. Gráfica resultado MUCT técnica SURF	64
6.13. Gráfica resultado Yale técnica SURF	65
6.14. Gráfica resultado Extended Yale B técnica SURF	66
6.15. Gráfica resultado BioID técnica SURF	67
6.16. Resultado obtenido de la técnica CBIR usando imágenes con ruido	68
6.17. Gráficas desempeño CBIR	70
1.1. Análisis de píxeles en vecindad 4	82
1.2. Análisis de píxeles en vecindad 8	82
1.3. Histograma de imagen	83
1.4. Ruido gaussiano	85
1.5. Imagen con tipos de ruido sal y pimienta	86
1.6. Ventana	87
1.7. Ventana deslizante	88
1.8. Filtro promedio	89
1.9. Filtro gaussiano	90
1.10. Aplicación de filtro gaussiano	92
1.11. Filtro contra-harmónico	93
1.12. Filtro promedio geométrico	94
1.13. Filtro promedio-harmónico	94
1.14. Filtro YP	95
1.15. Filtro Laplaciano	95
1.16. Filtro diferenciador	96
1.17. Filtro enfatizador	97
1.18. Cubo RGB	98
1.19. Color HSI	99
2.1. Imagen arquitecturas paralelas	103
2.2. Imagen SISD	103

Capítulo 1

Introducción

Las computadoras a través del tiempo son cada vez más poderosas a la hora de procesar información; esta característica las ha convertido en la herramienta perfecta para tratar de implementar algunas de las capacidades humanas. Dado que la visión es el sentido por el cual nosotros, los humanos, adquirimos la mayor porción de la información que nos es útil para el desarrollo de nuestras actividades diarias. En particular una de las tareas mas sencillas y comunes para nosotros como es la identificación de personas conocidas a través su rostro. El rostro juega uno de los papeles más importantes en nuestra vida diaria, por medio del rostro podemos determinar la identidad de una persona, su sentido del humor, la edad, el sexo, etc. Por este motivo los investigadores de diversas áreas de la computación, buscan desarrollar metodologías para poder implementar esta tarea dentro de una computadora; con esto surge lo que hoy en día se conoce como el reconocimiento de rostros.

El reconocimiento de rostros es un área de investigación activa que abarca diferentes disciplinas, tales como la visión por computadora, el reconocimiento de patrones, el procesamiento digital de imágenes, y el reconocimiento de objetos, ya que podremos ver al rostro como un objeto tridimensional, que puede tener forma, profundidad, textura, etc. Desde 1882 se comenzaron a desarrollar los primeros sistemas de reconocimiento de rostros, siendo el principal de ellos y podría decirse que el padre de los sistemas de reconocimiento de rostros, el llamado *Sistema Bertillon* [19], el cual comenzó a implementar imágenes de rostro de personas, junto con las características antropométricas para la identificación de criminales. Hoy en día los sistemas de reconocimiento de rostros no solo se utilizan para identificar criminales, estos sistemas sirven para más propósitos, e.g. sistemas de seguridad [35], sistemas de identificación de personas [9], localización

[3], entretenimiento (Google, Facebook), etc. Al hablar de reconocimiento de rostros primero debemos mencionar que el rostro¹ se basa fundamentalmente en la distribución de los rasgos físicos como características del mismo, a partir de la percepción visual que tenemos de una persona; es así que la técnica CBIR se tomará de apoyo para mejorar el proceso de clasificación de los rostros, con respecto a los trabajos estudiados en la literatura. Los tejidos blandos de la cara expresan las emociones del individuo, por la acción de los músculos de la mímica. Asimismo, la apariencia facial es un elemento esencial de la identidad de cada ser humano. De los cinco órganos de los sentidos, cuatro están en la cabeza, y más específicamente en la cara; la vista, el oído, el olfato y el gusto. Aunque el quinto sentido, el tacto, también está ampliamente representado, con una zona de altísima sensibilidad en los labios. La sensibilidad en dichos sentidos puede afectar la expresión en un rostro y en el caso de nuestro sistema, disminuir el porcentaje de clasificación esperado.

La tarea del reconocimiento de rostros no es una tarea simple, ya que como se mencionó aun es un área de investigación en donde no hay una única manera de desarrollar un sistema de reconocimiento de rostros, por lo que es un reto abierto, donde se esta trabajando muy duro, una de las ramas de la computación que estudia el reconocimiento de rostros es la inteligencia artificial, principalmente haciendo uso del de aprendizaje automático. Se menciona que, el aprendizaje automático tiene como objetivo desarrollar técnicas que le permitan a las máquinas (computadoras) aprender sobre algún objetivo en específico [27]; en otras palabras, desarrollar programas que sean capaces de generalizar un comportamiento a partir de la información provista en forma de ejemplos. En esta tarea del reconocimiento de rostros también se utiliza la clasificación, ya que para asignar una identidad de una persona es necesario generar clases o agrupar los rostros que sean parecidos y no buscar comparar un rostro contra todos los rostros en una base de datos ya que esto hace que sea una tarea muy difícil cuando se tienen grandes cantidades de rostros.

Actualmente vivimos en un mundo en donde “lo que se mira” determina muchos aspectos de nuestra vida diaria. Por ello el sentido más explotado en los últimos años es el de la vista, para diferentes fines y en el caso de la computación es para la visión por computadora. Las estructuras y propiedades que se quieren deducir en la visión por computadora no solo incluyen propiedades geométricas (forma, tamaño y localización de un objeto), si no también sus propiedades materiales (color, iluminación, composición,

¹“face”. Encyclopedia Britannica. <http://global.britannica.com/>

textura). Estas características nos sirven para determinar de cierto modo la forma en que se generan los grupos o clases dentro de un sistema de clasificación.

Un sistema de clasificación aplicado a imágenes hace uso de las técnicas de la visión por computadora: una de las técnicas más utilizadas en la visión por computadora es el procesamiento digital de imágenes, donde una imagen está definida como la representación visual de un objeto, persona, un paisaje o cualquier otra cosa, capaz de ser captada por el ojo humano a través de diferentes técnicas como, la pintura, el diseño, la fotografía, el video, etc. [39]. Hoy en día, el avance de las tecnologías relacionadas a la proyección, almacenamiento y manejo de imágenes ha crecido de manera exponencial. Esta forma excesiva de generación de imágenes tiene sus aspectos positivos, como puede ser el hecho de que en tan solo un instante, se puede captar una realidad y compartirla con aquellos más próximos (actualmente el mayor contenido de información en la Internet son imágenes, aproximadamente 73% [23]). Los objetos captados por las imágenes pueden ser analizados y clasificados para un uso en específico dentro, uno de los métodos de que pueden detectar objetos a través de sus características geométricas está implementado en la técnica SURF, esta técnica extrae información geométrica de un objeto a través de una serie de puntos, de esta manera si el objeto está parcialmente tapado, rotado, o dentro de una imagen con una resolución diferente, cuando se pasa la imagen con el objeto dentro de esta, el algoritmo SURF localiza los puntos y si existe uno o más puntos del objeto a buscar entonces puede decirse que el objeto está contenido en la imagen de consulta, por lo que el sistema dará un resultado de la consulta.

Generalmente, el objetivo de un sistema de clasificación de rostros es: dada una imagen, en este caso de una cara “desconocida”, o imagen de test, encontrar una clase a la que ésta pertenece en un conjunto de imágenes “conocidas”, o imágenes de entrenamiento. Los sistemas de clasificación se dividen principalmente en dos grupos: los supervisados y los no-supervisados. En los supervisados las características de un conjunto de imágenes ya están organizadas y cuando se consulta una nueva imagen, se comparan sus características contra las características contenidas en la base de entrenamiento, para así ser asignada a una clase correspondiente. Por otro lado en los no supervisados, el conocimiento se genera a partir de los datos en bruto, estos datos se obtienen por medio de la técnica de extracción de características, una vez obtenida la información, esta se trata de agrupar a través de características similares, uno de los algoritmos de agrupamiento no supervisado más utilizados es el k -medias[34], en donde se generan k racimos, y se busca que estos racimos estén lo más separados posible, me-

diante una serie de iteraciones en donde los centros de cada grupo se mueven en cada iteración hasta que se llega a una distancia mínima o la iteración máxima.

La mayoría de los artículos revisados en la literatura, no mencionan de manera clara si realizan una clasificación o no, por lo que para nosotros es una oportunidad de definir una metodología de clasificación de rostros y que sirva de base para el desarrollo de otro tipo de sistemas, no solo aplicado a los rostros, ya que esta metodología fue tomada de base de un análisis a escenarios naturales. En el análisis de escenarios naturales toman como característica principal analizar las características propias de la imagen, por medio de la textura.

La textura es una propiedad innata de todas las superficies, por ese motivo cuando se desea detectar la presencia de un objeto éste puede identificarse por sus características de textura la cual puede ser evaluada como fina, suave, gruesa o tosca según corresponda [25]. La textura se ha utilizado en diversas áreas como la ingeniería, el arte y el diseño para identificar y describir diferentes objetos que los caracterizan, tales como agua, vegetación, montañas, enfermedades o detección de fallas, ya sea en suelos o en cualquier superficie, es por ello que en cada área tiene un significado diferente y esto lleva a que existan diversas definiciones de la misma. En la literatura relativa al procesamiento de imágenes se puede encontrar una gran cantidad de definiciones de este concepto, como se ve en [36] donde se describe a la textura como un arreglo de píxeles cuya relación es la variación espacial de los tonos de grises.

Una de las técnicas que hacen uso del análisis de textura es la recuperación de imágenes mediante imágenes, también llamada “por contenido” (Content-Based Image Retrieval o CBIR). CBIR es una metodología de base estadística para organizar y recuperar imágenes mediante la combinación de la extracción de características obtenidas directamente de la propia imagen, sin anotaciones manuales y sin supervisión. En los inicios de los sistemas CBIR se utilizaban metadatos que consisten en palabras o frases cortas que tratan de definir lo que contiene una imagen. La inclusión de estos metadatos se realizaba manualmente y se usaba como índice y parámetro de recuperación. En la actualidad este proceso es muy costoso ya que se debe indexar la información para poder recuperarla de forma rápida y efectiva; es por eso que los metadatos dejan de tener sentido para colecciones grandes de imágenes y toman fuerza los sistemas automáticos de indexación y recuperación, basados en lo que contiene la imagen. Un ejemplo del funcionamiento de la técnica CBIR se muestra en [23], donde la técnica se aplica exitosamente a imágenes de escenarios naturales.

La técnica CBIR vista en [23], analiza diferentes tipos de escenarios y consta de tres

partes principales Figura 1.1:

1. A la imagen de consulta se le extraen las características estadísticas CBIR.
2. Las características estadísticas de la imagen de consulta son comparadas con las características estadísticas de toda la base de datos ya entrenada y agrupada en clases.
3. Por último se muestran las N imágenes más parecidas a la imagen de consulta, respecto a las características estadísticas CBIR que se analizaron.

La metodología CBIR ha demostrado tener excelentes resultados, en el análisis de escenarios naturales. Esta tarea normalmente se realiza en 4 etapas, como se muestra en la Figura 1.2, estas 4 etapas son los pasos generales que siguen las metodologías de detección y clasificación de escenarios naturales [38]:

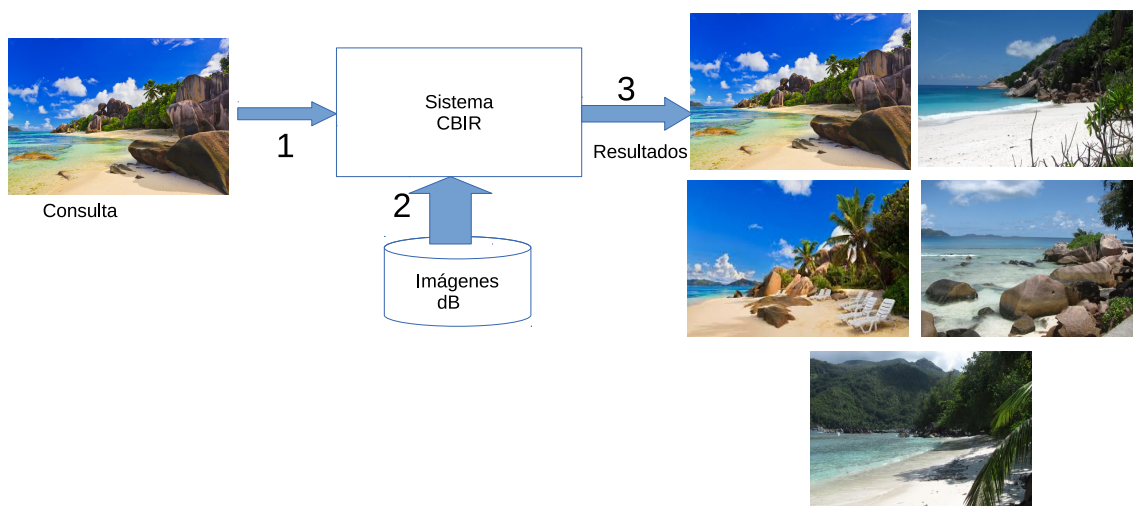


Figura 1.1: Pasos del sistema CBIR aplicado en escenarios naturales.

1. Acondicionamiento de los datos/Pre-procesamiento.
2. Extracción de características.
3. Organización de características
4. Clasificación.

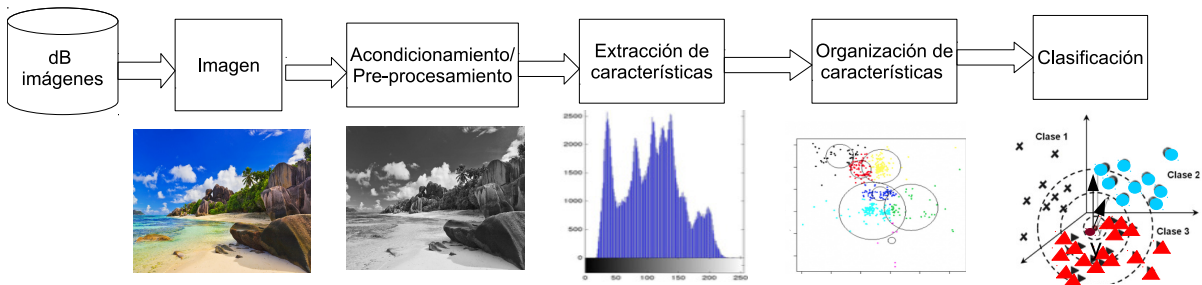


Figura 1.2: Metodología tradicional de CBIR en escenarios naturales.

CBIR aplicada a los escenarios naturales funciona muy bien como se ve en la literatura, pero si se tratara de clasificar un solo escenario natural y decidir si es un lugar en específico, podría ser más interesante ya que entonces no se buscaría separar clases diferentes, si no una misma clase en diferentes lugares, como por ejemplo, tipos de playa, tipos de bosques, tipos de montañas etc. Sería un reto mayor la aplicación de la técnica CBIR a este tipo de análisis.

La implementación de la técnica CBIR en el análisis de imágenes es una tarea muy costosa y tardada, ya que analizar una cierta cantidad de puntos de interés en cada una de las capas que componen la imagen hace sumamente pesado el procesamiento en un proceso secuencial, por otro lado si el número de imágenes contenidas en la base de datos es muy grande, la etapa de entrenamiento tardaría demasiado y esto en una aplicación de la vida real no sería muy eficiente, por lo que se asume necesario el uso del cómputo paralelo en las etapas donde el procesamiento es muy pesado. Los trabajos revisados en la literatura no mencionan como solucionan el problema el procesamiento digital de imágenes cuando la base de datos es muy grande, ya que las bases de datos con las que se realizan las pruebas no son muy grandes, por lo que aquí se encuentra una oportunidad de explotar el uso del paralelismo.

En esta tesis se presenta el desarrollo de un sistema de clasificación de rostros no supervisado utilizando el análisis de puntos de interés SURF, junto con el análisis de textura local CBIR a partir de puntos de interés marcados de manera manual dentro del rostro de las personas y el uso del cómputo paralelo para agilizar el proceso de entrenamiento. Usualmente para generar sistemas de clasificación de rostros se utilizan las características geométricas, por lo que, para darle apoyo a estas características se planea comparar como funciona la técnica CBIR aplicado a rostros. En esta propuesta se evaluará el desempeño y robustez de nuestro sistema, en la clasificación y posible detección de rostros, desde el análisis de características geométricas del rostro humano,

y además las características de textura que conformen el patrón de la imagen, de tal forma que incluya información de rasgos geométricos del rostro, con esto el patrón será robusto.

Los sistemas de recuperación de imágenes basados en contenidos presentan dos módulos prioritarios:

- a) Representación y almacenamiento de imágenes.
- b) Recuperación de imágenes ante una consulta del usuario.

El uso de la técnica CBIR en el reconocimiento de rostros es una nueva línea de investigación ya que existen muy pocos estudios del uso de esta técnica aplicada al reconocimiento de rostros, por lo que es una gran oportunidad de explotar esta técnica y determinar que tan bien se desempeña al trabajar con un mismo conjunto de imágenes (rostros) y diferentes clases (personas). Para delimitar nuestro trabajo, las imágenes de rostros a usar para esta propuesta serán sencillas, es decir de frente, con el rostro descubierto y con un nivel de luminosidad suficiente para distinguir las facciones, tal como se muestra en la Figura 1.3.



Figura 1.3: (a) Rostro en condiciones ideales. (b) Rostro en condiciones no-ideales.

1.1. Objetivos

1.1.1. Objetivo general

Propuesta un sistema eficiente no supervisado de clasificación de rostros que integre técnicas CBIR, para manejar patrones robustos que consideren información adicional de textura como apoyo a los rasgos geométricos.

1.1.2. Objetivos específicos

- Obtener un patrón de características CBIR lo suficientemente robusto para mejorar la clasificación de rostros.
- Contribuir en el diseño de técnicas robustas de clasificación de imágenes de rostros.
- Proponer un sistema de clasificación eficiente usando arquitecturas de cómputo paralelo.

El resto de la tesis se compone de la siguiente manera: En el Capítulo 2 se presenta el estado del arte del reconocimiento del rostro, con una revisión de las investigaciones más importantes reportadas en la literatura. En el Capítulo 3 se presenta el marco teórico utilizado para el desarrollo del sistema, en el Capítulo 4 se presenta la metodología de nuestro clasificador para resolver la clasificación de rostros tanto para las características geométricas del rostro puntos SURF y las características de textura local CBIR. En el Capítulo 5 se presenta un análisis de la complejidad numérica que representa el algoritmo de la metodología propuesta, todo esto para proponer los lineamientos de paralelización de la metodología. En el Capítulo 6 se muestran los resultados obtenidos durante el desarrollo de la implementación de la metodología, así como el análisis de los resultados esperados; y finalmente en el Capítulo 7 se presentan las conclusiones del trabajo y las líneas de trabajo futuro.

Capítulo 2

Estado del arte

En este capítulo se muestra el desarrollo del estado del arte sobre el proceso de detección del rostro humano en una imagen digital, la detección de puntos de interés, como área de desarrollo de nuestra propuesta, el resumen de los diferentes trabajos analizados sobre el reconocimiento de rostros.

2.1. El rostro visual

El rostro juega un papel muy importante en nuestra vida diaria [20], gracias a su amplia variedad de funciones. Al visualizar la expresión facial podemos determinar, de alguna manera, ciertos elementos como: el modo de juzgar, el carácter o las intenciones de la persona. Otras señales faciales, más explícitas, nos permiten estimar factores como: la edad, género, origen étnico e identidad. La identidad como señal visual para reconocimiento, es una de las tareas principales de investigación en el área de reconocimiento de rostros o reconocimiento facial.

Los cambios del rostro

El proceso de la identificación de una persona a partir de la apariencia facial tiene que ser realizada en la presencia de una serie de factores que entran en conflicto. El rostro puede cambiar su apariencia debido a diversos factores como pueden ser: *intrínsecos* o *extrínsecos*. Los factores *intrínsecos* toman lugar muy independiente de cualquier observador y es debido a la naturaleza física del rostro. De igual manera, los factores *extrínsecos* aparecen cuando el rostro es observado vía la interacción: luz, rostro y

observador.

La apariencia de un rostro se altera de manera considerable dependiendo de las condiciones de iluminación y, en particular, de la proyección de la sombra propia. Las características de la cámara (o lente) usadas para observar el rostro también afectan la calidad de la imagen. Otros objetos en la escena pueden causar oclusión, y sombras, así como alteraciones en la luz incidente. Una de las fuentes de variación más significativas es el cambio en la posición. Es de notar que, la posición de un rostro está determinada por la orientación y posición tridimensional (3D) del observador. La causa principal del cambio en la posición es relativa al *movimiento rígido* entre el observador y el sujeto. Un rostro sufre movimiento rígido cuando este cambia su posición y orientación en el espacio tridimensional relativo al observador. Un rostro también puede sufrir un *movimiento no rígido* cuando su forma tridimensional cambia debido, por ejemplo, al hablar o por su expresión facial. Esto resulta en una variación intrínseca de la apariencia. Aunque estos tipos de movimiento ocurren juntos, en general nos conviene tratarlos de forma separada. En este trabajo se considera el rostro en forma rígida con una expresión neutra y de frente.

2.1.1. Reconocimiento de rostros

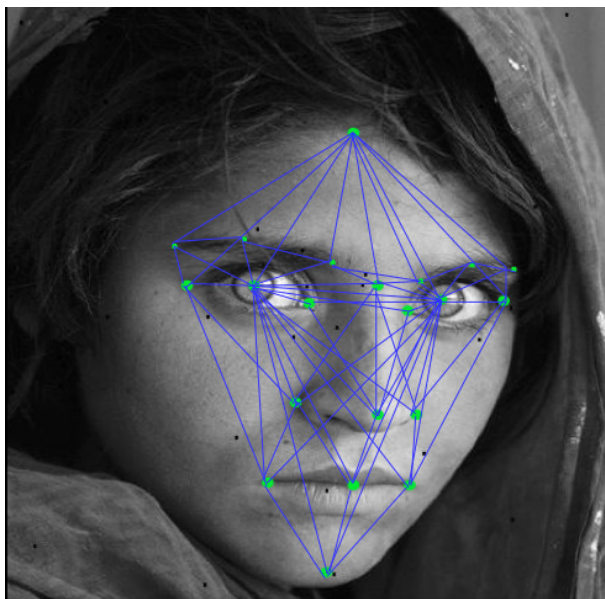


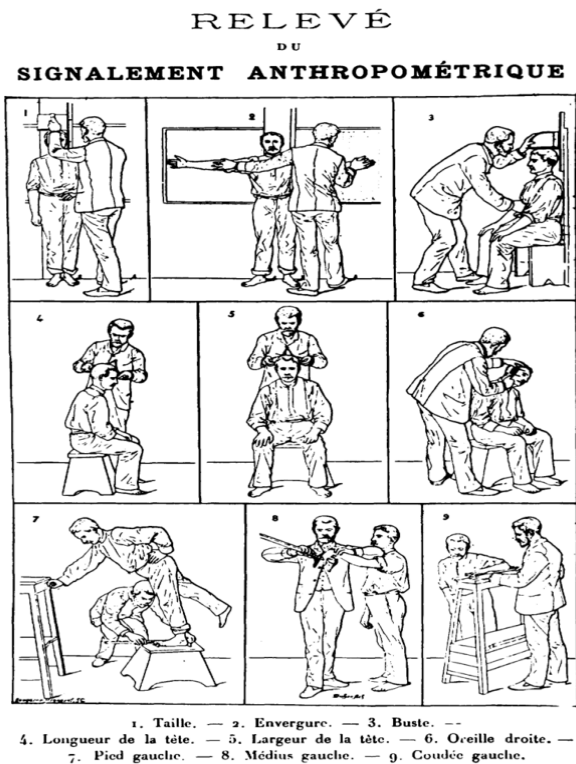
Figura 2.1: Reconocimiento de rostros características geométricas.

El reconocimiento de rostros es un área dentro del reconocimiento de patrones [11] que ha sido investigada desde hace varios años. Este problema de reconocimiento se hace difícil por la gran variabilidad en la rotación de la cabeza, su inclinación, la iluminación, el ángulo, la expresión facial, envejecimiento, etc.

Dentro del procesamiento digital de imágenes aplicado a reconocimiento de rostros, se trabaja con los llamados “puntos de interés” los cuales se definen, como la intersección de dos o más contornos. La evaluación antropométrica facial se basa en la extracción de puntos característicos del rostro, los cuales se determinan a través de las características visibles o palpables del espacio facial. La mayor parte de los estudios determinan que los ojos, la boca y la nariz, se nombran como las partes más importantes para recordar el rostro de las personas. Esto significa que existen características distinguibles que no se pueden encontrar en otras partes como la frente, mentón y mejillas[1].

Dentro del reconocimiento de rostros se mencionan diferentes problemas que se han detectado, a la hora de generar técnicas para automatizar este problema:

- **variabilidad intra-clases.** Una misma persona puede cambiar de manera drástica para un sistema de reconocimiento de rostros, ya que este puede tener diferentes poses, diferente intensidad de iluminación y diferentes expresiones; cubrir el rostro con algunos elementos (lentes, bufandas, gorras, etc.); la edad también juega un papel muy importante dentro del reconocimiento de rostros.
- **Similitud entre-clases.** Cuando las personas tienen rasgos semejantes, se tienen que encontrar los rasgos que hacen diferente una persona de otra.
- **La Base de Datos.** Otro de los grandes problemas en el reconocimiento de rostros, es la base de datos, esto quiere decir que, para poder evaluar los métodos de reconocimiento de rostros, la base de datos debe ser adquirida en condiciones equivalentes para todas las personas y el número de personas debe ser lo suficientemente representativo, pero hay que tener en cuenta que un aumento en el número de datos, aumenta el costo computacional.
- En 1882 en el Reino Unido el *Sistema Bertillon* [19] comenzó a hacer reconocimiento de rostros en presos Figura 2.2; ya que las simples medidas biométricas (altura, peso, color de piel, etc.) no eran suficientes, se comenzaron a tomar medidas geométricas del cráneo, cara, pies, brazos, etc. Lo que sea necesario para que la persona no pudiera cambiar de identidad con el paso del tiempo.



C. L. Brown

Height	1m 79.6	Head l'gth	19.8	L. Foot	27.1	Circle	ch	Age	22	Born in	
Eng. H'ght	5-10 3/4	Head width	16.3	L. Mid. F.	11.2	Periph Z		Apparent Age			
Outs. A	1m 75.5	Cheek width	14.4	L. Lit. F.	8.7	Color of Left Eye	Ch. Mel	Nativity	Louisville, Ky.		
Trunk	94.9	R. Ear	6.8	L. Fore A.	46.6	Pecul		Occupation	showman		
Remarks Incident to Measurement											

DESCRIPTIVE

Incl.	<i>Redd</i>	Ridge	<i>Very</i>	R. Ear	<i>—</i>	Beard	<i>Shaved</i>
Height	<i>M</i>	Base	<i>(Eu)</i>	Rostr	<i>Thin</i>	Hair	<i>Black</i>
Width	<i>Red</i>	DIMENSIONS				Teeth	<i>Upper good</i>
Pecul		Length	<i>er</i>	Projection	<i>m</i>	Complexion	<i>M. Dark</i>
		Pecul		Breadth	<i>m</i>	Weight	<i>165</i>
				Chin	<i>M. Prom</i>	Build	<i>M. Slim</i>

BUREAU OF IDENTIFICATION
Department of Police,
Tulane Ave. and Saratoga St.
New Orleans, La.

Measured *Feb 1 1912*
By *Geo. B. Jones*

Figura 2.2: El valor de fotografiar a un preso fue reconocido por la Corte Penal Habitual, Reino Unido en 1869, imagen tomada de [2].

- Dada la misma persona detectarla en múltiples facetas, es una tarea muy difícil, en un experimento se trata de empatar 700,000 rostros con una galería de 5,000,000 (florida DMV) y se obtuvieron 5,000 rostros duplicados [2], esto nos puede servir para clasificar los rostros y no tener datos redundantes, se puede generar una proceso de similitud .

■ Aplicaciones

En la actualidad, existen diversas aplicaciones cuyo funcionamiento está basado en el reconocimiento facial, dividiéndose en tres grandes grupos.

- El grupo más conocido por la gente, es el **comercial**, es decir, aplicaciones para la introducción de un usuario en redes de computadoras, seguridad electrónica, acceso a internet, cajeros automáticos, controles de acceso, teléfonos móviles, etc. Entre las aplicaciones que tienen cabida en este primer grupo, son remarcables los ejemplos de *Google* y *Facebook*.

La última innovación de *Google*, ha sido un nuevo sistema facial que permite al usuario el acceso a la información de su Smartphone o tableta. Dicho reconocimiento facial permite crear a la vez distintos perfiles en un mismo aparato. Solo hace falta que registre sus datos biométricos para que sus dispositivos los reconozca, después de configurar el acceso a sus datos y documentos personales para que solo el usuario tenga acceso a ellos. La localización de los rasgos biométricos de cada persona se basaría en un análisis de la forma y el tamaño de la nariz, la mandíbula y los pómulos. *Google* ha patentado este desbloqueo de dispositivos con un simple reconocimiento facial ¹.

Por otra parte, *Facebook* ha creado una herramienta de reconocimiento facial, que permite reconocer al usuario mediante una foto. Sin embargo, en este caso, han tenido que desactivar su tecnología de reconocimiento facial y borrar los datos incluidos en los perfiles de sus usuarios, debido a la Comisión de Protección de Datos (DPC), ya que exigen una mayor transparencia para los usuarios sobre la manera en que se tratan sus datos ².

- El segundo gran grupo al que se hace referencia, es el **gubernamental**, cuyas aplicaciones son las vinculadas a documentos identificativos (DNI, pasaporte, permiso de conducir), Seguridad Social, control de fronteras, control de aeropuertos, etc. La aplicación más conocida y moderna, perteneciente a este segundo grupo es el ABC system, utilizado para el chequeo de pasaportes (se pone un pasaporte y se comprueba que la foto coincide con la del mismo, registrada en la base de datos). Este sistema se ha implementado en el Aeropuerto de Barajas, como un nuevo método de reconocimiento facial. A partir de la imagen del rostro, se determinan unos cuantos números que representen únicamente a esa persona y se compara con los de una base de datos para conocer quién es ese individuo. Para obtener la información adecuada, lo mejor es que la imagen se capte lo más frontal posible.
- En el tercer grupo, el **forense**, tienen lugar investigaciones criminales, identificación de cadáveres, terrorismo, identificación de personas desaparecidas, etc.

¹<http://yosoyandroid.com/2012/09/google-obtiene-patente-para-desbloqueo-facial-de-un-dispositivo-para-multiples-usuarios/>

²<http://www.abc.es/tecnologia/redes/20140321/abci-facebook-reconocimiento-facial-preciso-201403211737.html>

Para el reconocimiento de rostros se han propuesto varias técnicas, entre las más populares se encuentran [1]:

- **Basados en geometría** [16][4]. En esta técnica las características son extraídas basándose en información geométrica tal como la posición y tamaño de los componentes de la cara; generalmente estos métodos requieren una selección muy cuidadosa para los umbrales.
- **Basados en Plantillas** [42][37][28]. En esta técnica, se realiza una correspondencia con alguna de las plantillas previamente diseñadas, la mejor correspondencia debe producir energía mínima; además de los altos costos computacionales estos métodos presentan dificultades cuando las características varían de manera significativa, por ejemplo en los ojos cerrados, los ojos con gafas y bufandas 2.3.
- **Basados en apariencia**[24][43]. Esta técnica hace uso del color de la piel para aislar la región facial; dentro de la región facial, cualquier región que no tenga el color de la piel es aceptada como una posible zona para la ubicación de los ojos o la boca. Sin embargo el rendimiento de esta técnica se limita a causa de la diversidad étnica.

2.1.2. Automatización del reconocimiento de rostros

En la automatización del reconocimiento de rostros, el esfuerzo ha sido multidisciplinario y en esta tarea participan tantas áreas variadas como las ciencias de la computación, la ciencia cognitiva, las matemáticas, la física, la psicología y la neurobiología. Dentro de las aplicaciones deseables del reconocimiento de rostros basado en computadora, se incluyen interfaces hombre-máquina, multimedia, vigilancia, seguridad, entretenimiento, etc. Por esto mismo, durante años recientes se han realizado enormes esfuerzos de investigación en el estudio y desarrollo de modelos computacionales, algoritmos y sistemas de visión por computadora para el reconocimiento automático de rostros [5].

Una de las principales ventajas del reconocimiento de rostros, es que se trata de un método no intrusivo, es decir, los datos pueden ser adquiridos incluso sin que la persona se de cuenta de ello. Además, el aspecto facial es el método más utilizado de manera natural por los seres humanos para reconocerse unos a otros .

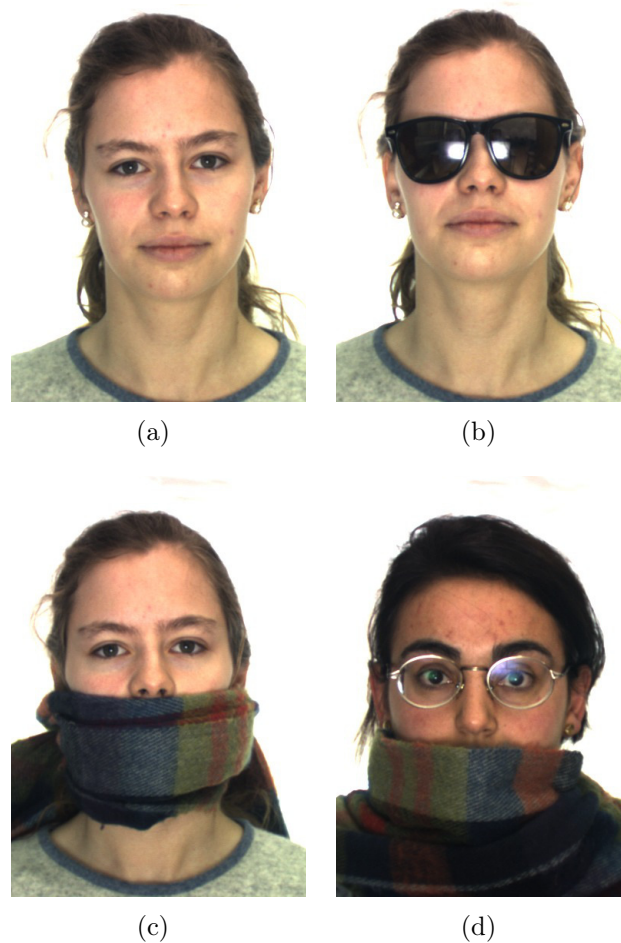


Figura 2.3: (a)Rostro ideal, (b)Rostro con lentes, (c)Rostro con bufanda, (d)Rostro con lentes y bufanda.

Luis Blázquez [5] presenta un sistema de reconocimiento facial basado en puntos de interés dentro de entornos no controlados, en este trabajo se realiza la comparación de diferentes métodos de extracción de puntos de interés, a partir de bases de datos utilizadas en la mayor parte de la literatura, algunas bases de datos extraen sus imágenes en entornos no controlados, y una vez extraídos los puntos de interés se generan los rasgos geométricos, los cuales serán comparados con los rasgos obtenidos en la base de datos y así determinar las personas más parecidas con respecto a los rasgos geométricos, consultados.

González y Prieto [1] proponen una técnica para la extracción de 22 de puntos de interés de la cara, orientado en aplicaciones de antropometría. La técnica se basa en las transformaciones waveles-Gabor y el uso del algoritmo EBGM (del término en in-

glés: Elastic Bunch Graph Matching). El EBGM se modificó para hacer que los puntos extraídos corresponden a puntos de interés que se utilizan comúnmente en las medidas antropométricas del rostro. Las modificaciones realizadas consisten en un conjunto de restricciones geométricas para ajustar inicialmente la ubicación de los centros de búsqueda, y posteriormente para la definición de la región de esta búsqueda. Los resultados mostraron que los puntos centrales del rostro presentan errores de ubicación inferiores al milímetro, lo cual es consistente con las medidas en antropometría facial directa, estas determinaciones se hacen respecto a unas escalas que se tienen a la hora del análisis de la imagen, ya que determinan que un pixel tiene cierta medida.

Vezzetti y Marcolin [40] muestran un análisis de las fortalezas y limitaciones de lo que está disponible en el dominio de la gemoetría, proporcionan lo que ellos llaman la primera guía para apoyar el desarrollo de un enfoque automático de reconocimiento de rostros. A partir de la geometría diferencial, como los coeficientes de las formas fundamentales, las curvaturas principales, media y curvaturas de Gauss, los derivados y la forma e índices “curvedness” introducidos por Koenderink y VanDoorn, analizan las actuaciones de los descriptores geométricos en el estudio facial tridimensional mediante el uso de un conjunto de indicadores (similitud entre diferentes caras, sensibilidad al ruido, etc.).

Bereta, Pedrycz y Reformat [7] generan un análisis comparativo entre diferentes descriptores locales. Realizan el análisis integral en profundidad comparativo de las técnicas de reconocimiento de rostros recientes basadas en descriptores locales. En particular, destacan sus principales rasgos en el contexto de problemas de reconocimiento de rostros. Identifican las principales ventajas y limitaciones de los métodos discutidos. Muestran que las técnicas presentadas son adecuadas para los sistemas de autenticación facial de gran escala, mencionan que generar un entrenamiento con la base de datos global es computacionalmente prohibido. Muestran una variedad de enfoques que se utilizan para realizar una fusión de los descripciones locales con los globales y estas se discuten junto con sus pros y sus contras. Finalmente también proponen diferentes medidas de similitud y posibles extensiones e hibridaciones con técnicas de aprendizaje estadístico.

Mehta, Yuan y Egiararian [31] presentan un nuevo enfoque para hacer frente al problema de reconocimiento de rostros, haciendo uso de la información direccional y la textura de las imágenes del rostro. Esta técnica captura la direccionalidad, utilizando la técnica de aproximación polinómica local, hicieron uso de filtros digitales adaptativos

de escala. Para la extracción de características de textura, se utiliza un descriptor local de dimensiones y computacionalmente. Características texturales y direccionales son capturados en los niveles basados holísticos y parte como resultado un descriptor robusto del rostro.

Sultana y Gavrilova [29] presentan un proceso de reconocimiento de rostros basado en contenido por medio de la técnica CBIR, en donde se propone una combinación del color, la textura y la forma, con estas características se pretende mejorar la precisión de recuperación del sistema, cada una de las características es extraída por separado de la imagen y son comparadas con cada una de las características de la base de datos de características, el color se compara por medio de la intersección del histograma, la textura y la forma se calculan por medio de la distancia euclidiana.

Martinez, Ortega y Arreguin [32] proponen un método modificado de autenticación de rostro basado en el pre-procesamiento de la imagen (principalmente en la ecualización del histograma, HE) y el uso de la técnica SURF en la etapa de extracción de características. El objetivo particular es determinar la autenticidad de una persona cuando él / ella tiene un par de expresiones faciales, diferentes fondos o una variación en la iluminación, ellos toman la imagen por completo y a esta le pasan el algoritmo SURF, para extracción de puntos de interés y así poder comparar una imagen de consulta con las imágenes contenidas en la base de datos de entrenamiento.

Dreuw, Steingrube, Hanselmann y Ney [13] presentan un análisis del uso de la técnica SURF aplicada al reconocimiento facial, donde comparan diferentes formas de extraer los puntos de interés dentro de la imagen del rostro, aplican tres métodos diferentes, matching máximo en donde extraen y comparan toda la imagen por completo y en este existen grandes errores de transformación, grid-Based matching en donde la imagen es dividida en rejillas y cada rejilla es comparada con su similar, aquí cada punto puede tener varios puntos consistentes y el Grid-Based Best matching es similar al anterior solo que en este un único punto solo tiene un punto que coincide con el punto de búsqueda, esta es la mejor de las 3 técnicas y se muestra en el análisis.

La metodología general para la automatización de un sistema de reconocimiento de rostros con base a lo visto en el capítulo 1 Figura 1.2.

Con los bloques de la forma:

1. La base de datos de rostros, la cual contiene las imágenes en condiciones ideales para el uso del sistema.

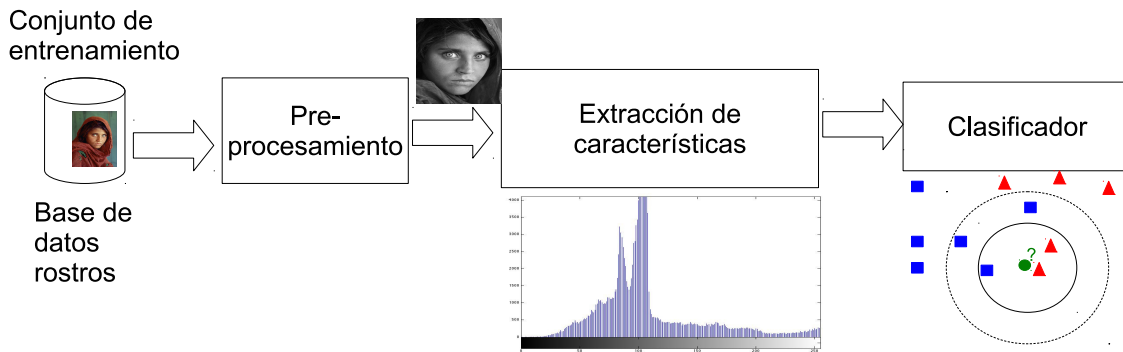


Figura 2.4: Esquema general para un sistema de reconocimiento de rostros.

2. La etapa de Pre-procesamiento se encarga de acondicionar la imagen en cuestión, e.g Eliminar ruido, disminuir brillo, aumentar contraste, realzar bordes, etc. elementos del procesamiento digital de imágenes necesarios para eliminar cualquier impureza de la imagen.
3. La extracción de características, parte de la imagen perfectamente acondicionada, para extraer los rasgos geométricos a partir de los puntos de interés, así como las características de textura, dentro de las vecindades al rededor de un punto de interés.
4. La etapa de clasificación, crea las clases correspondientes, con las cuales el sistema genera el conocimiento a partir de la extracción de características de las imágenes procesadas.

2.2. CBIR

En los sistemas CBIR(Content based image retrieval), una imagen de la base de datos es representada por una serie de características visuales que describen el contenido propio de la imagen. Para esto hay que considerar dos procesos para representar las imágenes:

1. La extracción de características.
2. La construcción de patrones visuales para almacenamiento y recuperación [10].

La extracción de características consiste en obtener información cuantitativa interesante o que sea fundamental para diferenciar una clase de objetos de otra. Estas características se dividen en dos tipos:

- **Características primitivas o de bajo nivel.** Se conoce así a la información visual por ejemplo del color, forma o textura.
- **Características de alto nivel.** Estas definen conceptos como montaña, cielo, playa, nube, etc.

Existen dos métodos de extracción de características:

- **Global.** Este se realiza sobre la imagen entera.
- **Local.** Se aplica a un grupo pequeño de píxeles de la imagen.

Por otro lado se tiene que un descriptor de una imagen es una forma de representar a una imagen a partir de sus características, con fines de almacenamiento y recuperación, en donde normalmente el descriptor es una formulación matemática. La formulación del descriptor se realiza tanto para características locales como globales.

Es así que esto lo aterrizamos en nuestra propuesta de trabajo de tesis, con la base de la metodología a implementar, y que se describe con detalle en el capítulo 3.

Capítulo 3

Marco teórico

En este capítulo se describen las bases del procesamiento digital de imágenes, y sus conceptos básicos que se manejan, así como lo que es el reconocimiento de patrones, estas técnicas son de las más importantes para el desarrollo de nuestro sistema de clasificación.

3.1. Procesamiento digital de imágenes

3.1.1. Definiciones previas

Imagen

Una imagen puede ser definida matemáticamente por una función bidimensional $f(x, y)$, donde x y y son coordenadas espaciales (en un plano), y f en cualquier par de coordenadas es la intensidad o nivel de gris de la imagen en esa coordenada [18]; $\forall x, y \in R$

Imagen digital

Cuando x , y , f los valores de f son todas cantidades finitas, discretas, tal que $x, y \in N$ y $f(x, y) \in N$ si $0 \leq f \leq L - 1$, decimos que la imagen es una *imagen digital* Figura 3.1 [18].

Una imagen digital se compone de un número finito de elementos, cada uno con un lugar y valor específicos.



Figura 3.1: Ejemplo de imagen continua (izquierda) vs imagen digital (derecha), siendo ésta última una aproximación discreta en puntos dispuestos en formación matricial, y valores de frecuencia

Pixel

El pixel es la unidad de medida más pequeña de video que tiene una luminosidad y color específico. Sus proporciones están limitadas por el número de líneas que componen la trama de barrido (el patrón de puntos que forman la imagen) y la resolución a lo largo de cada línea. En la forma más común de gráficas por computadora, los miles de diminutos píxeles que componen una imagen individual se proyectan en una pantalla de visualización en forma de puntos iluminados que desde la aparecen como una imagen continua. Un haz de electrones crea la cuadrícula de píxeles mediante el trazado de cada línea superior a la línea de fondo. Un pixel también puede ser el elemento más pequeño de un dispositivo sensible a la luz, como las cámaras que utilizan dispositivos de carga acoplada¹.

3.1.2. Paradigma de la imagen digital

1. *Resolución espacial.* Tiene que ver con el tamaño de la imagen Figura 3.2. Usualmente decimos que una imagen digital de tamaño $M \times N$ con L niveles de gris

¹“pixel”. Encyclopedia Britannica.2015. Web. 09 Jul. 2015. <http://global.britannica.com/technology/pixel>

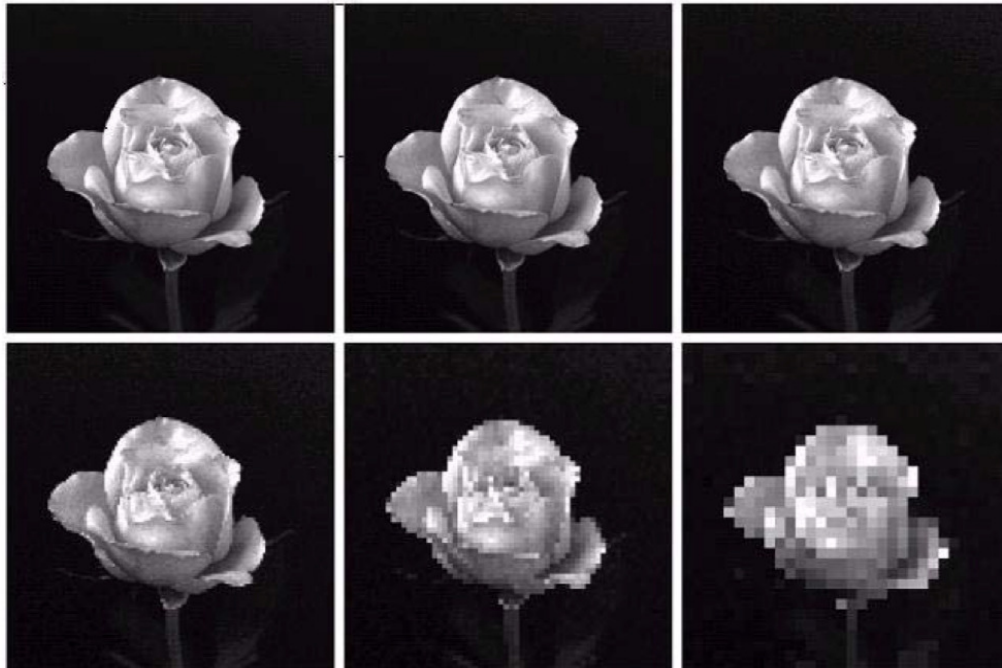


Figura 3.2: Imagen de 1024x1024 original y sus submuestreos de 512x512, 256x256, 128x128, 64x64 y 32x32.

tiene una resolución espacial de $M \times N$ píxeles y una resolución de nivel de gris de L niveles Figura.

e.g. Tenemos los estándares de tamaño en resolución de pantallas VGA (640x480), XGA (1024x768), etc. véase Figura 3.3.

2. *Cuantización del tono.* Se refiere al más pequeño cambio discernible en nivel de gris aunque, medir los cambios discernibles en niveles de intensidad es un proceso altamente subjetivo véase la Figura 3.4. Este tiene que ver con la discretización del nivel de gris asociado a cada punto de la imagen. La potencia de 2 que determina la cuantización, el número de niveles de gris, es usualmente 8 bits, es decir, 256 diferentes niveles de grises. Algunas aplicaciones especializadas utilizan 16 bits, esto depende de la aplicación ya que si necesitamos obtener muchos rasgos o información, es necesario tener mayor detalle del objeto. Veamos la tabla 3.1:

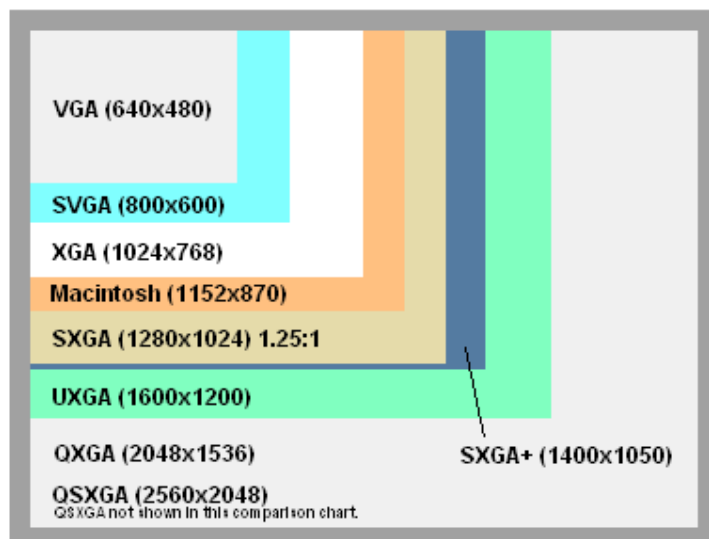


Figura 3.3: Tamaños estándar en resolución.

Número de bits	potencia	valores
1	2^1	2
2	2^2	4
4	2^4	16
8	2^8	256

Tabla 3.1: Tabla de valores potencia de 2.

Procesamiento digital de imágenes

El procesamiento de imágenes es el conjunto de técnicas computacionales para el análisis, la mejora, la compresión y la reconstrucción de imágenes. Sus principales componentes son importadores, en el que se captura una imagen a través de la exploración o la fotografía digital; el análisis y la manipulación de la imagen, se realiza utilizando diversas aplicaciones de software especializados; y de salida (por ejemplo, a una impresora o monitor). El procesamiento de imágenes tiene amplias aplicaciones en muchas áreas, incluyendo la astronomía, la medicina, la robótica industrial, y la teledetección por satélites².

Para el procesamiento digital de imágenes, la segmentación de imágenes es una de las principales problemáticas a resolver, permitiendo la extracción de contornos, texturas

²“image processing”. Encyclopedia Britannica.2015. Web. 09 Jul. 2015. <http://global.britannica.com/technology/image-processing>

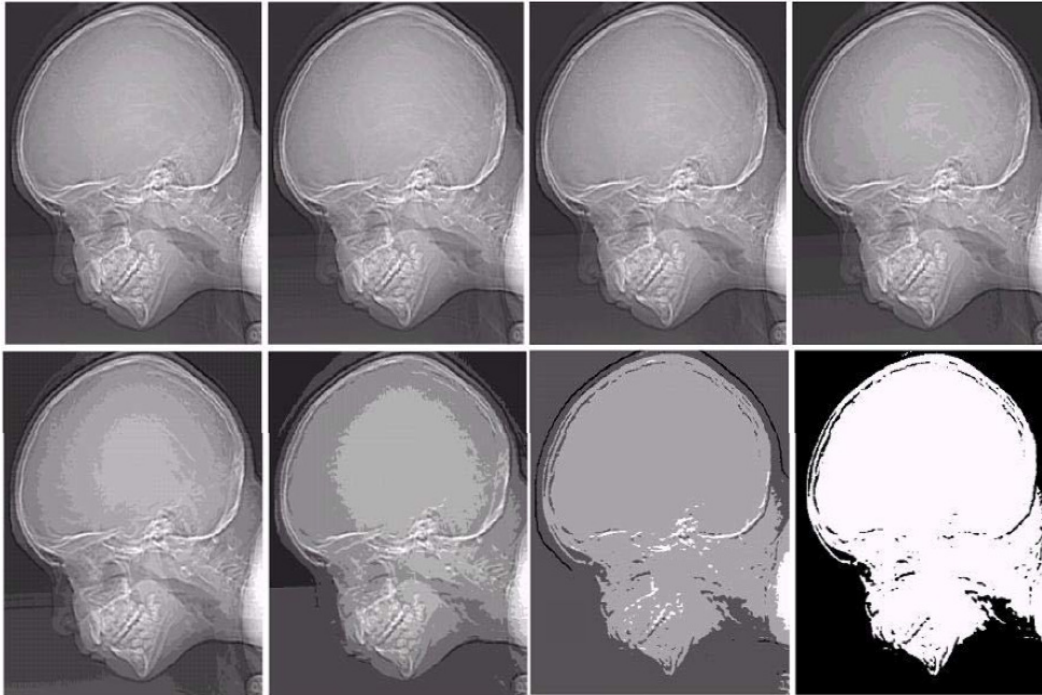


Figura 3.4: Imagen de 472 x 374 píxeles de resolución, en 256, 128, 64, 32, 16, 8, 4 y 2 niveles de gris.

y regiones presentes y facilita etapas posteriores del procesamiento relacionadas con la representación e interpretación de la información contenida. Existen diferentes formas de realizar la segmentación de una imagen. El enfoque basado en reconocimiento de patrones consiste en el uso secuencial de:

- Técnicas de extracción de características, que identifican o describen numéricamente píxeles o regiones de la imagen.
- Técnicas de reconocimiento de patrones, que permiten separar regiones en la imagen, identificando las mismas según sus diferentes características.

Una gran cantidad de algoritmos ha sido desarrollada para la extracción de características y reconocimiento de patrones, con enfoques y criterios muy variados. Cada uno se basa en un determinado paradigma de análisis del problema y un conjunto de parámetros que lo definen. La gran cantidad de parámetros involucrados en el uso de cada técnica hace a menudo difícil la optimización del algoritmo que se desea utilizar. Frecuentemente, además, es necesaria la combinación de más de una técnica para una tarea específica de segmentación.

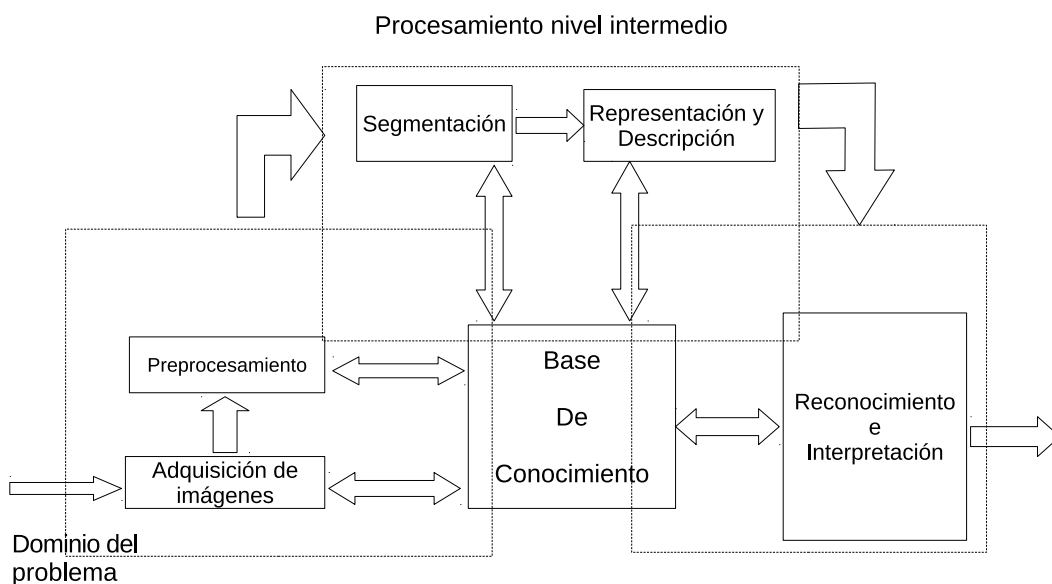


Figura 3.5: Sistema de Procesamiento Digital de Imágenes.

Nuestro sentido más avanzado es la vista, y no es sorprendente que las imágenes jueguen el papel más importante en la percepción humana de su entorno, y a partir de ello conforma muchas de sus tareas. La Figura 3.5 muestra un sistema del procesamiento digital del imágenes de propósito general propuesto en [18], el cual es el que se toma de referencia para todo el procesamiento de la imagen digital. A continuación vemos los bloques:

- **Adquisición de imágenes.** Generalmente incluye preprocesamiento e.g. Escalar la imagen.
- **Preprocesamiento.** Después de que la imagen digital ha sido obtenida, el siguiente paso es el preprocesamiento. Su función es mejorar la imagen de manera que se incremente la oportunidad de éxito de los siguientes procesos. El preprocesamiento casi siempre trata con técnicas para realzar el contraste y eliminar cualquier elemento de ruido.
- **Segmentación.** Definida de forma amplia, la segmentación particiona una imagen de entrada en sus partes constituyentes u objetos. En general, la segmentación automática es una de las tareas más difíciles en el procesamiento digital de imágenes. Un proceso de segmentación robusto lleva el proceso un largo camino hacia la solución satisfactoria de los problemas de imagen que requieren los objetos para

ser identificados individualmente. Por otro lado, los algoritmos de segmentación débiles o críticos casi siempre garantizan el fracaso final. En general, cuanto más precisa es la segmentación, el reconocimiento más probable es tener éxito.

- **Segmentación y Descripción.** La segmentación y la descripción casi siempre siguen después de una etapa de segmentación, que por lo general son píxeles en formato RAW³, que constituye ya sea el límite de una región (es decir, el conjunto de píxeles que separan una región de imagen de otra) o todos los puntos en la misma región. Descripción, también llamada función de selección, se ocupa de extraer los atributos de la información, los cuales resultan en alguna información cuantitativa de interés o son básicos para la diferenciación de una clase de objetos de otro.
- **Reconocimiento e interpretación.** El reconocimiento es el proceso que etiqueta, o asigna un nombre, a un objeto basándose en la información que proveen sus descriptores. La interpretación involucra la asignación de significado a un conjunto de objetos reconocido. En términos de nuestro ejemplo, identificar un rostro de una persona, requiere la asociación de los descriptores para este *rostro* con la etiqueta *R1* por ejemplo.
- **Base de Conocimiento.** La base de conocimiento mantiene toda la información (conocimiento) que se tiene acerca del problema. Este conocimiento puede ser simple, teniendo solo las regiones de la imagen donde la información de interés se sabe que esta localizada, con lo que se reduce la búsqueda que gua el camino a esa información. Además de guiar la oración de cada modulo del proceso, la base del conocimiento también controla la interacción entre módulos. Como se representa con flechas de doble sentido en la Figura 3.5.

Es importante mencionar que los resultados del procesamiento pueden verse en la salida de cualquier paso de la Figura 3.5. También debe hacerse notar que no todas las aplicaciones del procesamiento digital de imágenes requieren la complejidad de interacciones de la Figura 3.5. Muchas aplicaciones caen fuera de este esquema. De hecho, no todos los módulos, son siempre necesarios. Por ejemplo, el realce de imágenes para interpretación visual humana rara vez vas más allá de la etapa de

³El formato de imágenes RAW (entiéndase como "formato de imagen sin modificaciones") es un formato de archivo digital de imágenes que contiene la totalidad de los datos de la imagen tal y como ha sido captada por el sensor digital de la cámara fotográfica.

pre-procesamiento. En general, las funciones de pre-procesamiento que incluyen reconocimiento e interpretación están asociadas con aplicaciones de análisis de la imagen en la cual el objetivo es extraer las características o información necesaria de una imagen de manera automática o semiautomática

Aunque los seres humanos estemos limitados a la banda visible del espectro electromagnético (EM), las máquinas pueden percibir casi el espectro completo, desde los rayos gamma, a las ondas de radio. Las máquinas también pueden procesar imágenes generadas por fuentes que los humanos no asociamos con imágenes; como es el caso del ultrasonido, la microscopía de electrones, etc. véase la Figura⁴ 3.6.

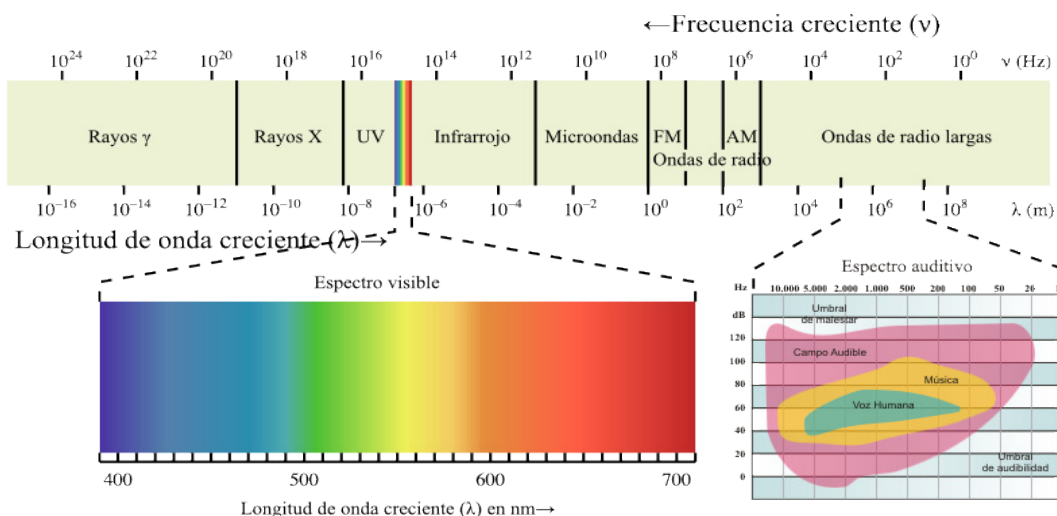


Figura 3.6: Espectro electromagnético

Los autores de muchos artículos y libros aun no se ponen de acuerdo para decir dónde termina el campo del Procesamiento Digital de Imágenes y dónde empiezan otros campos como el Análisis de Imágenes y la Visión por Computadora. La tarea de esta última disciplina es utilizar computadoras para emular la visión humana, incluyendo el aprendizaje, hacer inferencias y actuar basándose en entradas visuales. La Visión Computacional, es claramente, un área de la Inteligencia Artificial[39].

Algunos de los conceptos utilizados en este proyecto respecto al procesamiento digital de imágenes se pueden ver en el apéndice de la página 81.

⁴Espectro electromagnético: <http://www.paz-digital.net/imagenes/haarp/haarp004.png>

3.2. Reconocimiento de patrones

El reconocimiento de patrones es la ciencia que se encarga de la descripción y clasificación (reconocimiento) de objetos, personas, señales, representaciones, etc. Esta ciencia trabaja con base en un conjunto previamente establecido de todos los posibles objetos (patrones) individuales a reconocer [15].

Con el paso de los años se han diseñado e implementado algoritmos para simular la habilidad humana de reconocer objetos. A la par con el desarrollo de métodos capaces de emular las diferentes formas de reconocimiento de patrones, ha evolucionado la construcción de sistemas automáticos dotados de cierta inteligencia, que procesan la información de los objetos. Las tareas en reconocimiento de patrones pueden ser descriptivas o predictivas. Las tareas descriptivas son para caracterizar el conjunto de datos de cierta clase y las tareas predictivas son para asignar una etiqueta clase a una muestra a partir de un conjunto de muestras cuya clase es conocida [15].

3.2.1. Patrones

La representación que describe a un objeto se le llama patrón [12]. Existen dos tipos de patrones: abstractos y concretos, los patrones abstractos son aquellos que no podemos visualizar por decir las ideas, pensamientos, estados de ánimo y los patrones concretos son aquellos que podemos visualizar por ejemplo, símbolos, letras, fotografías, firmas, rostros, señales auditivas, etc. [8]. Un patrón está formado por un conjunto de n valores o medidas para las características o atributos que representan a un objeto. Dependiendo del dominio del problema, un patrón puede ser tratado como un vector de características en un espacio n -dimensional, como se puede ver la ecuación 3.1; o como una cadena de símbolos, como se puede ver en la ecuación 3.2, donde S es el conjunto de todas las cadenas formadas a partir de los símbolos. El conjunto de los valores posibles para las características de un patrón es el espacio de características o espacio de representación.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_3 \end{bmatrix} \in X \subset \mathbb{R} \quad (3.1)$$

$$x = x_1, x_2, \dots, x_n \in S \quad (3.2)$$

El reconocimiento de patrones se enfoca de dos maneras: supervisado y no supervisado. En el supervisado se cuenta con información de la categoría a la que pertenecen las muestras de un conjunto de entrenamiento. Y en el no supervisado no contamos con la información referente a las clases.

3.2.2. Elementos de un sistema de reconocimiento de patrones

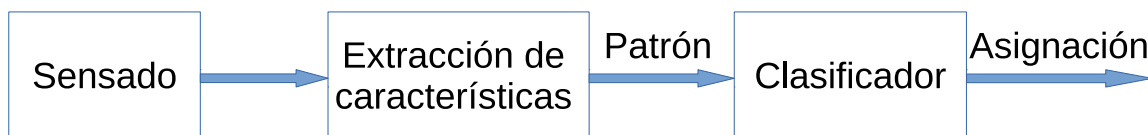


Figura 3.7: Elementos de un sistema de reconocimiento de patrones.

En la Figura 3.7, se describen las etapas de un sistema de reconocimiento de patrones, el objetivo de estas etapas es ajustar el sistema para que sea capaz de clasificar señales u objetos de entrada en una de las clases predefinidas. Para ello deberá analizar un cierto número de características y para poder clasificar satisfactoriamente señales de entrada, es necesario un proceso de aprendizaje en el cual el sistema crea un modelo de cada una de las clases a partir de una secuencia de entrenamiento o conjunto de vectores de características de cada una de las clases.

Como puede verse en el diagrama de la Figura 3.7, el sistema de reconocimiento de patrones de manera general se divide en tres etapas:

1. *Sensado*: El objetivo de esta etapa es convertir y representar los datos obtenidos para su futura clasificación. Generalmente el costo de este depende de la calidad del sensor o electrodo ocupado, por lo tanto entre mayor calidad se tenga mayores características tendrá, por ejemplo menor ruido o mayor fiabilidad en las señales obtenidas.
2. *Extractor de características*: A partir de la señal obtenida por el sensado, a este se le extraerá información conocidas como características relevantes y así eliminará redundancias e información irrelevante.
3. *Clasificador*: de acuerdo a las características extraídas en la etapa anterior, el clasificador asignará una categoría apropiada a los patrones obtenidos.

3.2.2.1. Diseño de un sistema de reconocimiento de patrones

En la Figura 3.8 se muestran las 3 etapas de diseño que incluye una serie de tareas que se deben realizar para su construcción. Contiene: la etapa de datos, la etapa de características y la etapa de clasificación.

El proceso inicia con la adquisición y preparación de los datos (tarea 1.1). En la tarea (1.2), preparación de los datos, se plantea una estrategia de pre-procesamiento, que puede ser de diferente índole dependiendo del tipo de patrones, con el objetivo de limpiar, seleccionar e integrar los datos para su procesamiento.

En la etapa de características se realiza primero la tarea (2.1), análisis de datos con el objetivo de plantear una estrategia de extracción de características. Posteriormente se aplican algoritmos de selección y transformación en la tarea (2.2), este procesamiento en el dominio de las características es con dos objetivos: incrementar la separabilidad de las clases y reducir la dimensión de los patrones. A continuación se realiza la tarea (2.3), una evaluación de la separabilidad de las clases, en caso de no obtenerse los resultados deseados, podemos replantear la estrategia de formación de características o de procesamiento en el dominio de las características. La etapa termina cuando se consigue un conjunto de características con un criterio de separabilidad conveniente.

La última etapa es para la construcción del clasificador. En la tarea (3.1) se selecciona y se configura un algoritmo de clasificación. Posteriormente en la tarea (3.2), se entrena el clasificador, esta tarea depende del algoritmo seleccionado y consiste en utilizar un conjunto de datos, llamado conjunto de entrenamiento, para especificar el clasificador.

Finalmente una evaluación del clasificador en la tarea (3.3), si el resultado de la evaluación no es satisfactorio, hay diferentes opciones de retorno en el proceso de diseño, se puede cambiar la estrategia de evaluación, el conjunto de datos de entrenamiento, el tipo de clasificador utilizado o regresar a las etapas anteriores para replantear la estrategia de formación de características o de adquisición y preparación de los datos.

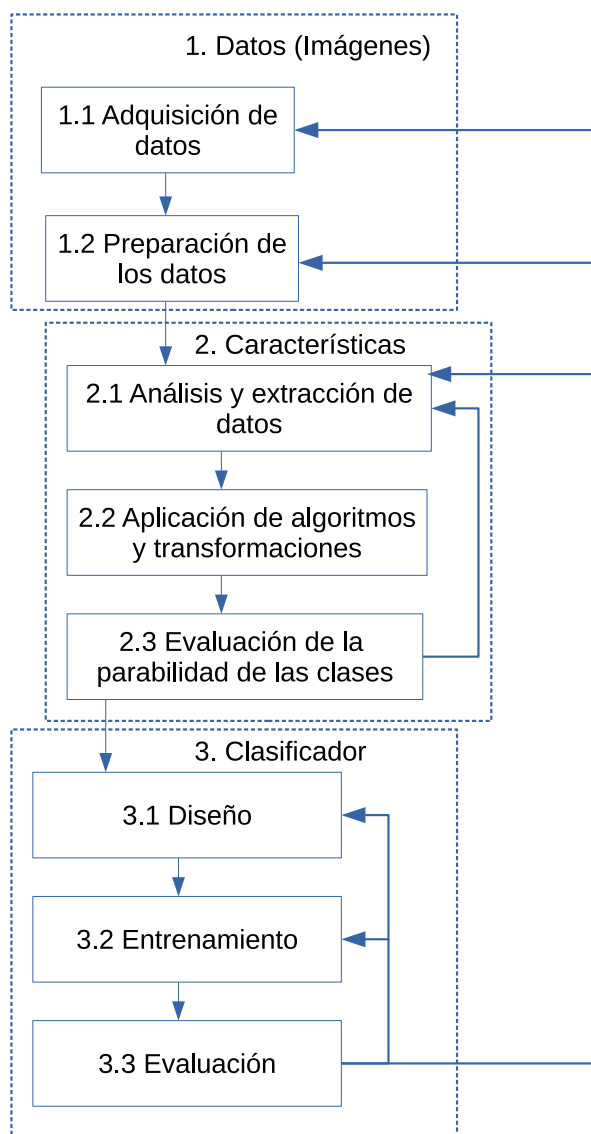


Figura 3.8: Diseño de un sistema de reconocimiento de patrones.

Capítulo 4

Metodología propuesta

En este capítulo se describe con detalle cada uno de los pasos involucrados en la aplicación de la metodología propuesta, para nuestro sistema de clasificación, este sistema implementa dos metodologías mencionadas con anterioridad como son: CBIR y SURF.

La clasificación y reconocimiento de imágenes se ha estudiado con diferentes enfoques, desde el etiquetado manual de las imágenes hasta el etiquetado automático por análisis semántico del texto de las páginas web en que aparecen las imágenes.

Otro enfoque es el análisis automatizado del contenido de una imagen refiriéndose en su contexto a colores, formas, texturas o cualquier otra información que pueda derivarse de la propia imagen.

Específicamente para la detección de rostros la metodología tradicional marca 4 etapas: detección de la cara, alineación de la cara, extracción de las características y reconocimiento, tal como se muestra en la Figura 4.1. Diversos trabajos han mostrado el uso de la técnica CBIR en el problema de analizar el contenido de imágenes como Vogel [21] , Caltech [26] y [23] .

Los factores principales que se deben tener en consideración en la detección de rostros son:

- Luminosidad
- Orientación
- Escala

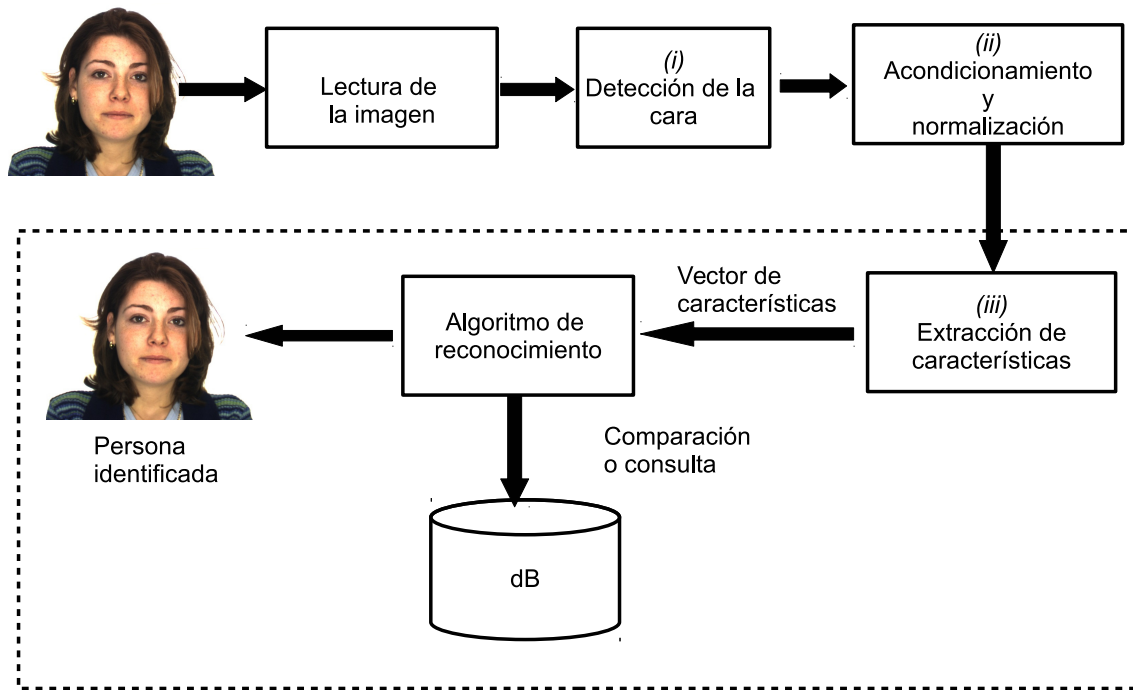


Figura 4.1: Metodología clásica de la identificación de rostros.

- Textura
- Accesorios u objetos que puedan afectar a la geometría esperada (gafas, bufandas, etc.), véase la Figura 2.3.

4.1. Arquitectura propuesta

En esta propuesta las imágenes de rostros a trabajar consideran el último factor, como una variante de pruebas finales de comparación, centrándose en los cuatro primeros para la extracción de características globales, y en los puntos geométricos de disposición de elementos del rostro para extracción de características locales.

La forma de trabajar de la metodología CBIR que planteamos en la Figura 4.1 consta de una entrada, y una salida al sistema. En nuestra propuesta el sistema de identificación y clasificación consta de dos etapas: entrenamiento y recuperación. El sistema propuesto se muestra en la Figura 4.2, junto con su operación.

Entradas del sistema:

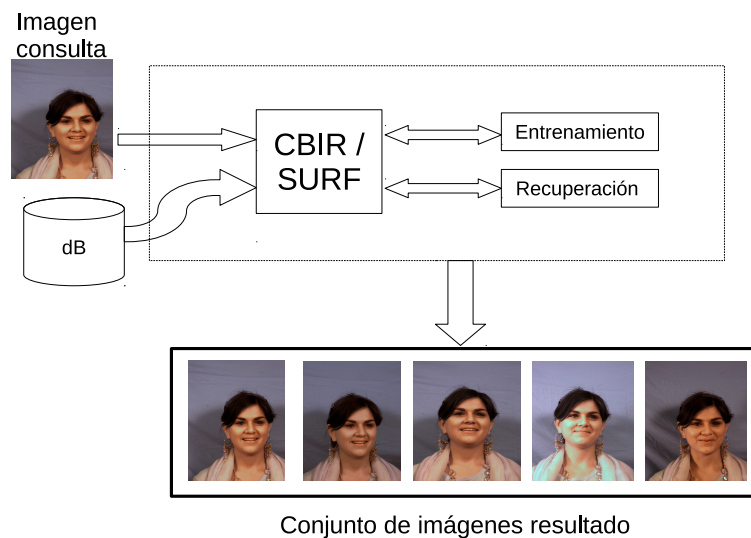


Figura 4.2: Arquitectura propuesta del sistema para la clasificación y recuperación de rostros.

- **Entrenamiento del sistema CIBR propuesto con imágenes de rostros:** Bases de imágenes de rostros en condiciones ideales. (Pruebas con 10, 50 y 100 imágenes).

Salidas del sistema:

- **Recuperación:** Las tres o cinco imágenes de las bases de datos que más se aproximen a la imagen consultada.

4.1.1. Entrenamiento CBIR

La arquitectura del sistema en el entrenamiento para el análisis CBIR parte de una arquitectura como se muestra en la Figura 4.3, la cual se describe de la siguiente manera:

- **dB.** Representa el conjunto de las imágenes a considerar, como una base de datos se simboliza con un cilindro, contiene las imágenes de los rostros en condiciones ideales.
- **Lectura de la imagen.** Se encarga de extraer las imágenes de la base de datos, que serán procesadas, la lectura de la imagen da como resultado una imagen en tres capas, correspondientes al espacio de color RGB.

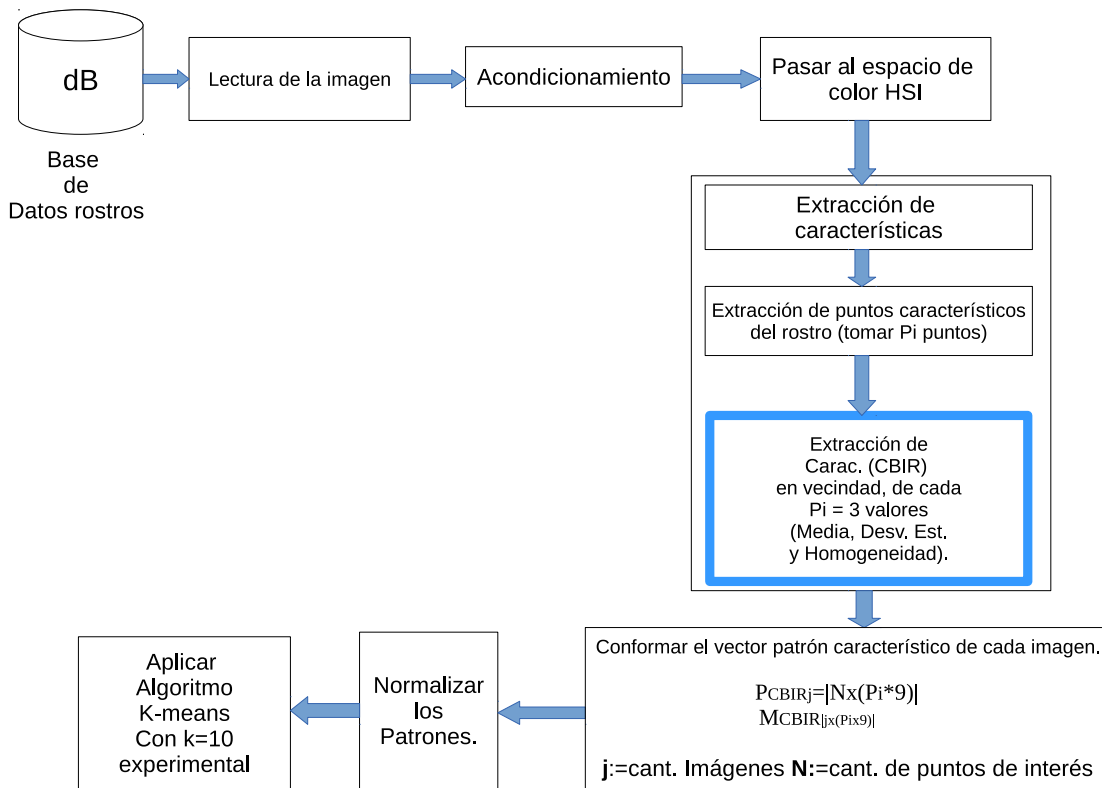


Figura 4.3: Arquitectura entrenamiento CBIR.

- Acondicionamiento.** Realiza el procesamiento previo de la imagen en cuestión en cada una de las capas, ya que no se puede trabajar directamente con una imagen recién adquirida, esta debe pasar por la etapa de acondicionamiento para que sea que sea eliminada cualquier impureza generada, tanto por el dispositivo de adquisición, la mala iluminación, el fondo, etc. Una vez eliminadas todas las impurezas la imagen esta lista para su correcto análisis.
- Pasar al espacio de color HSI.** El espacio de color RGB no nos da la información necesaria o clara para la extracción de características, por lo que la imagen RGB se debe pasar al espacio de color HSI; este espacio de color nos proporcionará la información que necesitamos saber acerca de la imagen, como es la textura, luminosidad, saturación de color, etc. información más clara que la que nos proporciona el espacio de color RGB, al final de este proceso tendremos nuevamente la imagen dividida en tres capas pero ahora en el espacio de color HSI.
- Extracción de características.** Este bloque es el más importante ya que es donde implementaremos la mayor parte del trabajo consta de tres etapas:



Figura 4.4: Ejemplo de los puntos de interés en una imagen de rostro. Aquí cada punto esta numerado indexado.

- **Extracción de los puntos de interés.** Las bases de datos con las que se trabaja tienen los puntos de interés marcados de manera manual, la localización de los puntos de interés se muestra dentro de un documento de texto para cada una de las imágenes, por lo que nosotros debemos seleccionar los puntos de interés con los que trabajará nuestro sistema, como se observa en la Figura 4.4, es una muestra de una base de datos de como están marcados los puntos de interés. Como se menciona en la literatura la mayor cantidad de puntos de interés con los que se trabaja están en la parte interna del rostro, por lo que tomaremos esos puntos y se trabajara con ellos como se observa en la Figura 4.4.
- **Extracción de los rasgos estadísticos .** Los rasgos estadísticos son generados para cada uno de los puntos de interés, de cada una de las capas del espacio de color HSI. A partir de un punto de interés determinado en un pixel se genera una ventana p de tamaño $\{P_i + p \times P_i + p\}$ pixeles con el punto de interés en el centro, dentro de esta ventana (*vecindad*) Figura 4.5 se extraerán tres valores de información estadística, la media, desviación estándar y homogeneidad, por lo que al final obtendremos N puntos de interés,

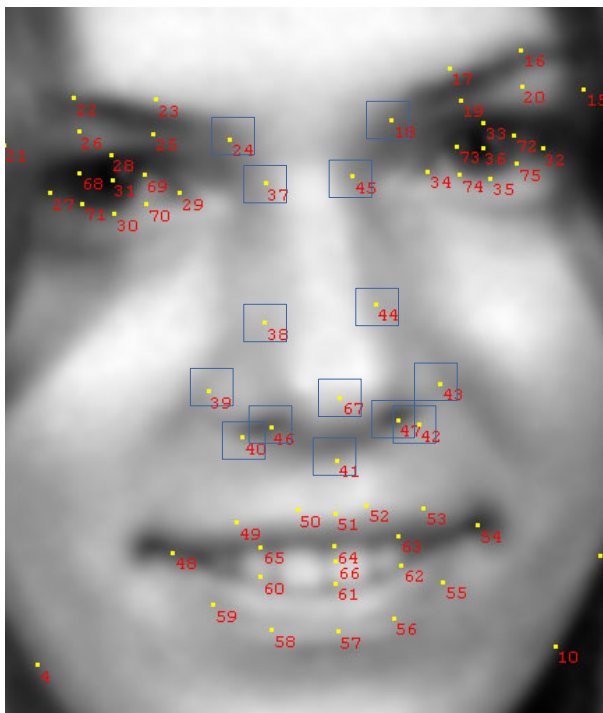


Figura 4.5: Ejemplo de construcción de las ventanas al rededor de un punto de interés del rostro. Sobre c/u de ellas se extraen las características a estudiar de la textura, como rasgos estadísticos.

3 valores estadísticos para cada vecindad del punto de interés y 3 capas por imagen, tendremos $\{N \times 3 \times 3\}$ valores estadísticos de la imagen en cuestión.

Al final de este proceso se obtendrá una matriz $M_{CBIR|j \times (N_{P_i} * 9)}$, con j =Número de imágenes, N_{P_i} =Número de puntos de interés, 9 =características estadísticas $\{media, homogeneidad, \sigma\}$ en cada capa HSI para un tamaño p de ventana al rededor del punto de interés .

- **Normalizar los patrones.** A partir de los valores de la matriz M_{CBIR} , se toman los valores máximos y mínimos de cada componente para normalizar los valores de la matriz entre 0 y 1, así se obtendrá una nueva base de datos con la información del contenido de las imágenes en valores normalizados.
- **Aplicar algoritmo k-Means con $k = 10$ experimental.** Se generan 10 clases (*grupos*), a partir de los valores de la base de datos normalizada, este algoritmo se aplica a la matriz normalizada, véase Figura 4.6, el valor de 10 es tomado como

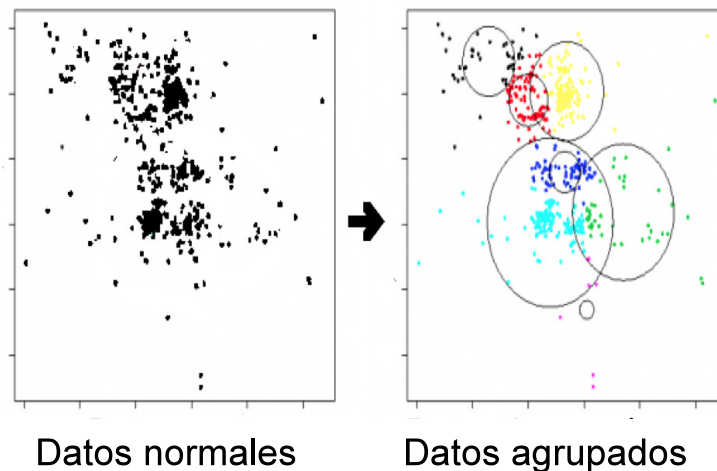


Figura 4.6: Ejemplo de una agrupación k-means.

referencia ya que este valor es aplicado en la literatura y funciona muy bien para el análisis de imágenes de escenarios naturales, por lo que para comenzar con las pruebas tomaremos este valor de $k = 10$.

4.1.2. Entrenamiento SURF

La arquitectura del sistema en el entrenamiento para el análisis SURF parte de una arquitectura como se muestra en la Figura 4.7, la cual se describe de la siguiente manera:

- **dB.** Contiene las imágenes de los rostros en condiciones ideales.
- **Lectura de la imagen.** Se encarga de extraer las imágenes de la base de datos, que serán procesadas, la lectura de la imagen da como resultado una imagen en tres capas, correspondientes al espacio de color RGB.
- **Acondicionamiento.** Realiza el procesamiento previo de la imagen en cuestión en cada una de las capas, ya que no se puede trabajar directamente con una imagen recién adquirida, esta debe pasar por la etapa de acondicionamiento para que sea que sea eliminada cualquier impureza generada, tanto por el dispositivo de adquisición, la mala iluminación, el fondo, etc. Una vez eliminadas todas las impurezas la imagen esta lista para su correcto análisis.

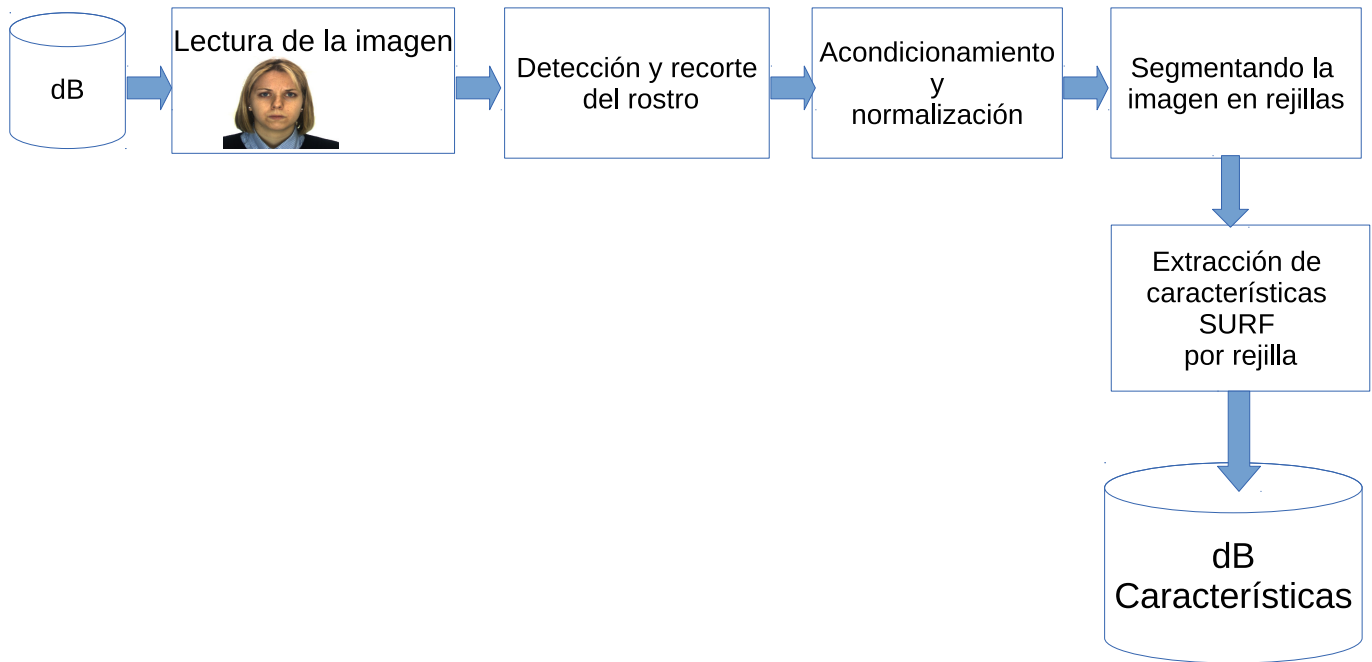


Figura 4.7: Arquitectura entrenamiento SURF.

- **Detección y recorte del rostro.** Utiliza el algoritmo de detección de rostro de la biblioteca OpenCV, una vez detectado el rostro se recorta de la imagen y se realiza un ajuste de tamaño a la imagen correspondiente.
- **Segmentación de la imagen en rejillas.** Cada una de las imágenes será segmentada en 225 rejillas, como se describe en [13], con un tamaño de 40×40 pixeles como lo muestra la Figura 4.8, esta metodología de análisis a mostrado mejores resultados que la propuesta por [32], en donde utilizan la imagen del rostro completa sin segmentarla en rejillas.
- **Extracción de características por rejilla.** A cada una de las rejillas de la imagen se le extraen los puntos de interés Figura 4.9 y el descriptor SURF para cada uno de esos puntos. Para cada imagen se obtendrán 225 contenedores de puntos de interés y 225 descriptores.
- **dB características.** Contiene la información de los puntos de interés y los descriptores de cada una de las imágenes, por lo que al final se tendrá un arreglo de puntos de interés y un arreglo de descriptores SURF: $vectorPuntosInterés[N_{imagen} * 225]$ y $vectorDescriptores[N_{imagen} * 225]$

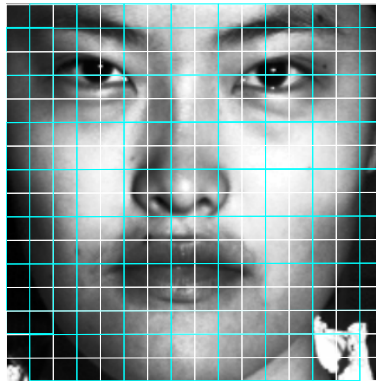


Figura 4.8: Imagen dividida en rejilla.

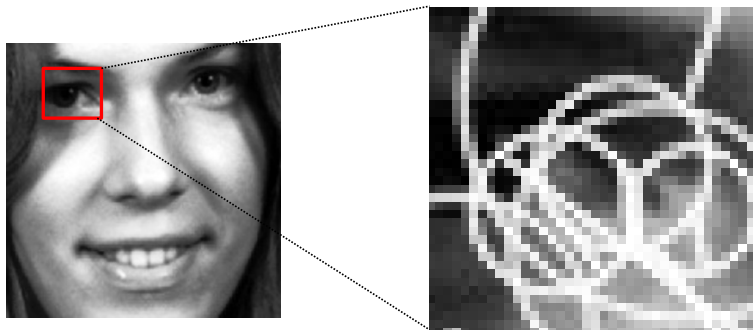


Figura 4.9: Extracción de los puntos de interés SURF

4.1.3. Recuperación CBIR

La arquitectura del sistema CBIR en la recuperación consta parte de una arquitectura como se muestra en la Figura 4.10, la cual se describe de la siguiente manera:

- **Lectura de la imagen.** Se encarga de extraer la imagen que será procesada, la lectura de la imagen da como resultado una imagen en tres capas, correspondientes al espacio de color RGB.
- **Acondicionamiento.** Realiza el procesamiento previo de la imagen en cuestión en cada una de las capas, ya que no se puede trabajar directamente con una imagen recién adquirida, esta debe pasar por la etapa de acondicionamiento para que sea que sea eliminada cualquier impureza generada, tanto por el dispositivo de adquisición, la mala iluminación, el fondo, etc. Una vez eliminadas todas las impurezas la imagen esta lista para su correcto análisis.

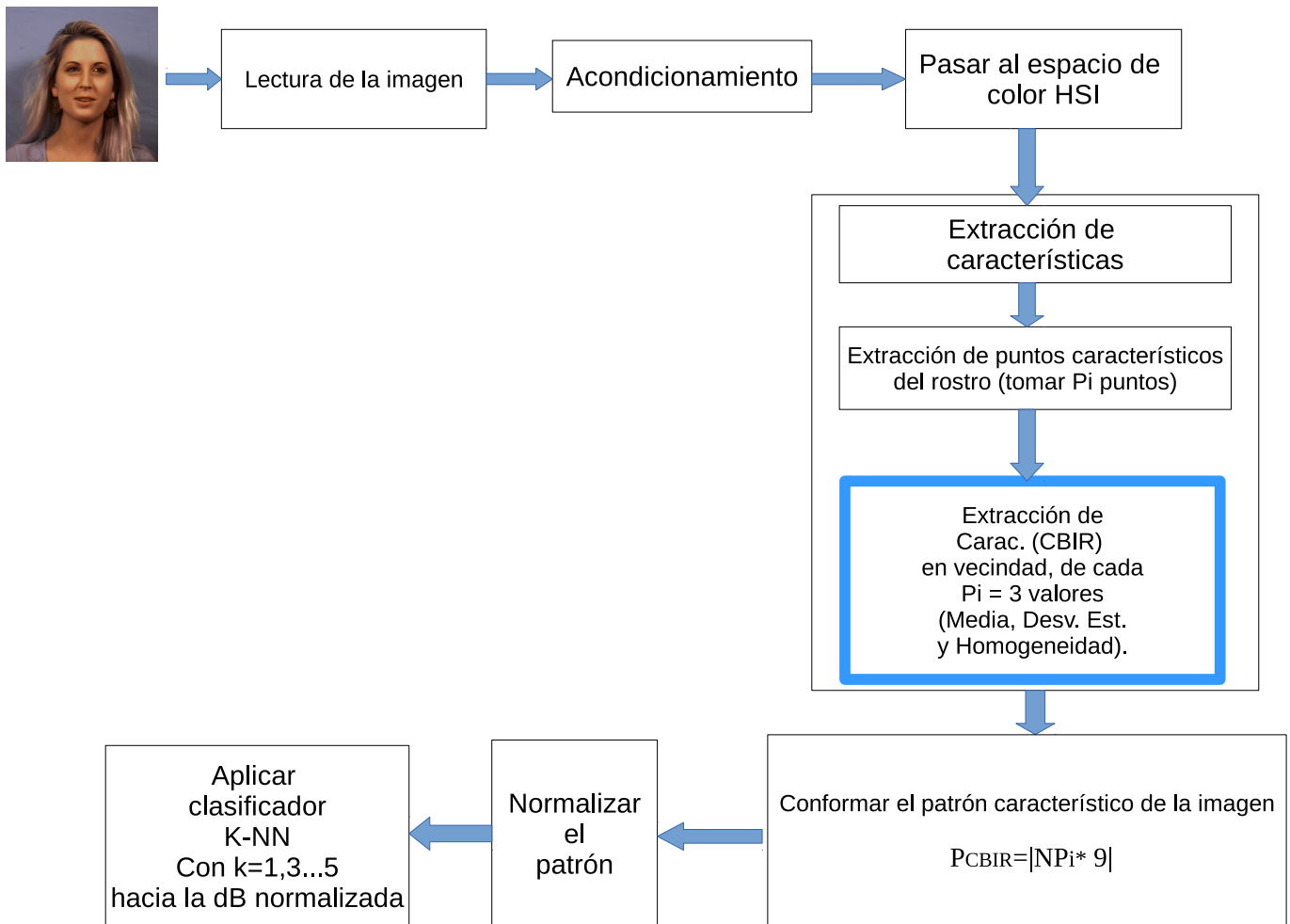


Figura 4.10: Arquitectura recuperación.

- **Pasar al espacio de color HSI.** El espacio de color RGB no nos da la información necesaria o clara para la extracción de características, por lo que la imagen RGB se debe pasar al espacio de color HSI; este espacio de color nos proporcionará la información que necesitamos saber acerca de la imagen, como es la textura, luminosidad, saturación de color, etc. información más clara que la que nos proporciona el espacio de color RGB, al final de este proceso tendremos nuevamente la imagen dividida en tres capas pero ahora en el espacio de color HSI.
- **Extracción de características.** Este bloque es el más importante ya que es donde implementaremos la mayor parte del trabajo consta de tres etapas:
 - **Extracción de los puntos de interés.** Las bases de datos con las que se trabaja tienen los puntos de interés marcados de manera manual, la lo-

calización de los puntos de interés se muestra dentro de un documento de texto para cada una de las imágenes, por lo que nosotros debemos extraer los puntos de interés con los que trabajara nuestro sistema, no importa como estén marcados en las diferentes bases de datos, para nuestro sistema el orden de cada uno de los puntos de interés se muestra en la Figura 4.4. Como se menciona en la literatura la mayor cantidad de puntos de interés con los que se trabaja están en la parte interna del rostro, por lo que tomaremos esos puntos y se trabajara con ellos como se muestra la Figura 4.5

- **Extracción de los rasgos estadísticos** . Los rasgos estadísticos son generados para cada uno de los puntos de interés, de cada una de las capas del espacio de color HSI. A partir de un punto de interés determinado en un píxel se genera una ventana p de tamaño $\{(P_i + p) \times (P_i + p)\}$ píxeles con el punto de interés en el centro, dentro de esta ventana (*vecindad*) se extraerán tres valores de información estadística, la media, desviación estándar y homogeneidad, por lo que al final obtendremos N puntos de interés, 3 valores estadísticos para cada vecindad del punto de interés y 3 capas por imagen, tendremos $\{N \times 3 \times 3\}$ valores estadísticos de la imagen en cuestión.
- **Normalizar el patrón.** A partir de los valores de la matriz de entrenamiento M_{CBIR} se toman los valores máximos y mínimos para normalizar el patrón obtenido entre 0 y 1.
- **Aplicar algoritmo k-NN con $N = 5$.** El valor del patrón es comparado contra los valores de las clases generadas en la etapa de entrenamiento, esta operación no es más que una simple distancia euclidiana entre dos puntos Figura 4.11, al terminar las operaciones se obtienen las 5 imágenes más parecidas a la imagen de consulta.

4.1.4. Recuperación SURF

La arquitectura del sistema en la recuperación para el análisis SURF parte de una arquitectura como se muestra en la Figura 4.12, la cual se describe de la siguiente manera:

- **imagen de consulta.** Imagen de prueba que entra a la etapa de recuperación.

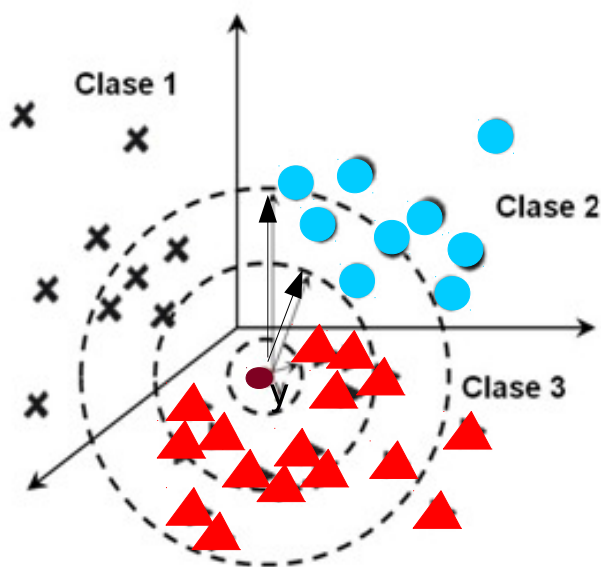


Figura 4.11: Ejemplo de algoritmo K-NN.

- **Acondicionamiento.** Realiza el procesamiento previo de la imagen en cuestión en cada una de las capas, ya que no se puede trabajar directamente con una imagen recién adquirida, esta debe pasar por la etapa de acondicionamiento para que sea que sea eliminada cualquier impureza generada, tanto por el dispositivo de adquisición, la mala iluminación, el fondo, etc. Una vez eliminadas todas las impurezas la imagen esta lista para su correcto análisis.
- **Detección y recorte del rostro.** Utiliza el algoritmo de detección de rostro de la biblioteca OpenCV, una vez detectado el rostro se recorta de la imagen y se realiza un ajuste de tamaño a la imagen correspondiente.
- **Segmentación de la imagen en rejillas.** La imagen será segmentada en 225 rejillas, con un tamaño de 40×40 píxeles como lo muestra la Figura 4.8.
- **Extracción de características por rejilla.** A cada una de las rejillas de la imagen se le extraen los puntos de interés Figura 4.9 y el descriptor SURF para cada uno de esos puntos. Para cada imagen se obtendrán 225 contenedores de puntos de interés y 225 descriptores.
- **Aplicar algoritmo k-NN con $N = 5$.** Los 225 descriptores generados, son comparados uno a uno con el descriptor correspondiente en cada una de las imágenes en la base de datos de características y al final se extraen las imágenes más

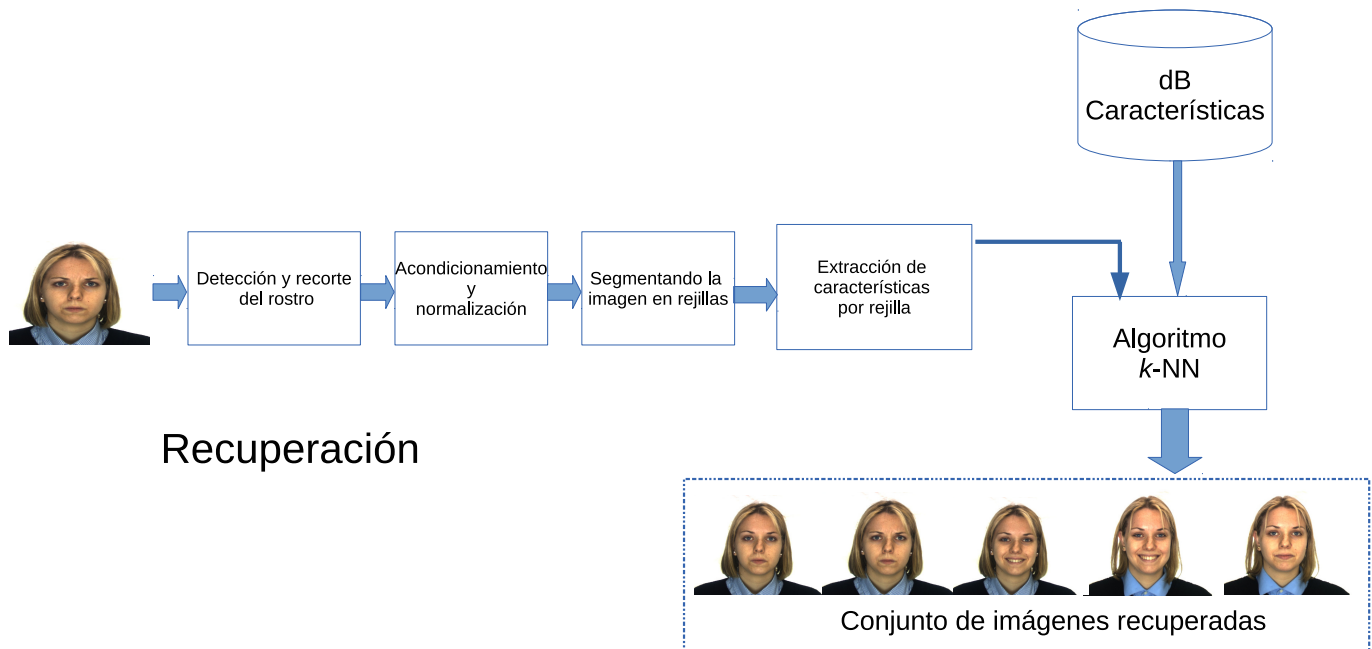


Figura 4.12: Arquitectura de recuperación SURF.

cercanas a la imagen de consulta que tuvieron el máximo número de puntos que realizaron el mejor “matching” en cada una de las rejillas. El resultado de los puntos que hacen “matching” se les aplica una operación de distancia Euclidiana para los puntos de interés involucrados, por medio de un radio mínimo para tener una tolerancia y asegurarse que la comparación realizada corresponde con la región analizada y no esta en una región más alejada como se muestra en la Figura 4, donde se observa un punto que esta dentro de la zona de búsqueda.

4.2. Bases de datos

Las bases de datos utilizadas para el estudio de esta metodología son las mencionadas en los artículos de investigación, y que más se apegan a la descripción de condiciones ideales, que nosotros propusimos, a continuación en la tabla 4.1 se muestra la descripción del contenido de cada una de las bases de datos.

dB	N. imágenes	N. personas	T. de imagen	Formato	Resolución
AR [30]	4000	126	Color	jpg	768 × 576
Ext. Yale B [17]	16128	10	Escala de grises	pgm	640 × 480
Yale [6]	165	15	Escala de grises	gif	320 × 243
Bio ID [22]	1521	23	Escala de grises	pgm	648 × 286
MUCT [33]	3755	276	Color	jpg	640 × 480
MEDS II [41]	1307	518	Color	jpg	variable

Tabla 4.1: Descripción de las bases de datos.

El resultado de las clases más parecidas de las dos técnicas, análisis de textura local CBIR y rasgos geométricos SURF, entraran a una sistema de voto (*clasificador conjunto*), donde se analizará el resultado del conjunto de las imágenes recuperadas, la salida de este sistema de voto generara las k clases más cercanas a la que corresponde la imagen consultada.

Capítulo 5

Arquitectura paralela del clasificador

En el Anexo 2 pagina 101, tenemos una introducción a lo que es el Cómputo paralelo, con algunas definiciones y términos que se utilizarán dentro de este capítulo, donde se describe la implementación de la arquitectura paralela generada a partir de la arquitectura mencionada en el capítulo anterior, la modificación de la metodología se realiza para poder ser utilizada dentro de un cluster en donde pretendemos reducir el tiempo de generación de la etapa de entrenamiento, por lo que es necesario desarrollar una arquitectura paralela que mejore el desempeño de este sistema.

5.1. Paralelización del la versión secuencial

Como se mencionó al inicio de este capítulo, el cómputo paralelo nos sirve entre otras cosas para reducir el tiempo de ejecución de un programa que trabaja con mucha información; también se mencionó que una de las líneas que de investigación que trabajan en cómputo paralelo es el procesamiento digital de imágenes, por lo que tenemos dos características que describen la necesidad de utilizar un sistema de cómputo paralelo. El cuello de botella detectado en la arquitectura previamente desarrollada esta en el proceso de extracción de características, el cual es demasiado costoso, este proceso es el que deseamos agilizar, por lo que se realizó un análisis del sistema y ver que partes se podrían realizar en paralelo, para obtener el mejor desempeño del sistema sin afectar el proceso de clasificación ya que es el objetivo principal del sistema. La arquitectura del sistema mostrada en el capítulo anterior se modificará para que se pueda utilizar en una arquitectura paralela tipo SPMD(single program-multiple data), por medio de

un sistema MIMD utilizando la comunicación entre las maquinas por medio del paso de mensajes, utilizando openMPI. Esta arquitectura se puede observar en la Figura 5.1 la cual se describe de la siguiente manera:

1. Dividir el número de imágenes contenidas en la base de datos entre el número de procesos.
2. Cada proceso genera una matriz parcial que contiene características estadísticas CBIR $\{prom, des. Std, homogeneidad\}$, en las 3 capas HSI, para cada punto de interés, del número de imágenes que le toco procesar $|m_{|N \times (NPi*9)|}|$, donde N =número parcial de imágenes a procesar, NPi =número de puntos de interés y 9 =características estadísticas CBIR.
3. Cada proceso extrae los máximos y mínimos de cada una de las columnas de su matriz parcial y los guardan en 2 arreglos $Max_{|1 \times NPi*9|}$ y $Min_{1 \times NPi*9}$
4. Los procesos calculan los máximos y mínimos globales por medio de la operación MIP_ALL_Reduce , y guardan ese valor en 2 nuevos arreglos $MaxGral_{|1 \times NPi*9|}$ y $MinGral_{1 \times NPi*9}$.
5. Los procesos normalizan cada elemento de la matriz entre 0 y 1 con los valores máximos y mínimos globales por medio de la operación $||m_i|| = \frac{m_i + |MinGral|}{MaxGral}$.
6. Cada proceso envía la matriz parcial normalizada al proceso 0, por medio de las operaciones MPI_Send y MPI_Recv . El proceso 0 concatena las matrices en una sola $|m * |$.
7. El proceso 0 ejecuta el algoritmo de agrupamiento $k-medias$, para agrupar la información de la matriz normalizada final.
8. el proceso 0 genera una matriz de índices del tamaño $k \times N_{imag}$, donde k =número de grupos y así termina la etapa de aprendizaje y la paralelización.

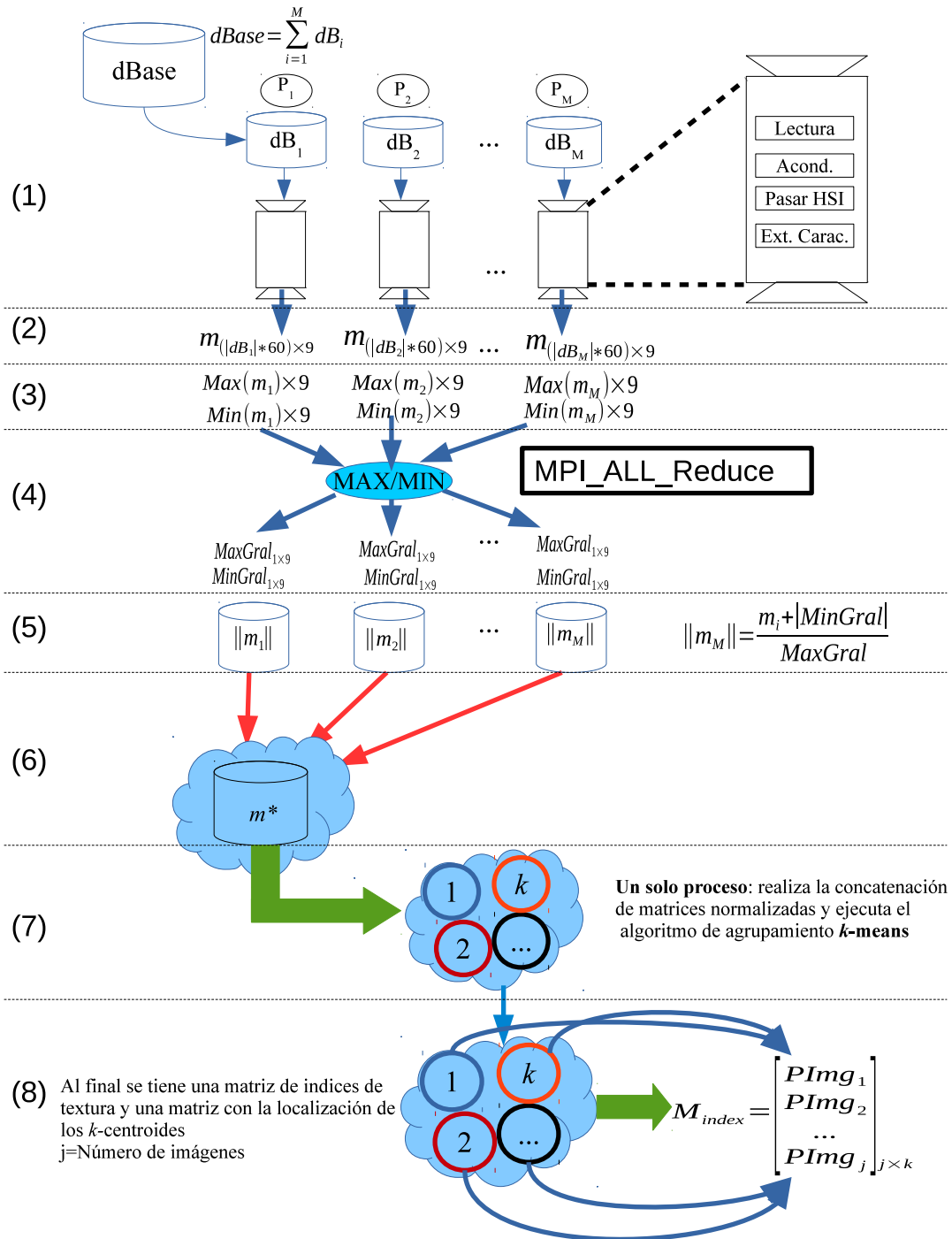


Figura 5.1: Arquitectura paralela del sistema de clasificación de rostros.

Capítulo 6

Resultados experimentales

En esta sección mostramos los resultados de los experimentos realizados para la metodología vista en el capítulo 3. Las pruebas utilizadas para el estudio de la robustez de nuestra metodología consta de tres tipos de pruebas, tal como es sugerido para el estudio del Reconocimiento de Patrones, a fin de poder validar la robustez y efectividad del clasificador y los resultados obtenidos; las tres pruebas son.

- **Resustitución.** En esta prueba se trabajo con el conjunto fundamental de la base de datos, es decir con todas las imágenes contenidas en la base de datos, tanto para el entrenamiento y la recuperación Figura 6.1(a). Sé probo la calidad del clasificador para las k imágenes más cercanas a la imagen de consulta y así determinar las clases a las que más se parece dicha consulta.
- **Dejar uno fuera (Leave One Out).** En esta prueba se entrena con $N - 1$ elemento de la base de datos Figura 6.1 (b), y para la etapa de recuperación se consulta la imagen que no conoce el clasificador, para determinar la clase a la que es asignada la imagen desconocida; esta es una de las pruebas que más tiempo se llevó ya que hay que hacer lo mismo para todas las imágenes contenidas en la base de datos.
- **Validación cruzada (Cross Validation).** En esta prueba se extrae un porcentaje de la base de datos Figura 6.1 (c), se entrena con el porcentaje restante y al final se realiza la consulta sobre el porcentaje de imágenes desconocidas. Esta prueba es una de las más difíciles ya que entre menos elementos tenga el clasificador para la etapa de entrenamiento menor será la precisión de clasificación a la

hora de entregar los resultados de cada una de las imágenes consultadas.

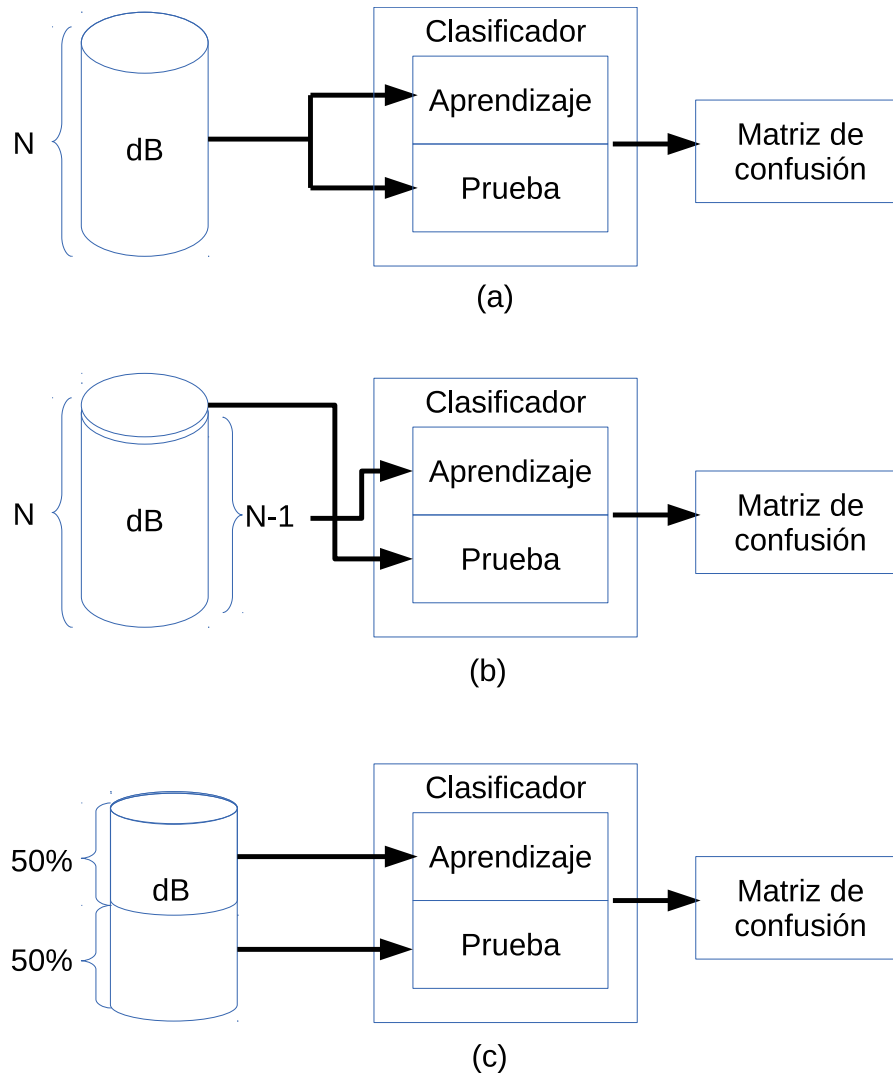


Figura 6.1: Diagramas de las metodologías de pruebas usadas: (a) Resustitución, (b) Dejar uno fuera (LOO), (c) Validación cruzada (CV). En cada figura podemos apreciar la característica de tamaño de la muestra para entrenamiento-aprendizaje del sistema, y la muestra con la que probamos lo aprendido.

Ruido. Se generan imágenes que simulan estar deterioradas con ruido tipo gaussiano al 1%, 3% y 5%, así como ruido mixto al 1%, 3% y 5%, algunos ejemplos de estas imágenes podemos verlas en la Figura 6.2. A estas imágenes se les realizara la prueba de resustitución, para ver el desempeño del sistema al trabajar con imágenes que contienen un poco de ruido.

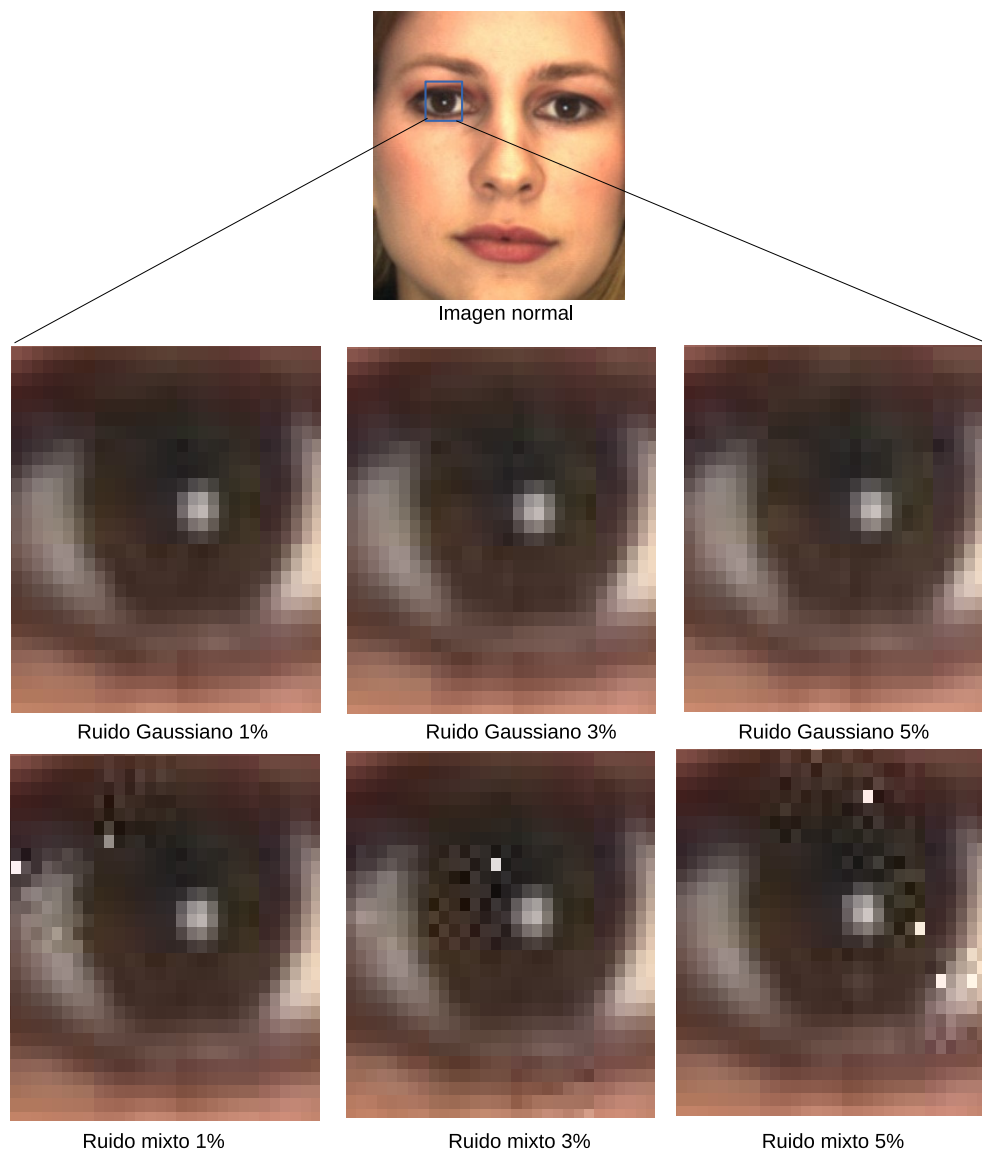


Figura 6.2: Ejemplo de imágenes contaminadas con ruido.

Cada una de estas pruebas nos arroja la “matriz de confusión”, la cual es con la que se calcularon los porcentajes de clasificación para cada una de las pruebas mencionadas, por medio del algoritmo k -NN.

6.1. Pruebas CBIR

Se realizarón las pruebas con el clasificador no supervisado k -means y el clasificador k -NN, construyendo las matrices de confusión con el método de Resustitución, LOO y validación cruzada para cada una de las bases de datos.

- El clasificador k -NN se tomo con $k = 5$.
- El calculo de las características estadísticas se genero a partir de los 60 puntos de interés con los tamaños de ventana $p = \{5, 7, 10, 15, 20\}$.
- El agrupamiento por medio del algoritmo k -means se obtuvo a partir de la variación del valor de $k = \{8, 10, 12, 14, 16, 18\}$.

Los resultados de esta variación de parámetros se muestra en las gráficas resultantes para la metodología de prueba, como se muestra en la Figura 6.4. En donde notamos que dependiendo el tamaño de la ventana y el número de grupos, se genera una mejora considerable para el desempeño de nuestro clasificador.

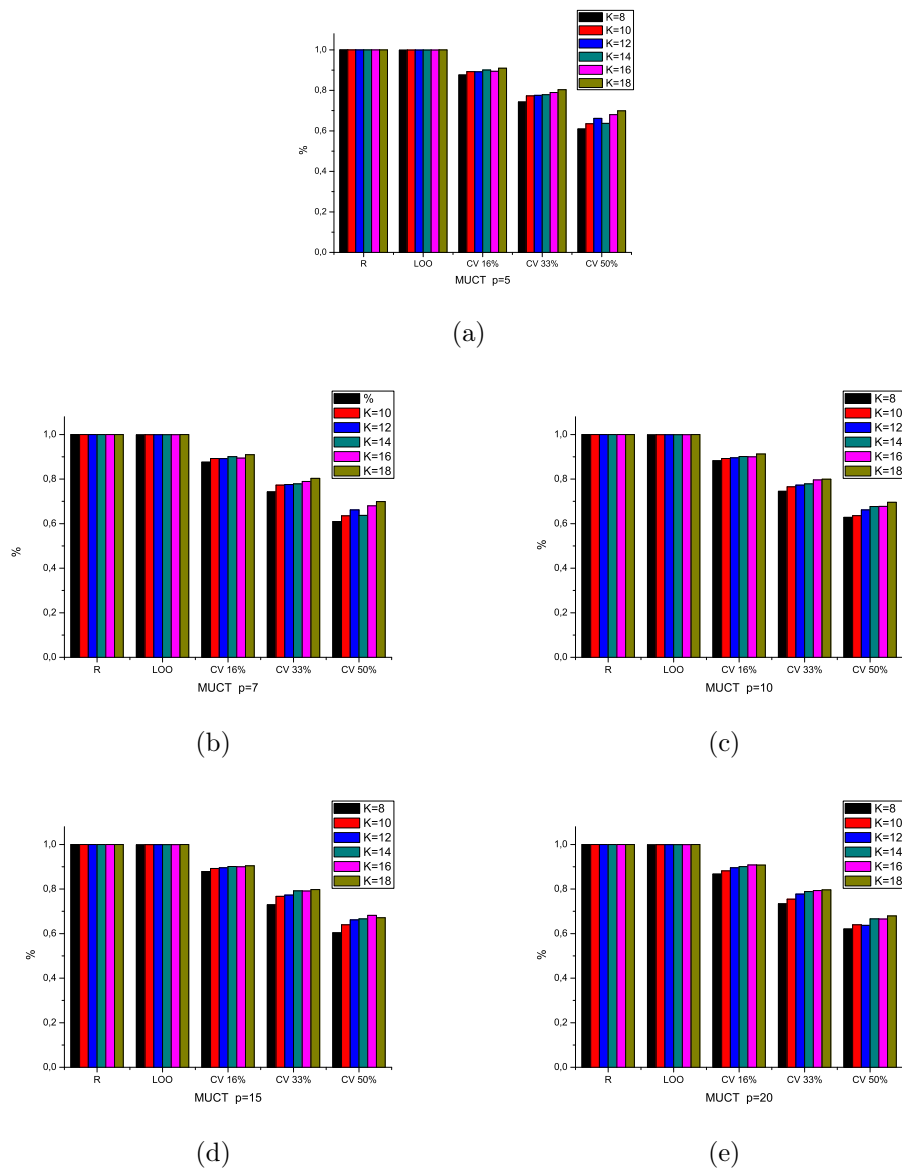


Figura 6.3: Porcentaje de recuperación con diferente tamaño de ventana, $p = \{5, 7, 10, 15, 20\}$ respectivamente.

Se decidió tomar el tamaño de ventana de $p = 10$ y el número de grupos como $k = 20$ después de el análisis de los resultados anteriores. Con esos valores de parámetros se obtuvieron los resultados siguientes para cada una de las bases de datos:

- AR. Esta base de datos se redujo a una cantidad de 6 imágenes por persona y contiene 46 personas (clases), esto hace un total de 276 imágenes contenidas en esta base de datos, las cuales cumplen mejor la descripción de condiciones ideales.

AR			
Prueba	k -NN=1	k -NN=3	k -NN=5
Res	100	100	100
LOO	64.13043478	83.33333333	89.13043478
CV17	63.04347826	80.43478261	93.47826087
CV33	66.30434783	89.13043478	92.39130435
CV50	48.55072464	64.49275362	70.28985507
R-Gaus1 %	86.5942029	96.73913043	98.55072464
R-Gaus3 %	80.43478261	91.66666667	96.73913043
R-Gaus5 %	71.37681159	86.5942029	93.84057971
R-Mixto1 %	88.04347826	96.73913043	98.1884058
R-Mixto3 %	85.14492754	96.01449275	99.27536232
R-Mixto5 %	84.7826087	94.20289855	96.73913043

Tabla 6.1: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos AR

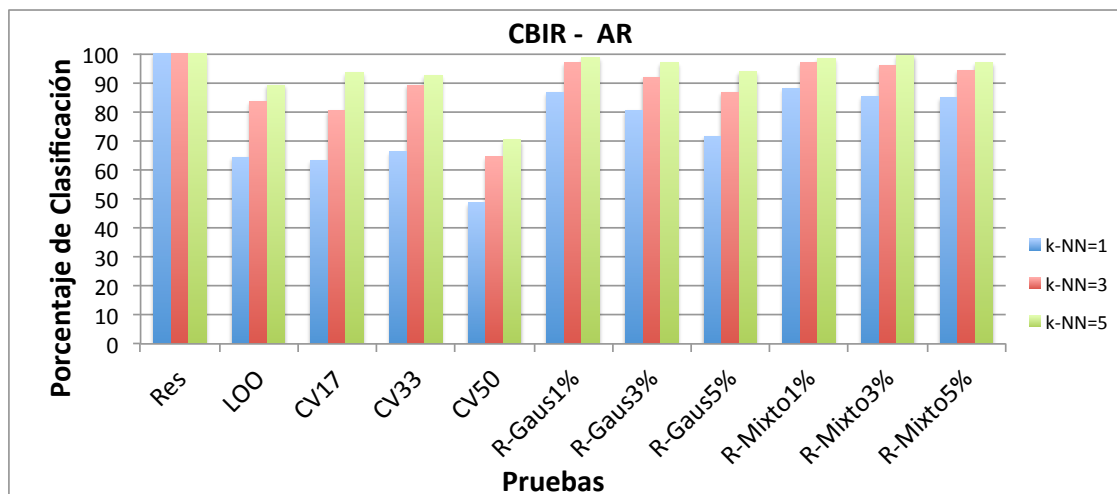


Figura 6.4: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos AR

- MUCT. Esta base de datos se redujo a una cantidad de 1368 imágenes, contiene 228 personas (*clases*) y 6 imágenes por persona, las cuales cumplen mejor la descripción de condiciones ideales.

MUCT			
Prueba	k-NN=1	k-NN=3	k-NN=5
Res	100	100	100
LOO	26.0964912281	42.3976608187	53.2894736842
CV17	16.88596491	32.01754386	39.69298246
CV33	50	57.69230769	65.38461538
CV50	15.05847953	25.14619883	31.87134503
R-Gaus1 %	50.14619883	68.20175439	75
R-Gaus3 %	44.66374269	61.33040936	68.49415205
R-Gaus5 %	40.93567251	56.7251462	65.35087719
R-Mixto1 %	56.79824561	76.31578947	82.23684211
R-Mixto3 %	52.85087719	72.44152047	80.70175439
R-Mixto5 %	48.83040936	68.71345029	75.73099415

Tabla 6.2: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos MUCT

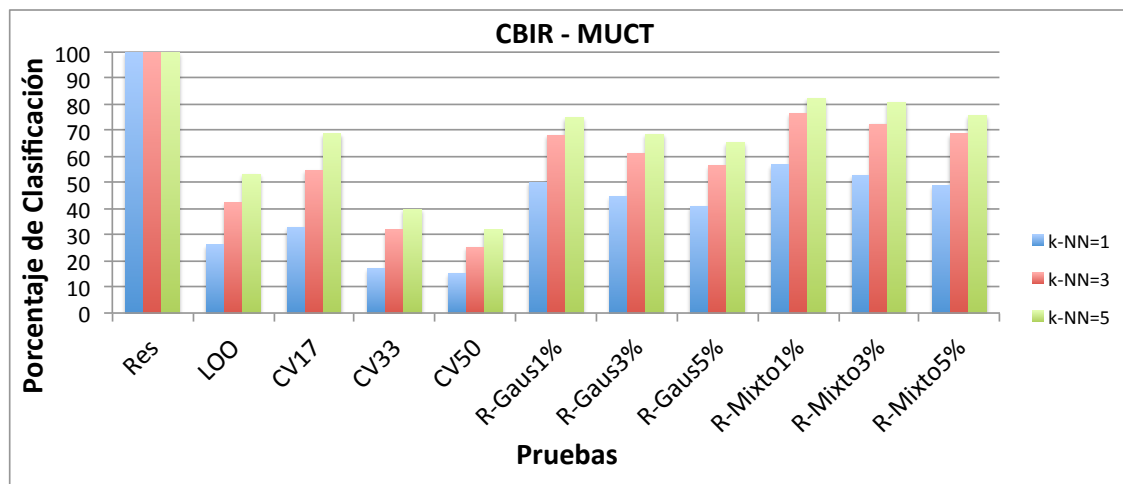


Figura 6.5: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos MUCT

- Yale. Esta base de datos se redujo a una cantidad de 104 imágenes, contiene 13 personas (*clases*) y 8 imágenes por persona, las cuales cumplen mejor la descripción de condiciones ideales, a esta base de datos se le redujo el número de grupos a $k = 10$, dejando el mismo tamaño de ventana $p = 15$.

Yale			
Prueba	k-NN=1	k-NN=3	k-NN=5
Res	100	100	100
LOO	50.96153846	67.30769231	75
CV13	15.38461538	30.76923077	46.15384615
CV25	50	57.69230769	65.38461538
CV38	58.97435897	74.35897436	79.48717949
CV50	53.84615385	69.23076923	80.76923077
R-Gaus1 %	100	100	100
R-Gaus3 %	89.42307692	98.07692308	98.07692308
R-Gaus5 %	81.73076923	94.23076923	98.07692308
R-Mixto1 %	100	100	100
R-Mixto3 %	97.11538462	99.03846154	100
R-Mixto5 %	88.46153846	97.11538462	99.03846154

Tabla 6.3: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos Yale

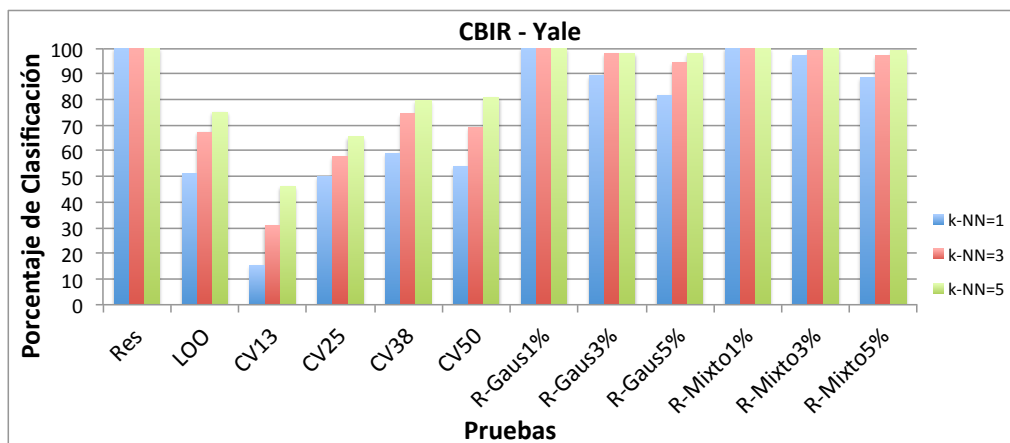


Figura 6.6: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos Yale

- Extended Yale B. Esta base de datos se redujo a una cantidad de 162 imágenes, contiene 27 personas (*clases*) y 6 imágenes por persona, las cuales cumplen mejor la descripción de condiciones ideales.

Ext Yale			
Prueba	k-NN=1	k-NN=3	k-NN=5
Res	100	100	100
LOO	19.75308642	35.80246914	47.5308642
CV17	7.407407407	40.74074074	51.85185185
CV33	9.259259259	31.48148148	50
CV50	7.407407407	18.51851852	32.09876543
R-Gaus1 %	91.30434783	96.89440994	98.13664596
R-Gaus3 %	78.88198758	90.0621118	94.40993789
R-Gaus5 %	64.59627329	81.36645963	88.19875776
R-Mixto1 %	100	100	100
R-Mixto3 %	99.37888199	100	100
R-Mixto5 %	93.16770186	97.51552795	97.51552795

Tabla 6.4: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos Extended Yale

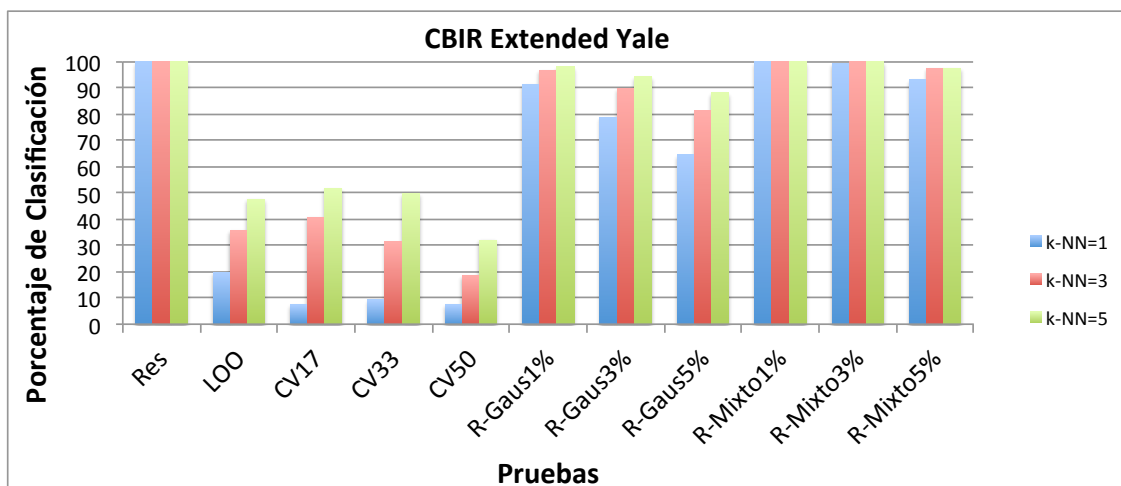


Figura 6.7: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos Yale

- BioID. Esta base de datos se redujo a una cantidad de 138 imágenes, contiene 23 personas (*clases*) y 6 imágenes por persona, las cuales cumplen mejor la descripción de condiciones ideales.

BioID			
Prueba	k-NN=1	k-NN=3	k-NN=5
Res	100	100	100
LOO	81.88405797	92.75362319	94.20289855
CV17	73.91304348	82.60869565	82.60869565
CV33	69.56521739	78.26086957	80.43478261
CV50	50.72463768	62.31884058	68.11594203
R-Gaus1 %	97.10144928	97.10144928	99.27536232
R-Gaus3 %	85.50724638	91.30434783	93.47826087
R-Gaus5 %	73.91304348	80.43478261	86.95652174
R-Mixto1 %	100	100	100
R-Mixto3 %	97.10144928	98.55072464	98.55072464
R-Mixto5 %	94.92753623	97.82608696	98.55072464

Tabla 6.5: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos BioID

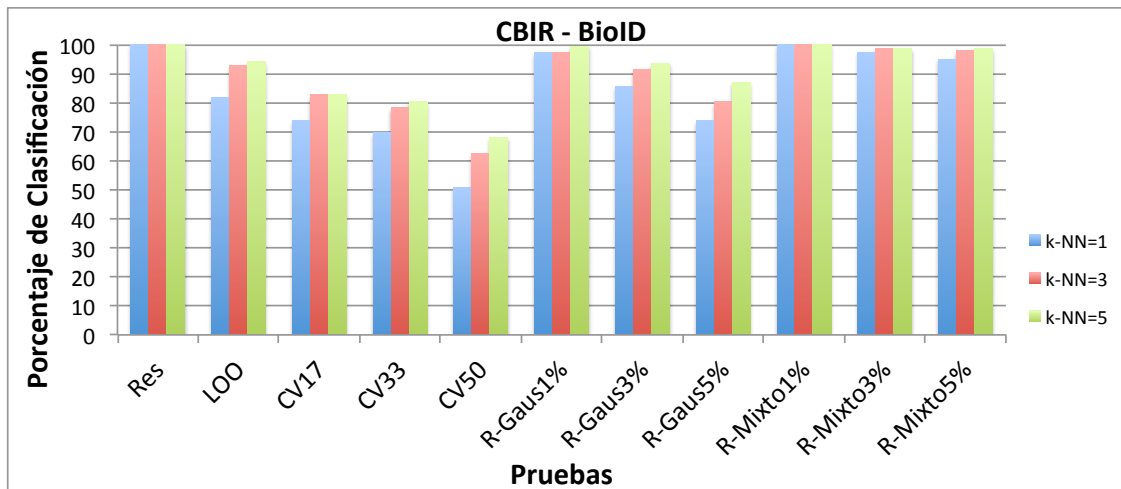


Figura 6.8: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos BioID

Como resultado de la técnica CBIR para las pruebas mencionadas anteriormente tenemos los resultados obtenidos para cada una de dichas pruebas como se muestra en la Figura 6.9, donde observamos que se logra una buena clasificación ya que la imagen consultada es asignada a la clase a la que corresponde.

El resultado de la técnica CBIR en la prueba de resustitución para imágenes contaminadas con ruido se puede observar en la Figura 6.10, en donde podemos observar que también existe una buena clasificación para la imagen de consulta contaminada con los diferentes valores de ruido, descritos en la sección anterior, para cada una de las imágenes de consulta se muestran las 5 primeras imágenes más parecidas, y que representan una clase correspondiente.



Figura 6.9: (a) Resultado resustitución, (b) Resultado quitar un elemento (LOO), (c) Resultado validación cruzada al 17% ,(d) Resultado validación cruzada al 33%, (e) Resultado validación cruzada al 50%



Figura 6.10: (a) Resultado ruido gaussiano al 1%, (b) Resultado ruido gaussiano al 3%, (c) Resultado ruido gaussiano al 5%, (d) Resultado ruido mixto al 1%, (e) Resultado ruido mixto al 3%, (f) Resultado ruido mixto al 5%

6.2. Pruebas SURF

Para trabajar en la técnica SURF, se recortaron las caras de las personas y se realizó un ajuste al tamaño de las caras para que estas imágenes tengan el mismo tamaño y así asegurarse que las comparaciones de cada una de las rejillas coincidan en la zona del rostro que se desea analizar. Los resultados obtenidos para las pruebas SURF para cada una de las bases de datos son:

- AR.

AR			
Prueba	k -NN=1	k -NN=3	k -NN=5
R-Gaus1 %	100	100	100
R-Gaus3 %	100	100	100
R-Gaus5 %	100	100	100
R-Mixto1 %	100	100	100
R-Mixto3 %	100	100	100
R-Mixto5 %	100	100	100

Tabla 6.6: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos AR con la técnica SURF

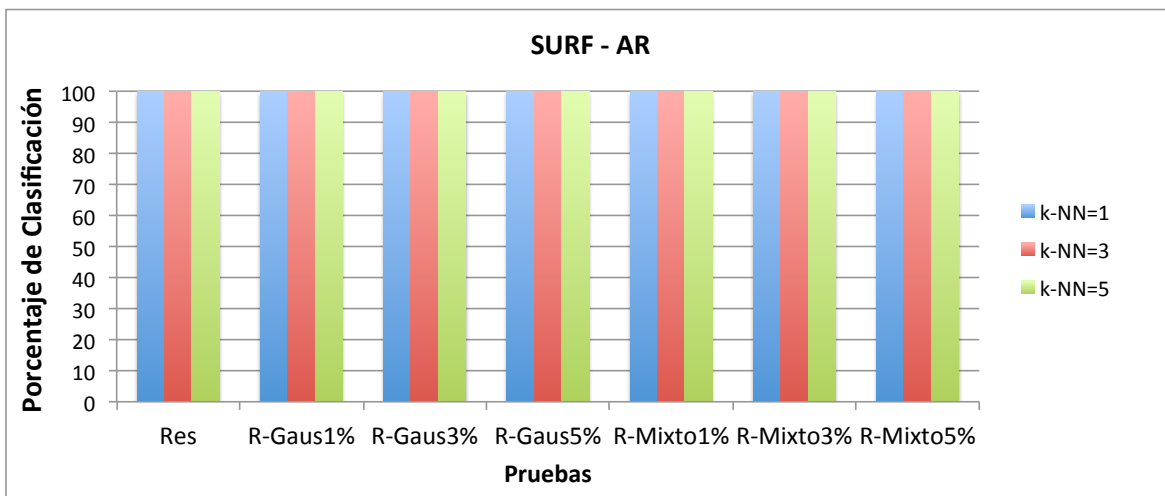


Figura 6.11: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos AR con la técnica SURF

- MUCT. Esta base de datos se redujo a una cantidad de 1368 imágenes, contiene 228 personas (*clases*) y 6 imágenes por persona, las cuales cumplen mejor la descripción de condiciones ideales.

MUCT			
Prueba	k-NN=1	k-NN=3	k-NN=5
R-Gaus1 %	100	100	100
R-Gaus3 %	100	100	100
R-Gaus5 %	100	100	100
R-Mixto1 %	100	100	100
R-Mixto3 %	100	100	100
R-Mixto5 %	100	100	100

Tabla 6.7: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos MUCT con la técnica SURF

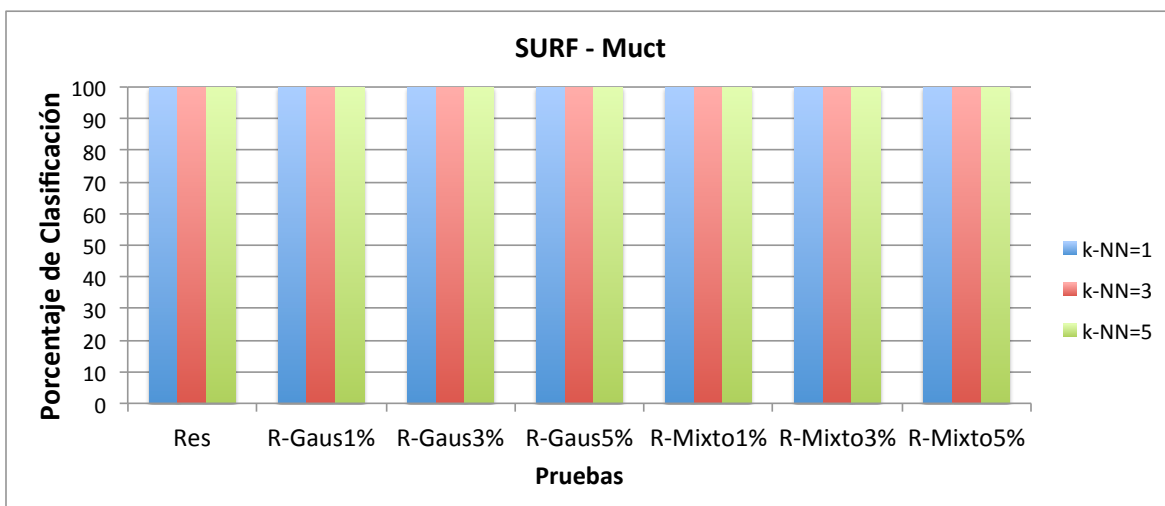


Figura 6.12: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos MUCT con la técnica SURF

- Yale. Esta base de datos se redujo a una cantidad de 104 imágenes, contiene 13 personas (*clases*) y 8 imágenes por persona, las cuales cumplen mejor la descripción de condiciones ideales.

Yale			
Prueba	k-NN=1	k-NN=3	k-NN=5
R-Gaus1 %	100	100	100
R-Gaus3 %	100	100	100
R-Gaus5 %	100	100	100
R-Mixto1 %	100	100	100
R-Mixto3 %	100	100	100
R-Mixto5 %	100	100	100

Tabla 6.8: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos Yale con la técnica SURF

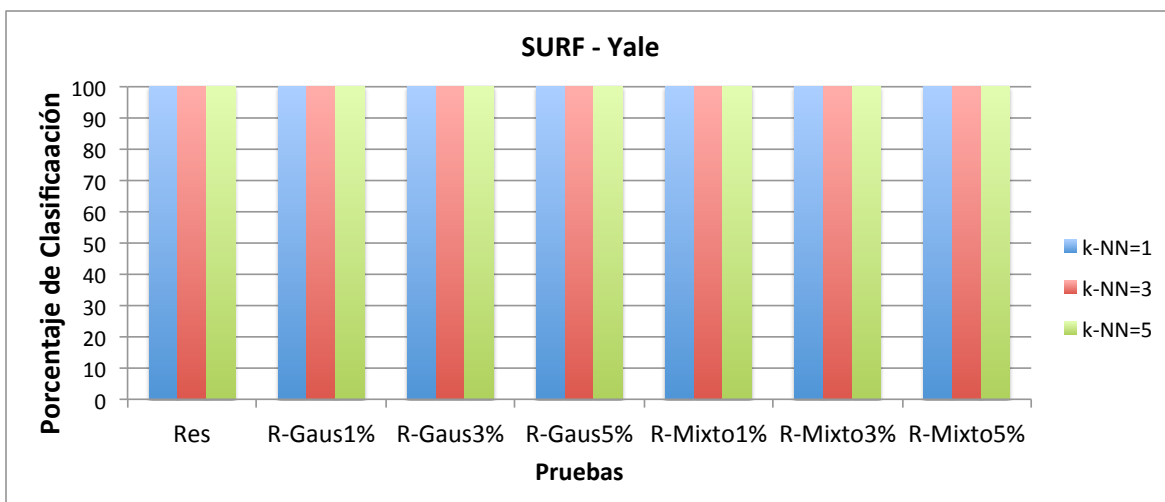


Figura 6.13: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos Yale con la técnica SURF

- Extended Yale B. Esta base de datos se redujo a una cantidad de 162 imágenes, contiene 27 personas (*clases*) y 6 imágenes por persona, las cuales cumplen mejor la descripción de condiciones ideales.

Ext Yale			
Prueba	k-NN=1	k-NN=3	k-NN=5
R-Gaus1 %	100	100	100
R-Gaus3 %	100	100	100
R-Gaus5 %	100	100	100
R-Mixto1 %	100	100	100
R-Mixto3 %	100	100	100
R-Mixto5 %	100	100	100

Tabla 6.9: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos Extended Yale con la técnica SURF

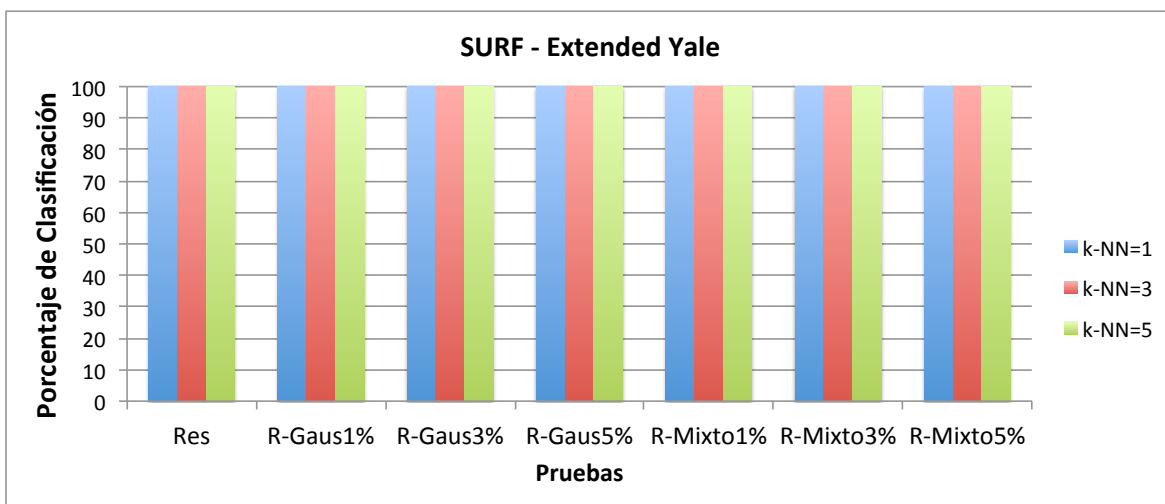


Figura 6.14: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos Extended Yale B con la técnica SURF

- BioID. Esta base de datos se redujo a una cantidad de 138 imágenes, contiene 23 personas (*clases*) y 6 imágenes por persona, las cuales cumplen mejor la descripción de condiciones ideales.

BioID			
Prueba	k-NN=1	k-NN=3	k-NN=5
R-Gaus1 %	100	100	100
R-Gaus3 %	100	100	100
R-Gaus5 %	100	100	100
R-Mixto1 %	100	100	100
R-Mixto3 %	100	100	100
R-Mixto5 %	100	100	100

Tabla 6.10: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos BioID con la técnica SURF

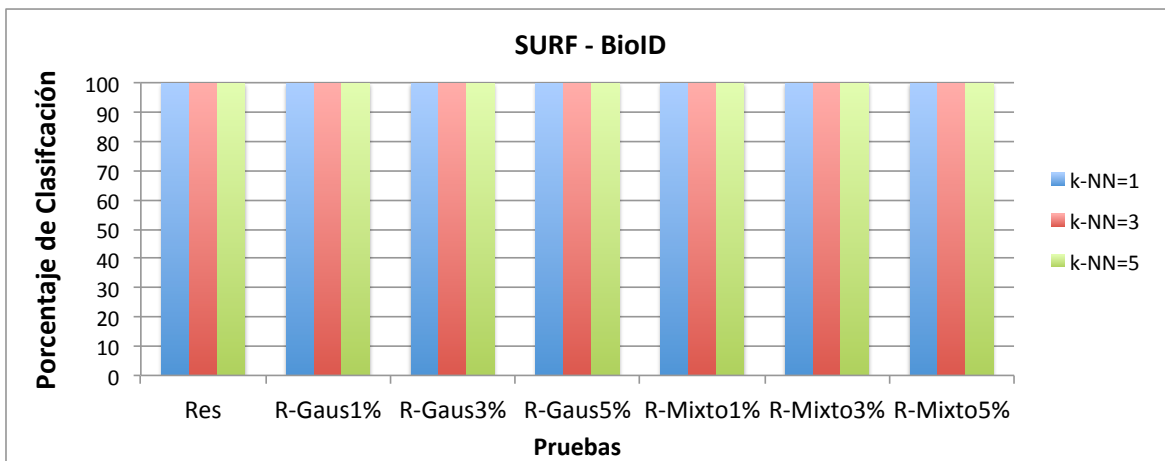


Figura 6.15: Porcentaje de clasificación obtenido para cada una de las pruebas señaladas para la base de datos BioID con la técnica SURF

El resultado de la técnica SURF en la prueba de resustitución para imágenes contaminadas con ruido se puede observar en la Figura 6.16, en donde podemos observar que también existe una excelente clasificación para la imagen de consulta contaminada con los diferentes valores de ruido. La Figura 6.16 (a)

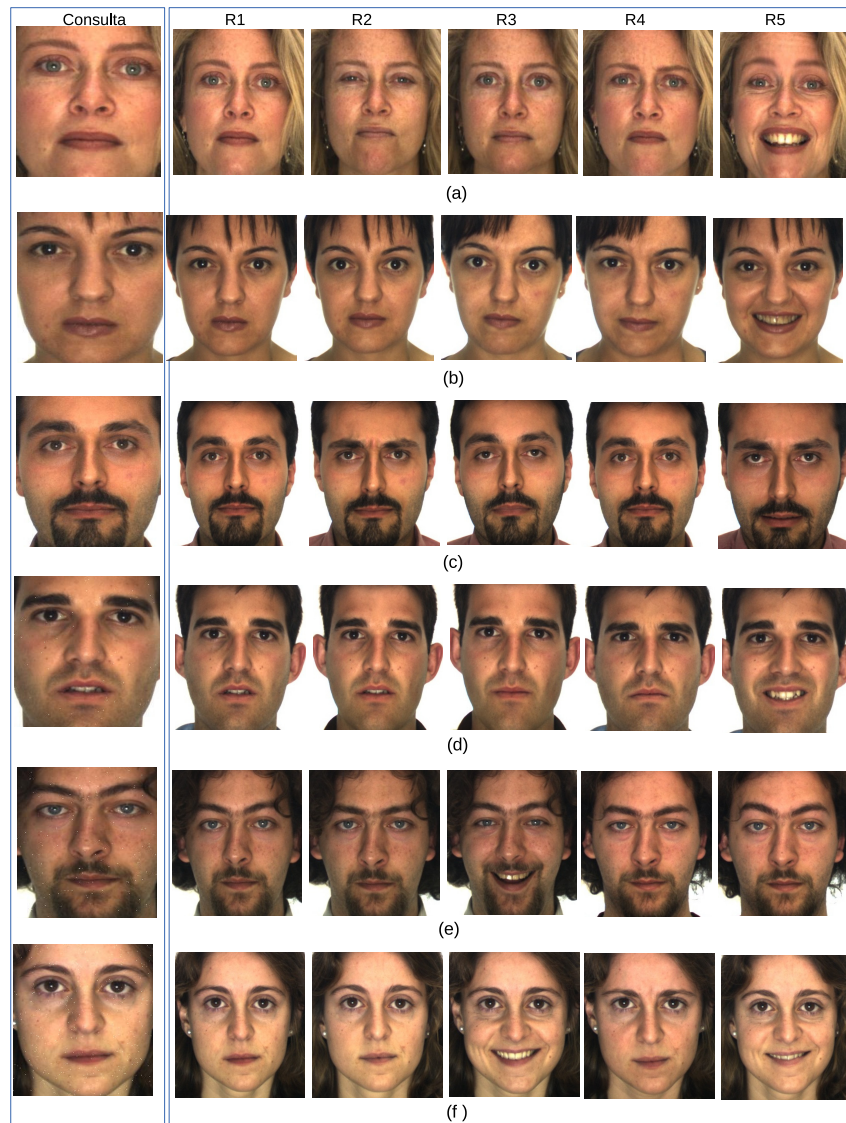


Figura 6.16: (a) Resultado ruido gaussiano al 1 %, (b) Resultado ruido gaussiano al 3 %, (c) Resultado ruido gaussiano al 5 %, (d) Resultado ruido mixto al 1 %, (e) Resultado ruido mixto al 3 %, (f) Resultado ruido mixto al 5 %

6.3. Rendimiento del sistema paralelo

El desempeño del clasificador utilizando la técnica CBIR se puede ver reflejado en la Figura 6.17.

1. Se observó que el tiempo de ejecución tuvo una reducción considerable mientras aumentaba el número de procesos Figura 6.17 (a).
2. En la Figura 6.17 (b) observamos el factor de aceleración (FA) ecuación 6.1 el cual va aumentando de manera lineal hasta el punto 6 donde se observa una caída y se vuelve a incrementar a partir del proceso 9.

$$FA = \frac{T_{Secuencial}}{T_p(N)} \quad (6.1)$$

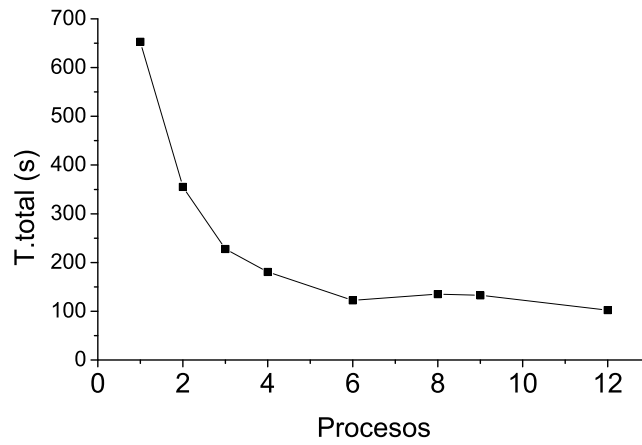
3. Por último en la Figura 6.17 (c) se observa la eficiencia la cual es calculada como:

$$Eficiencia = \frac{FA}{T_p(N)} \quad (6.2)$$

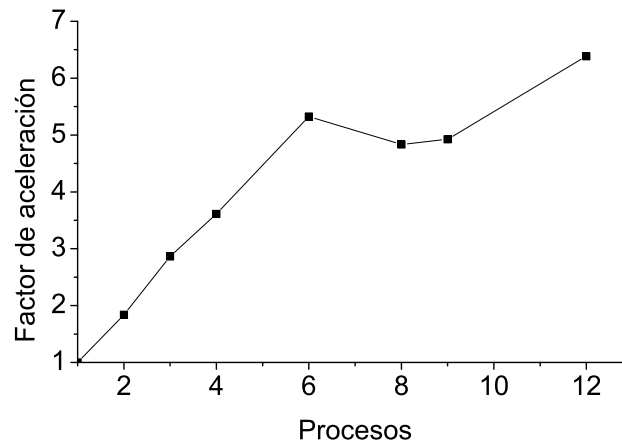
En donde se observa que la eficiencia decrece de manera considerable a partir de el proceso 6 y se observa que después del proceso 9 esta se mantiene estable.

Aunque en el análisis del desempeño se observa que a partir del proceso 6 el rendimiento decrece un poco, tanto en el factor de aceleración como en la eficiencia, por lo que parece que la máquina simula tener 12 núcleos y que se trata de un sistema *Hyper Threading*, por lo que tendría 6 microprocesadores reales y por este motivo, se observa un decremento en la evaluación de nuestro sistema dentro de esta máquina.

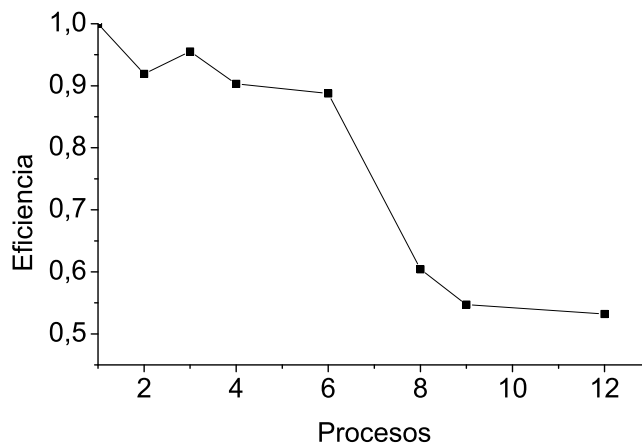
Para delimitar el trabajo solo se implementó la versión paralela sobre la técnica CBIR. Pero como las metodologías utilizadas para el entrenamiento de CBIR y SURF siguen la misma línea de desarrollo, se intuye que la versión paralela de la técnica SURF presentaría el mismo comportamiento que se encontró en CBIR.



(a)



(b)



(c)

Figura 6.17: (a) Tiempo paralelo vs procesos, (b) Factor de aceleración, (c) Eficiencia

Capítulo 7

Conclusiones y trabajo futuro

En este capítulo se muestra por un lado, el conjunto de conclusiones a las que se llegó después de del trabajo de investigación. Por otro lado se menciona el trabajo para las futuras investigaciones derivadas del desarrollo de esta tesis.

7.1. Conclusiones

Se propuso una nueva implementación a la técnica CBIR, utilizada ahora para el reconocimiento de rostros. Se tuvo que esta técnica genera muy buenos resultados cuando las imágenes con las que se trabaja cumplen con las condiciones ideales, como lo es la iluminación, la resolución y que las imágenes son de la misma sesión o tiempo, como en el caso de las bases de datos BioID y MUCT, no así en MEDS II puesto que algunas imágenes son tomadas en diferentes temporadas y las resoluciones de ellas hacen que el análisis de textura local tenga demasiadas variaciones.

Se pudo observar que la metodología CBIR al trabajar con imágenes con ruido, le pega directamente al desempeño de nuestro clasificador, ya que CBIR trabaja directamente con el contenido propio de la imagen (análisis de textura), una imagen alterada de esta manera, aunque sea trabajando con el conjunto fundamental, ya no es la misma imagen contenida en la base de datos, por lo que el sistema se confunde y asigna la imagen de consulta a una clase diferente de la esperada. Otro factor importante que se afecta al sistema es la captura de imágenes ya que al no estar en las mismas condiciones tanto de iluminación, como en distancia generan cambios drásticos en la extracción de

características, las cuales aunque sea la misma persona, pero las condiciones de obtención de imágenes varia, para nuestro sistema ya no es la misma persona.

Se llevo a deducir que CBIR funciona bien en escenarios naturales, ya que se comparan diferentes entornos: playas, bosques, montañas, etc. pero aplicado ala clasificación de rostros, tiene un cambio drástico ya que aquí se compara un mismo entorno en este caso rostros, por lo que llegar a una clasificación optima no es tan sencillo como lo es el clasificar escenarios naturales.

Se implementó la técnica SURF, utilizada ahora para el reconocimiento de rostros. Se tuvo que esta técnica genera muy buenos resultados aun cuando las imágenes con las que se trabaja no cumplan con algunas condiciones ideales, como lo es la iluminación, la resolución. Estas características no afectan el desempeño del clasificador ya que el análisis de la técnica SURF es invariante a escala, iluminación y rotación.

En el análisis de puntos SURF no existen los problemas encontrados en CBIR para la clasificación, por lo que es un buen complemento para nuestro sistema, ya que muestra tener excelentes resultados aunque sea con imágenes contaminadas, el problema que se detecto en esta técnica es que el costo computacional es demasiado alto ya que al segmentar una imagen en rejillas, la extracción de características de cada una de las rejillas tarda un tiempo significativo para la etapa de entrenamiento, por lo que si se quisiera trabajar con una cantidad mayor de imágenes, se tendría que optimizar el algoritmo paralelo para mejorar el desempeño del sistema.

Se propuso una arquitectura paralela del método CBIR que funciona en sistemas multicomputador. La arquitectura permitió que los procesamientos de entrenamiento de la consulta fueran distribuidos entre un conjunto de procesos, disminuyendo los tiempos hasta en un 85 %. La configuración que obtuvo el mejor rendimiento fue usando 6 procesos por nodo en un cluster de 1 nodo, para la base de datos que contenía la mayor cantidad de imágenes $MUCT = 1368$.

7.2. Trabajo futuro

Como trabajo futuro se planean dos tareas inmediatas: *(i)* aumentar el número de imágenes para poder medir el desempeño de la metodología propuesta en grandes volúmenes de información, y *(ii)* el algoritmo esta diseñado para trabajar dentro de un servidor por lo que queda implementar nuestro sistema dentro del laboratorio de su-

percomputo de la UAM-Iztapalapa y calcular el rendimiento de nuestro sistema dentro de un servidor real con problemas reales de comunicación y manejo de recursos.

Con base a las estimaciones experimentales acorde con las bases de datos hasta ahora probadas, será interesante reportar el desempeño de reducción de tiempo de cómputo con bases mucho mayores, sobre la versión de programación paralela, y estudiar el nivel de recuperación sobre un número mucho mayor de imágenes ya integrando la arquitectura paralela de SURF y generando un clasificador compuesto entre ambas técnicas.

Bibliografía

- [1] A-González and F-Prieto. Extracción de puntos característicos del rostro para medidas antropométricas. *Revista Ingenierías Universidad de Medellín*, 9(17):139–150, julio-diciembre 2010.
- [2] A-K-Jain. Automatic face recognition: State of the art. *Distinguished Lecture Series*, pages 0–44, Septiembre 2010.
- [3] Brendan Klare Anil K. Jain. Face recognition: Some challenges in forensics. *IEEE International Conference on Automatic Face and Gesture Recognition - FGR*, pages 726–733, 2011.
- [4] B-Amarapur and N-Patil. The facial texture extraction for face recognition based on geometrical approach. *IEEE CCECE/CCGEI*, pages 1936–1939, Mayo 2006.
- [5] B-P-Luis. Reconocimiento facial basado en puntos característicos de la cara en entornos no controlados. Master’s thesis, Universidad Autónoma de MADrid, Enero 2013.
- [6] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):711–720, July 1997.
- [7] Michał Bereta, Witold Pedrycz, and Marek Reformat. Local descriptors and similarity measures for frontal face recognition: A comparative analysis. *Journal of Visual Communication and Image Representation*, 24(8):1213–1231, 11 2013.
- [8] Sing-Tze Bow. *Pattern Recognition and Image Preprocessing*. Number 77 in Electrical Engineering and Electronics. Dekker, 1992.

- [9] Sergio Damas, Oscar Cordón, Oscar Ibáñez, Jose Santamaría, Inmaculada Alemán, Miguel Botella, and Fernando Navarro. Forensic identification by computer-aided craniofacial superimposition: A survey. *ACM Comput. Surv.*, 43(4):27:1–27:27, October 2011.
- [10] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):5:1–5:60, May 2008.
- [11] T de la Luz-Raul. Sistema de reconocimiento facial por medio de eigenfaces y redes neuronales. Master’s thesis, Esime IPN, junio 2009.
- [12] J.P.M. de Sa. *Pattern Recognition: Concepts, Methods, and Applications*. Springer, 2001.
- [13] Philippe Dreuw, Pascal Steingrube, Harald Hanselmann, and Hermann Ney. Surface: Face recognition under viewpoint consistency constraints. In *Proc. BMVC*, pages 7.1–7.11, 2009. doi:10.5244/C.23.7.
- [14] E.Cuevas and D.Zaldívar. *Procesamiento digital de imágenes usando MatLAB y Simulink*. Alfaomega, 1ra edition, Septiembre 2010.
- [15] Keinosuke Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
- [16] G-G-Yen and N-Nithianandan. Facial feaature extraction using genetic algorithm. *Congress on Evolutionary Computation*, 2:1895–1900, Mayo 2002.
- [17] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):643–660, June 2001.
- [18] Wintz González García. *Digital Image Processing*. Prentice Hall, 2002.
- [19] H-T-F-Rodes and Alphonse Bertillon. *Father of Scientific Detection*. 1956.
- [20] J-S-Pérez. Localización y reconocimiento de rostros en imágenes monoculares de frente con variación en escala. Master’s thesis, Universidad Autónoma Metropolitana-Cuajimalpa, 2009.

- [21] J-Vogel and B-Schiele. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 72(2):133–157, 2007.
- [22] Oliver Jesorsky, Klaus J. Kirchberg, and Robert W. Frischholz. Robust face detection using the hausdorff distance. pages 90–95. Springer, 2001.
- [23] J.F.Serrano-Talamantes, C.Aviles-Cruz, J.Villegas-Cortez, and J.H.Sossa-Azuela. Self organizing natural scene image retrieval. *Expert Systems with Applications: An International Journal*, 40(7):2398–2409, 2012.
- [24] K-A-Kim. Facial feature extraction using pca and wavelet multi-resolution images. *sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 17–19, Mayo 2004.
- [25] A Carolina Pinto L. Segmentación de imágenes por textura. Technical report, Universidad de Concepción, Septiembre 2006.
- [26] L.Fei-Fei, R.Fergus, and P.Perona. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories [en línea], disponible: <http://www.vision.caltech.edu/image-datasets/caltech101>, 2004.
- [27] Gonzalo Pajares M. and Jesús M. de la Cruz G. *Visión por computador imágenes digitales y aplicaciones*. Alfaomega, 2 edition, Marzo 2008.
- [28] M-H-Mahoor and M-Abdel-Mottaleb. Facial features extraction in color images using enhanced active shape model. *7th International Conference on Automatic Face and Gesture Recognition(FGR'06)*, pages 10–12, 2006.
- [29] M-Gavrilova M-Sultana. A content based feature combination method for face recognition. 226:197–206, 2013.
- [30] A.M. Martinez and R. Benavente. The ar face database. CVC Technical Report 24, The Ohio State University, junio 1998.
- [31] Rakesh Mehta, Jirui Yuan, and Karen Egiazarian. Face recognition using scale-adaptive directional and textural features. *Pattern Recognition*, 47(5):1846–1858, 5 2014.

- [32] Cyntia Mendoza-Martinez, JesusCarlos Pedraza-Ortega, and JuanManuel Ramos-Arreguin. A novel approach for face authentication using speeded up robust features algorithm. In Alexander Gelbukh, FlixCastro Espinoza, and Sofía N. Galicia-Haro, editors, *Human-Inspired Computing and Its Applications*, volume 8856 of *Lecture Notes in Computer Science*, pages 356–367. Springer International Publishing, 2014.
- [33] S. Milborrow, J. Morkel, and F. Nicolls. The MUCT Landmarked Face Database. *Pattern Recognition Association of South Africa*, 2010. <http://www.milbo.org/muct>.
- [34] B.Kotaiah R.A. Khan O. Nagaraju, M.RamiReddy. Implementing and compiling clustering using mac queens alias k-means apriori algorithm. *International Journal of Database Management Systems*, 2012.
- [35] S. Prasad, A. Sawant, R. Shettigar, and S. Sinha. Real-time face recognition system with dynamic training and enhanced multi-algorithm face recognition. In *Proceedings of the International Conference & Workshop on Emerging Trends in Technology*, ICWET '11, pages 119–122, New York, NY, USA, 2011. ACM.
- [36] William Prat. *Digital Image Processing*. John Wiley and Sons, 3 edition, 2001.
- [37] T-F-Cootes. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.
- [38] José Felix Serrrano Talamantes. *Recuperación de imágenes mediante rasgos descriptores globales y locales*. PhD thesis, Instituto Politécnico Nacional, Enero 2011.
- [39] V-S-Nalwa. *A Guided Tour of Computer Vision*, volume 1 of *TA1632*. Addison Wesley, 1993.
- [40] Enrico Vezzetti and Federica Marcolin. Geometrical descriptors for human face morphological analysis and recognition. *Robotics and Autonomous Systems*, 60(6):928–939, 6 2012.
- [41] Andrew P. Founds; Nick Orlans; Whiddon Genevieve; Craig I. Watson;. Nist special databse 32 - multiple encounter dataset ii (meds-ii). *NIST Interagency/Internal Report (NISTIR) - 7807*, July 2011.

- [42] Y-Yokogawa. A proposal of improved lip contour extraction method using deformable template matching and its application to dental treatment. *Systems and Computers in Japan*, 38(5):80–89, 2007.
- [43] Y-Zaho. Part-based pca for facial feature extraction and classification. *International Workshop in Haptic Audio visual Environments and Games*, pages 99–104, Noviembre 2009.

Anexo 1

Glosario de términos

A continuación se muestran los conceptos básicos para entender la manipulación de imágenes digitales, iniciando por el análisis de la disposición de los píxeles que la conforman, su métrica, región, etc.:

Métrica

Expresa la proximidad entre dos objetos:

Distancia entre dos puntos (distancia Euclidiana), distancia de un punto a una recta, distancia de un punto a un plano.

Ruta (Path). Se define como el camino entre dos puntos (pasos) moviendo entre vecindades.

Región conectada. $b(x, y) \in R$ es un conjunto de puntos $p(x, y)$, tales que existe un camino que comienza en p_1 y termina en p_2 para todo $p_1, p_2 \in R$.

Vecindades de un Pixel

Un pixel p de coordenadas (x, y) tiene dos vecinos horizontales y dos vecinos verticales cuyas coordenadas están dadas por:

$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)$$

Este conjunto se llama: vecindad-4 de p y se denota como $N_4(p)$, véase la Figura 1.1.

Un pixel p de coordenadas (x, y) tiene 4 vecinos diagonales cuyas coordenadas están dadas por:

		2		
	2	1	2	
2	1	0	1	2
	2	1	2	
		2		

Figura 1.1: Vecindad $N_4(p, q) = |x - s| + |y - t|$.

$$(x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1)$$

Se denota como $N_8(p)$. A estos puntos junto con los de vecindad-4, se les llama vecindad-8 de p , como se muestra en la Figura 1.2.

2	2	2	2	2
2	1	1	1	2
2	1	0	1	2
2	1	1	1	2
2	2	2	2	2

Figura 1.2: Vecindad $N_8(p, q) = \max(|x - s|, |y - t|)$.

1.0.1. Información estadística de una imagen.

Entendemos por información estadística de una imagen digital, al cúmulo de valores extraídos a partir de la información de la conformación de la propia imagen (histograma, media, varianza, desviación estándar) [18][15]. A continuación se mencionan las definiciones de cómo se definen los valores estadísticos más usados en el campo del procesamiento digital de imágenes.

1. Histograma.

El histograma de una imagen digital con niveles de gris en el rango $0, L - 1$ es una función discreta $h(r_k) = n_k/L$ donde r_k es el nivel de gris k -ésimo y n_k es el número de píxeles en la imagen con nivel de gris r_k , como se ejemplifica en la Figura 1.3. Con un simple vistazo, puede proporcionarnos una idea muy aproximada de la distribución de niveles de gris de la imagen, el contraste y el brillo que presenta [14].

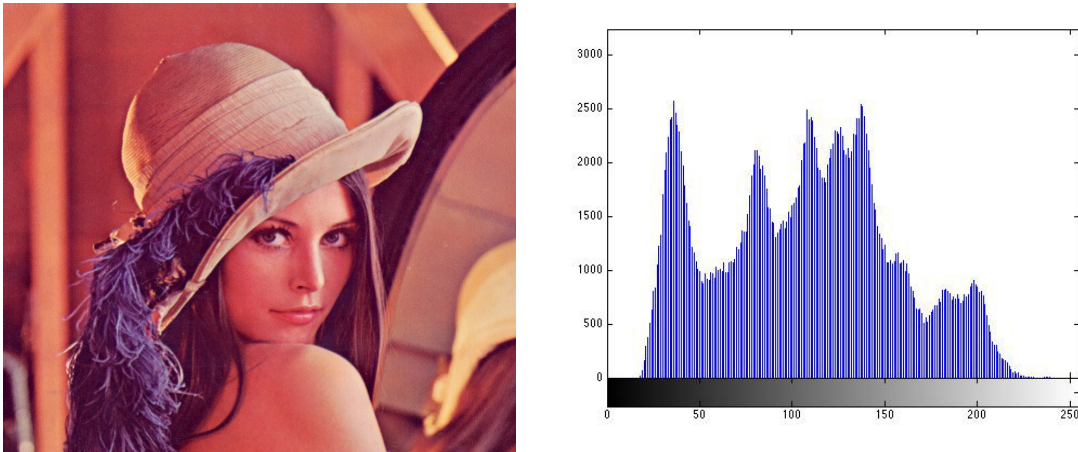


Figura 1.3: Histograma de un rostro Pre-procesado.

2. Media.

Consiste en generar una nueva imagen $g(x, y)$ cuya intensidad para cada pixel se obtiene promediando los valores de intensidad de los pixeles $f(x, y)$ incluidos en un entorno de vecindad predefinido.

$$\bar{f}(x, y) = \frac{1}{n} \sum_{k=1}^n f_k(x, y) \quad (1.1)$$

3. Varianza. Una medida simple del contraste de la imagen. Es una medida de la diferencia existente entre todos los valores de la muestra y la media, se define matemáticamente como el promedio del cuadrado de esas diferencias.

$$\alpha^2(x, y) = \frac{1}{n-1} \sum_{k=1}^n (f_p(x, y) - f_k(x, y))^2 \quad (1.2)$$

4. Desviación estándar.

$$\alpha(x, y) = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (f_p(x, y) - f_k(x, y))^2} \quad (1.3)$$

1.0.2. Acondicionado de la imagen.

Entendemos al acondicionamiento de la imagen como la adecuación de la misma, su preprocesamiento, acorde a un fin específico.

■ Ruido en imágenes.

El ruido en una imagen es cualquier entidad o resultado intermedio no interesante para los propósitos del cómputo principal. El ruido puede resultar de varias causas:

- La óptica y electrónica asociadas al sistema de adquisición.
- Las condiciones de iluminación.
- Las sombras creadas entre objetos.
- Las condiciones atmosféricas.

■ Algunos tipos de ruido y su ejemplo en la Figura 1.4.

- **Ruido aditivo.** $\tilde{f}(x, y)$ es obtenida a partir de $f(x, y)$, en este caso, al adicionar al azar a cada pixel de $f(x, y)$ un real c , tal que:

$$\tilde{f}(x, y) = f(x, y) + c \quad (1.4)$$

- **Ruido substractivo.** $\tilde{f}(x, y)$ es obtenida a partir de $f(x, y)$, en este caso, al restar al azar a cada pixel de $f(x, y)$ un real c , tal que:

$$\tilde{f}(x, y) = f(x, y) - c \quad (1.5)$$

- **Ruido mezclado.** $\tilde{f}(x, y)$ es obtenida a partir de $f(x, y)$, en este caso, al sumar o restar al azar a cada pixel de $f(x, y)$ un real c , tal que se tenga uno de los casos anteriores.
- **Ruido gaussiano.** Este tipo de ruido cumple con una distribución gaussiana, en otras palabras, la variable c toma valores como se observa en la ecuación 1.6 [18] véase la Figura 1.4:

$$c = (L - 1) * e^{\frac{-z^2}{2\sigma^2}} \quad (1.6)$$

con $z \geq 0$ real, y σ es el parámetro de amplitud de la Gaussiana.

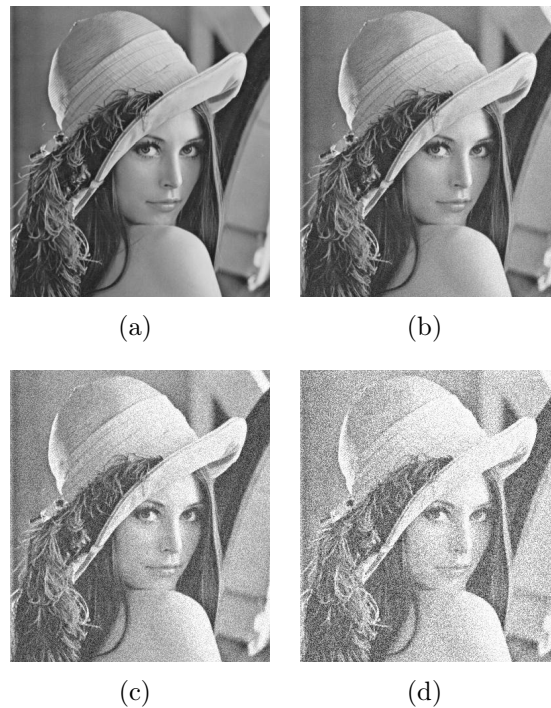


Figura 1.4: Ejemplo de una imagen (a) sin ruido, (b) con ruido gaussiano al 10 %, (c) con ruido gaussiano al 30 %, (d) con ruido gaussiano al 50 %.

- **Ruido sal.** Se caracteriza por saturar el valor de uno o más píxeles de la imagen hacia el límite superior[18] Figura 1.3 b).

$$\tilde{f}(x, y) = f(x, y) + b * c \quad (1.7)$$

$$\text{con } b = 1 \text{ y } c = L - 1$$

- **Ruido pimienta.** Se caracteriza por saturar el valor de uno o más píxeles de la imagen hacia el límite inferior[18] Figura 1.5 c).

$$\tilde{f}(x, y) = f(x, y) + b * c \quad (1.8)$$

$$\text{con } b = -1 \text{ y } c = 0$$



Figura 1.5: ejemplo de una imagen digital: (a) sin ruido, (b) contaminada con ruido sal, (c) contaminada con ruido pimienta.

■ El filtrado de una imagen.

Las imágenes al ser adquiridas a través de un medio óptico o electrónico es susceptible a ser degradada por el ambiente de captación.

El objetivo de las técnicas de filtrado consiste en reducir o eliminar, en el mejor de los casos, el ruido presente en una imagen.

Debido a que el modelado del ruido es algo muy complicado, en el mejor de los casos lo que se logra es una reducción del ruido presente en la señal.

Este puede llevarse a cabo en el *dominio de la frecuencia*, como en el *dominio espacial*.

● Filtrado en el dominio de la frecuencia

1. Se calcula la transformada de Fourier de la imagen a ser filtrada,
2. luego el resultado es multiplicado por una función de transferencia de filtrado adecuada.
3. Finalmente, al resultado obtenido se le calcula la transformada inversa de Fourier para obtener la imagen mejorada.

● Filtrado en el dominio espacial.

- En este caso se trabaja en el mismo dominio de los pixeles al cambiar el valor de un pixel tomando en cuenta el valor de dicho pixel y el de alguno de sus vecinos.
- → fácil y rápido de implementar.

■ Filtros lineales.

La operación de un filtro en el dominio espacial puede expresarse como:

$$s(x, y) = T[f(x, y)] \quad (1.9)$$

con f la imagen dada, s la imagen transformada y T es el operador sobre f , definido en una vecindad de (x, y) .

La vecindad puede ser cualquiera, las más usadas son la cuadrada Figura 1.6 y la circular, ambas con centro en (x, y) .

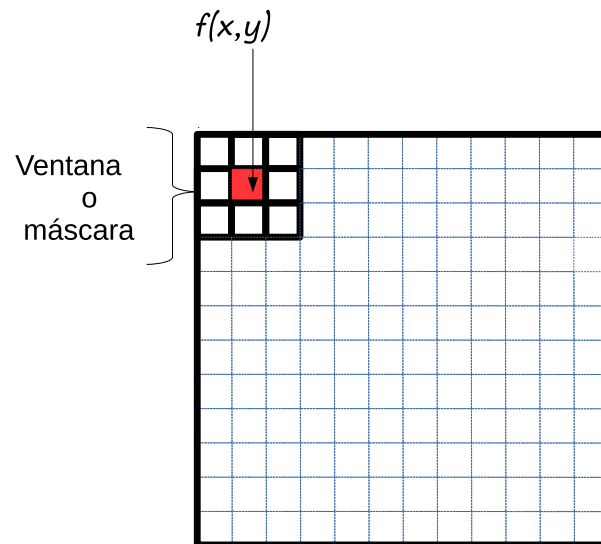


Figura 1.6: Ejemplo de una ventana o máscara cuadrada de 3x3 aplicada a una imagen.

El efecto de aplicar una operación T dentro de una ventana deslizante e.g 3x3, no es más que situar la ventana dentro de la imagen e irla recorriendo:

1. Se sitúa la ventana dentro de la imagen y se aplica el operador T al pixel del centro como se mostró en la Figura 1.6,
2. se desliza la venta una columna hacia la derecha y se ejecuta el operador de nuevo,

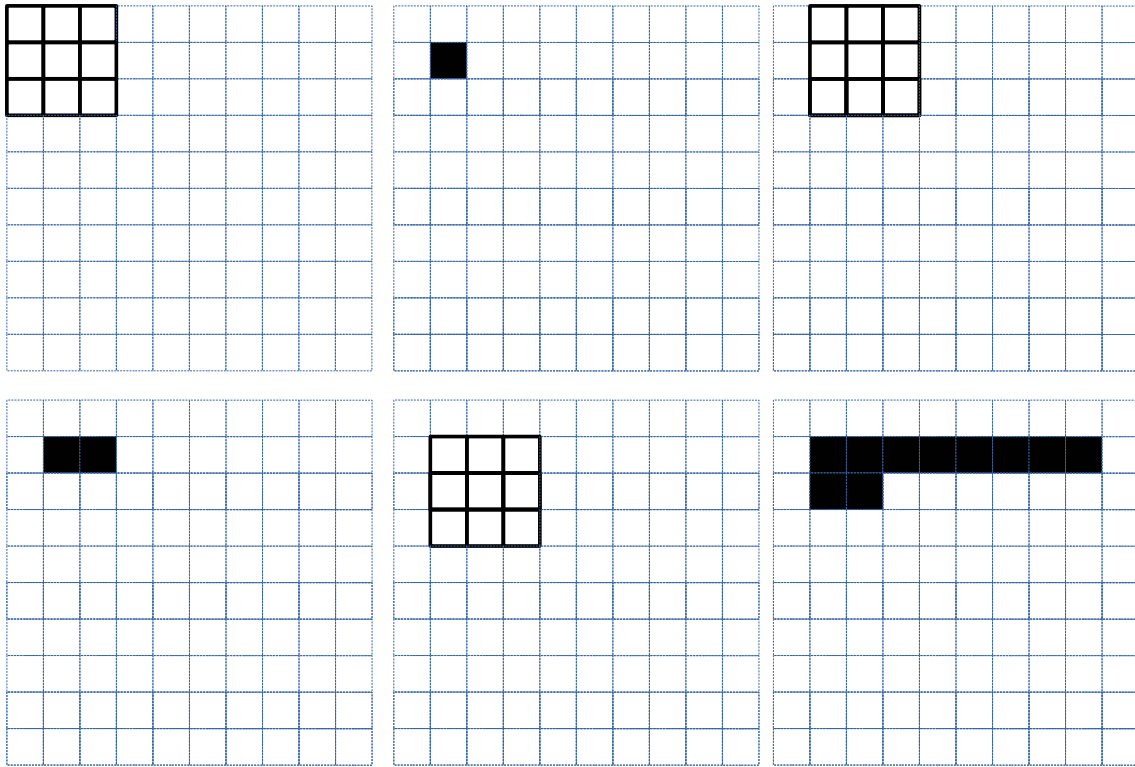


Figura 1.7: Efecto de la aplicación de *T-Ventana deslizante*.

3. una vez que el borde de la ventana llega al final de la imagen, esta se pasa al siguiente renglón Figura 1.7,
4. cuando los bordes de la ventana llega al último renglón y a la última columna, se termina de aplicar el operador T y se tiene la imagen procesada.

Los principales objetivos de la aplicación de los *filtros* son:

- Suavizar la imagen.
- Eliminar ruido.
- Detectar bordes.

■ Tipos de filtros

- **Filtro promedio aritmético.** Estos tipo de filtros encuentran algún tipo de promedio sobre una ventana de $n \times n$ pixeles. El *filtro promedio aritmético* es el filtro más simple de ellos, halla el valor promedio de los pixeles dentro de la ventana ec. (1.10):

$$f_{pa} = \frac{1}{n^2} \sum_{(x,y) \in V} f(x,y) \quad (1.10)$$

El resultado de aplicar un filtro promedio aritmético, es el de “suavizar” variaciones locales en la señal \rightarrow filtro “pasa-bajos” Figura 1.8. Tiende a desenfocar la imagen \rightarrow mitiga los efectos de ruido.

Este es un filtro muy fácil de entender, nos ayuda a comprender la forma de aplicación de los filtros, y la formación y aplicación de la ventana o máscara. Pero no es un filtro eficiente:

- Las señales con frecuencias comunes a las del ruido son atenuadas.
- El ruido de tipo impulsivo es solo atenuado más no removido e.g. Ruido sal o ruido pimienta.



Figura 1.8: (a) imagen original, (b) imagen con ruido gaussiano al 20 %, (c) imagen filtrada con ventana 3x3 filtro promedio.

- **Filtro gaussiano.** Es un tipo especial de filtro promedio, cuyos pesos siguen el comportamiento de una gaussiana.

En el dominio de la frecuencia, la transformada de Fourier de un filtro promedio aritmético uno-dimensional de ancho $2W$ Figura 1.8 esta dada por:

$$\text{sinc}(\omega) = \frac{2\text{sen}(\omega W)}{\omega} a \quad (1.11)$$

- **Observaciones**

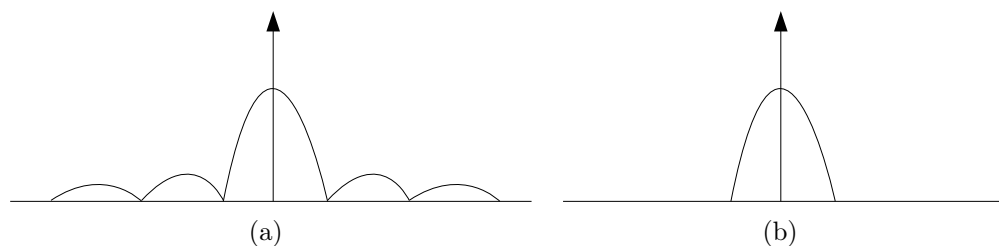


Figura 1.9: (a) Transformada de Fourier del núcleo promedio aritmético $1 - D$, (b) Transformada de Fourier del núcleo gaussiano $1 - D$.

Esto significa que convolucionar una imagen f con un núcleo gaussiano G es equivalente a convolucionar primero todas las filas de f , y después todas las columnas con un núcleo gaussiano uno-dimensional con la misma σ .

Una de las ventajas obvias de usar un núcleo uno-dimensional en lugar de su equivalente núcleo bidimensional es la reducción de tiempo obtenida.

o Construcción de núcleos gaussianos

Existen varias formas de calcular los coeficientes o pesos de un núcleo gaussiano. Una buena aproximación de los coeficientes de un núcleo gaussiano uno-dimensional se puede obtener a través de los coeficientes de la expansión binomial:

$$(1 + x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \dots + \binom{n}{n}x^n \quad (1.12)$$

→ equivalente a tomar la n -ésima fila del triángulo de pascal.

$$\begin{array}{ccccccc}
 & & & & 1 & & & & \\
 & & & & & 1 & & 1 & \\
 & & & & & & 1 & & 2 & & 1 \\
 & & & & & & & 1 & & 3 & & 3 & & 1 \\
 & & & & & & & & 1 & & 4 & & 6 & & 4 & & 1 \\
 & & & & & & & & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & & & & & & & & & & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 \\
 & & & & & & & & & & & & & & & & \cdot & & & & & & \\
 & & & & & & & & & & & & & & & & \cdot & & & & & &
 \end{array}$$

Así los coeficientes de un filtro uno-dimensional de tamaño cinco son:

$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

Esta máscara puede ser usada para alisar primero las filas de la imagen de entrada y luego las columnas.

Desventaja: no se tiene control sobre el parámetro σ .

- **Alternativa** Otra forma de diseñar un núcleo gaussiano consiste en calcular los pesos de la ventana al usar directamente la ecuación gaussiana (1.13).

$$G(x, y) = ce^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1.13)$$

Al escoger el valor de σ^2 se puede evaluar el núcleo para un tamaño de ventana n dado. e.g., para $\sigma = 2$ y $n = 7$ se tiene:

(x,y)	-3	-2	-1	0	1	2	3
-3	0.011	0.039	0.082	0.105	0.082	0.039	0.011
-2	0.039	0.135	0.287	0.368	0.287	0.135	0.039
-1	0.082	0.287	0.606	0.779	0.606	0.287	0.082
0	0.105	0.368	0.779	1.0	0.779	0.368	0.105
1	0.082	0.287	0.606	0.779	0.606	0.287	0.082
2	0.039	0.135	0.387	0.368	0.387	0.135	0.039
3	0.011	0.039	0.082	0.105	0.082	0.039	0.011

Es necesario re-escalar para que el valor más pequeño sea 1,0. para esto se calcula el valor: $k = \frac{1}{0,011} = 91$

1	4	7	10	7	4	1
4	12	26	33	26	12	4
7	26	55	71	55	26	7
10	33	71	91	71	33	10
7	26	55	71	55	26	7
4	12	26	33	26	12	4
1	4	7	10	7	4	1

Para evitar la saturación en los valores de la imagen resultante al aplicar el núcleo obtenido, se agrega un factor de normalización, que se obtiene

de la suma de todos los pesos del núcleo: $c = \sum_{h=-\frac{n}{2}}^{\frac{n}{2}} \sum_{k=-\frac{n}{2}}^{\frac{n}{2}} G(h, k) = 1115$

	1	4	7	10	7	4	1
	4	12	26	33	26	12	4
	7	26	55	71	55	26	7
$\frac{1}{1115}$	10	33	71	91	71	33	10
	7	26	55	71	55	26	7
	4	12	26	33	26	12	4
	1	4	7	10	7	4	1

Tabla 1.1: Filtro gaussiano con ventana de 7x7.



Figura 1.10: (a)Imagen original, (b)imagen contaminada con ruido gaussiano al 20%, (c)Imagen procesada con filtro gaussiano ventana 7x7.

- **Filtro promedio contra-harmónico**

Funciona muy bien en presencia de ruido sal o pimienta dependiendo del orden del filtro usado.

si \mathbf{R} es el orden del filtro entonces en la ec (1.14) tenemos:

$$f_{ca} = \frac{\sum_{(x,y) \in V} f(x,y)^{R+1}}{\sum_{(x,y) \in V} f(x,y)^R} \quad (1.14)$$

si $R < 0$, el filtro elimina el ruido tipo sal de la Figura 1.11 a) y se obtiene como resultado la Figura 1.11 c),

si $R > 0$, el filtro elimina el ruido tipo pimienta de la Figura 1.11 b) y se obtiene como resultado la Figura 1.11 d),

si $R=0$, realiza la media aritmética (*filtro promedio*).

Funciona de una u otra forma, pero no de forma simultánea.

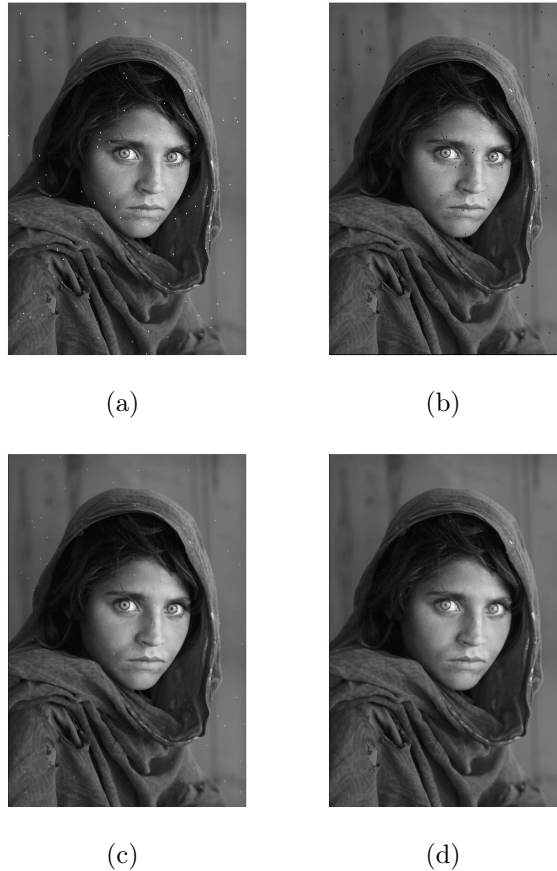


Figura 1.11: (a)Imagen con ruido sal, (b)imagen con ruido pimienta, (c)Filtro contra-harmónico con $R = -1$ para la Figura a, (d)filtro corta-harmónico con $R = 1$ para la Figura b.

- **Filtro promedio geométrico**

Funciona muy bien en presencia de ruido gaussiano, por encima del filtro promedio aritmético Figura 1.12.

$$f_{pg} = \prod_{(x,y) \in V} [f(x,y)]^{\frac{1}{n^2}} \quad (1.15)$$

- **Filtro promedio harmónico** Tiene un buen desempeño ante presencia de



Figura 1.12: (a)Imagen original, (b)Imagen con ruido gaussiano al 5 %, (c)Imagen con filtro promedio geométrico ventana 3x3.

ruido tipo sal Figura 1.13, pero no así con el ruido tipo pimienta.

$$f_{ph} = \frac{n^2}{\sum_{(x,y) \in V} \frac{1}{f(x,y)}} \quad (1.16)$$

Al igual que el filtro promedio geométrico, éste se comporta bien ante la presencia de ruido gaussiano, respetando también mejor el detalle de la imagen original que el filtro promedio aritmético Figura 1.13.



Figura 1.13: (a)Imagen original, (b)imagen contaminada con ruido SAL, (c)Imagen procesada con filtro Promedio Harmónico ventana 3x3.

- **Filtro Y_p**

$$f_{Y_p} = \left[\sum_{(x,y) \in V} \frac{f(x,y)^p}{n^2} \right]^{1/p} \quad (1.17)$$

- Remueve ruido tipo SAL para $p < 0$.
- Remueve ruido tipo PIMIENTA para $p > 0$

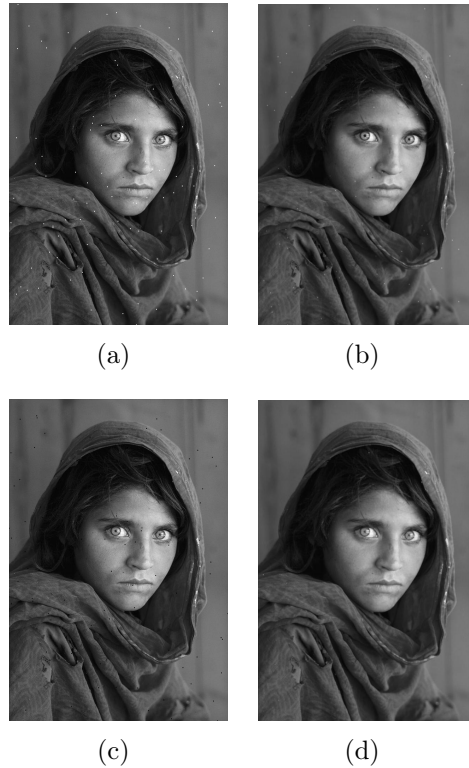


Figura 1.14: (a)Imagen contaminada con ruido SAL, (b)Imagen procesada con filtro Y_p ventana 7×7 y $p = -3$, (c)Imagen contaminada con ruido PIMIENTA, (d)Imagen procesada con filtro Y_p ventana 7×7 y $p = 3$.

- **Filtros laplacianos** Se usan comúnmente para resaltar los detalles en una imagen. Este tipo de filtros tiene a mejorar los detalles en todas las direcciones por igual. La Figura 1.15 muestra dos máscaras de 3×3 para un filtro del tipo Laplaciano que se puede aplicar a una imagen.

0	-1	0	1	2	1
-1	5	-1	-2	5	-2
0	-1	0	1	-2	1

Figura 1.15: Máscaras de 3×3 para resaltar detalles filtro Laplaciano.

- **Filtros diferenciadores.** Al contrario de los filtros Laplacianos, mejoran

los detalles de la imagen en la dirección especificada de acuerdo a la máscara seleccionada Figura 1.16

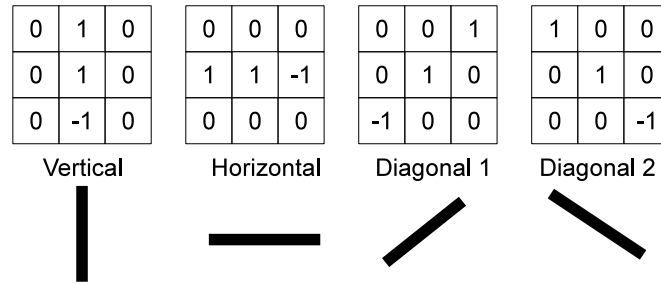


Figura 1.16: Máscaras de 3x3 para resaltar líneas filtro Diferenciador.

- **Filtros enfatizadores del altas frecuencias “High-Boost”.** Una imagen filtrada en sus altas frecuencias puede ser obtenida como la diferencia entre la imagen original y su versión pasa bajo.

$$f_a = f - f_b \quad (1.18)$$

Si se multiplica por un factor de amplificación A , tenemos un *filtro elevador* o *filtro enfatizador de altas frecuencias*.

$$f_e = A \cdot f - f_b \quad (1.19)$$

La ecuación (1.19) puede descomponerse como:

$$\begin{aligned} f_e &= A \cdot f - f_b \\ &= (A - 1)f + f - f_b \end{aligned}$$

$$f_e = (A - 1)f + f_a \quad (1.20)$$

Si $A = 1 \rightarrow$ da como resultado la versión pasa-alto estándar.

Si $A > 1 \rightarrow$ parte de la información de la imagen original es adicionada a la imagen filtrada en sus altas frecuencias \rightarrow permite restaurar parte de las componentes de bajas frecuencias perdidas durante el filtrado pasa-alto de la imagen \rightarrow mejora

$$\frac{1}{9} \begin{array}{|c|c|c|} \hline -1 & -1 & -1 \\ \hline -1 & P & -1 \\ \hline -1 & -1 & -1 \\ \hline \end{array}$$

Figura 1.17: Máscara de 3x3 para el filtro enfatizador.

los bordes de los objetos dependiendo el valor de A . El filtro elevador puede ser implementado por medio de la máscara ponderada que se muestra en la Figura 1.17 cuyo peso central puede ser obtenido como:

$$P = 9 \cdot A - 1 \text{ para } A \geq 0$$

El valor de A determina la naturaleza del filtro.

1.0.3. Modelos de color

El propósito de un modelo de color (también llamado espacio de color o sistema de colores) es facilitar la especificación de colores en alguna forma estándar de aceptación general. En esencia, un modelo de color es una especificación de un sistema coordinado y un subespacio dentro de ese sistema donde cada color se representa por un punto.

Los modelos de color que existen son: RGB, NTSC, YCbCr, HSV, CMY, CMYK y HSI (ver [13, 16]).

La mayoría de los modelos de color hoy en día están orientados ya sea al hardware (como monitores a color o impresoras) o a alguna aplicación donde el objetivo es la manipulación del color (como la creación de gráficas a color para animaciones). En términos de procesamiento digital de imágenes, los modelos orientados al hardware más comúnmente utilizados en la práctica son el modelo RGB (rojo, verde, azul) modelo para monitores a color y para una gran variedad de cámaras de video en color; el CMY (cyan, magenta, amarillo) y el CMYK (cyan, magenta, amarillo, negro), modelos para impresoras en color; y el HSI (brillo (hue), saturación, intensidad) que corresponde al modelo más cercano a la manera en que los humanos percibimos el color. A continuación se describen los espacios de color RGB y HSI.

Una imagen de color RGB es una matriz de píxeles de tamaño $M \times N \times 3$, donde cada pixel es una combinación de colores rojo, verde y azul. Por convención, cada matriz se

referencia como componente rojo, verde y azul de la imagen RGB. Una imagen RGB es de 8 bits (uint8) si su rango de valores es de $[0,255]$, o de 16 bits (uint16) para $[0,65535]$, o double para $[0,1]$. El número de bits empleados para representar el pixel de las componentes de la imagen determinan la profundidad de bit (*bit depth*). Por ejemplo, si cada una de las componentes de la imagen es de 8 bits, entonces la imagen RGB correspondiente se dice que es de 24 bits de fondo. El número total de colores en una imagen RGB de 24 bits es de $(2^8)^3 = 16,777,216$. El espacio de color RGB, generalmente se muestra como un cubo de colores como se muestra en la Figura 1.18. Los vértices del cubo son los colores primarios (rojo, verde y azul) y secundarios (cyan, magenta y amarillo) de la luz.

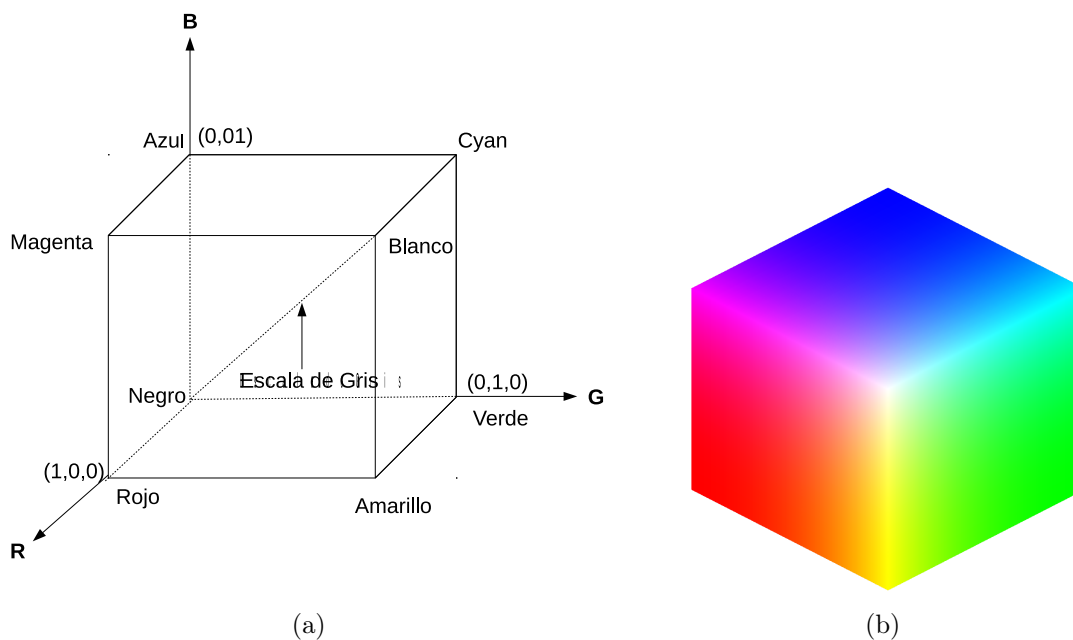


Figura 1.18: Cubo 24 bits RGB. (a) Esquema normalizado a valores entre $[0,1]$, (b) Cubo en colores.

El término nivel de gris (gray level) se emplea para referirse a la intensidad de las imágenes monocromáticas.

El modelo RGB está relacionado con el hardware, sin embargo no es útil para describir los colores en términos prácticos para la interpretación humana. Por ejemplo, uno no se refiere al color de un auto dando porcentajes del contenido de cada uno de los colores primarios. Cuando los humanos vemos un color, lo describimos en términos de su tono (H), saturación (S) y su brillo o intensidad (I). El modelo HSI debe su utilidad a dos hechos:

1. La componente de intensidad (I) esta desacoplada de la información cromática.
2. Las componentes de tono (H) y saturación (S) están íntimamente relacionadas con la forma en que los humanos percibimos el color.

El modelo de color HSI es una herramienta ideal para el procesamiento de imágenes en color. Los componentes importantes del espacio de color HSI, son el eje vertical de intensidad, el tamaño del vector al punto de color y el ángulo de ese vector hace respecto a la referencia (eje rojo). Por lo tanto no es inusual el plano HSI está definido en términos de los cortes en círculos vistos en Figura 1.19.

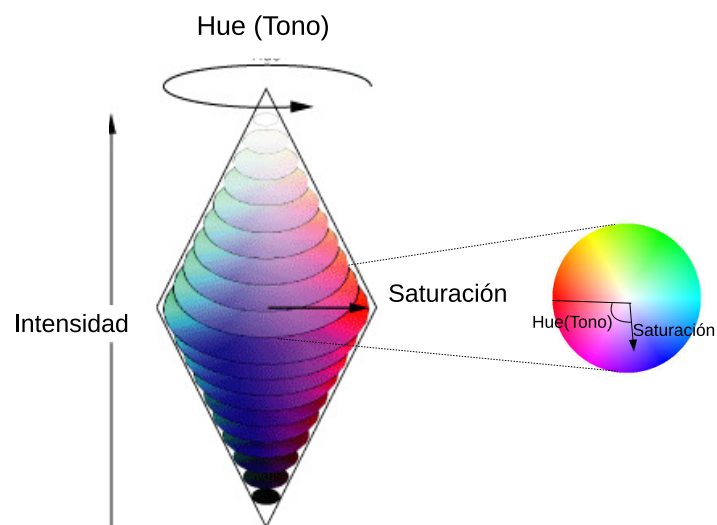


Figura 1.19: Ejemplo de una representación del espacio de color HSI.

Anexo 2

Cómputo paralelo

2.1. Introducción

Durante varios años el modo habitual del desarrollo computacional se ha hecho de forma secuencial. De esta manera, una computadora con un único procesador sólo cuenta con la capacidad de ejecutar una instrucción en un momento determinado. Esto es debido a la adopción del modelo que propuso Von Neuman, donde la estabilidad es la base para el desarrollo de diversas aplicaciones. Sin embargo, desde el comienzo de la computación secuencial se vislumbraba que, tarde o temprano, se necesitaría recurrir a la computación paralela para resolver problemas que demandaban muchos recursos de cómputo por lo que su implementación en máquinas secuenciales no sería lo más óptimo o en su caso posible. Por ejemplo:

- Sistema de tamaño N
- Atracción de $N - 1$ cuerpos $\rightarrow N * (N - 1) \approx N^2$
- Predicción del clima.

A la ejecución de dos o más procesos al mismo tiempo usando más de un procesador se le llama cómputo paralelo. Procesar un programa en paralelo hace que se ejecute de manera más rápida ya que existen más procesadores trabajando en forma simultánea y coordinada en diferentes partes del problema. El objetivo es reducir al mínimo el tiempo total de cómputo, esto se hace mediante la distribución de la carga de trabajo entre los procesadores disponibles en ese momento.

Las aplicaciones paralelas se pueden encontrar en una amplia gama de disciplinas, que van desde la predicción del clima, el modelado de la biósfera, el procesamiento de imágenes, el modelado de océanos, etc. La velocidad de procesamiento no es la única razón para utilizar el paralelismo. La construcción de aplicaciones más complejas ha requerido computadoras más rápidas, y las limitaciones en el desarrollo de computadoras seriales han llegado a ser más y más evidentes. Anteriormente, el procesamiento paralelo se empleaba para resolver problemas que requerían una gran escala simulación, un gran número de cálculos y procesamiento de datos (por ejemplo, el cómputo de los datos de un censo). En la actualidad, el costo del hardware disminuye rápidamente, por lo que el procesamiento paralelo se está utilizando cada vez más en tareas comunes. No obstante, además de las facilidades de hardware que permiten llevar a cabo este tipo de procesamiento, es necesario contar con software que soporte la ejecución y la coordinación de procesos en forma paralela.

En 1966 Michael Flynn propuso un mecanismo de clasificación de las computadoras paralelas. Aunque Flynn no cubre todas las posibles arquitecturas, su taxonomía facilita de manera significativa la comprensión del funcionamiento, basándose en el número de instrucciones y en la secuencia de datos simultáneos que utiliza una computadora para procesar información. Los tipos más sobresalientes dentro de esta clasificación son SISD (Single Instruction Single Data), SIMD (Single Instruction Multiple Data), y MIMD (Multiple Instruction Multiple Data).

La Figura 2.2 muestra una clasificación general de las arquitecturas paralelas, las cuales se pueden dividir en dos grandes grupos, los multiprocesadores y las multicomputadoras.

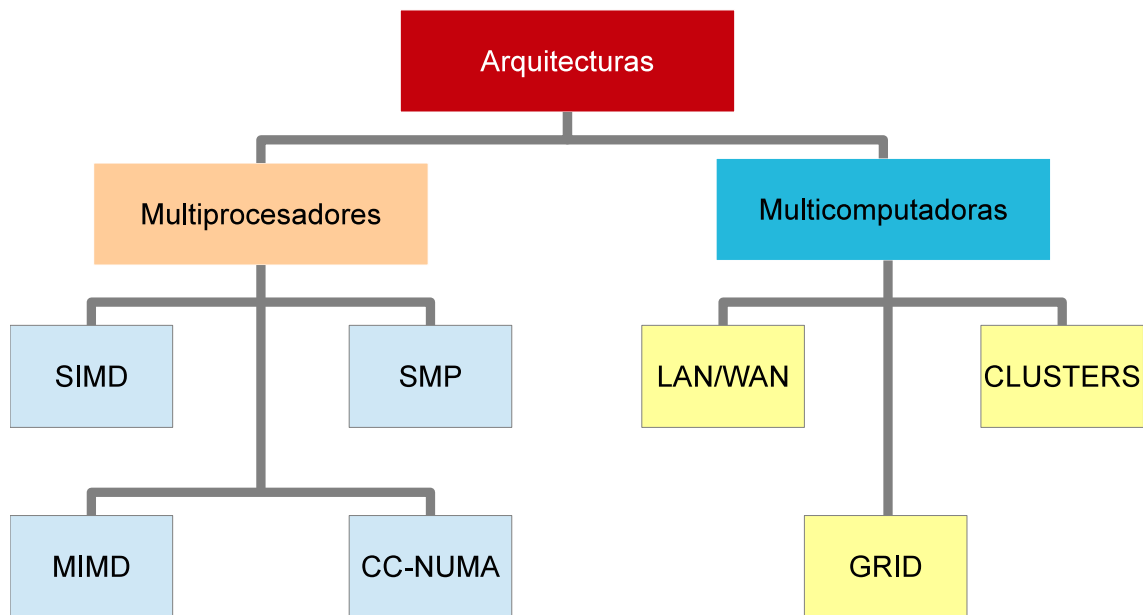


Figura 2.1: Arquitecturas paralelas.

2.1.1. SISD(Single instruction-Single data)

Este es el modelo tradicional de computación secuencial donde una unidad de procesamiento recibe una sola secuencia de instrucciones que opera en una secuencia de datos Figura2.2.

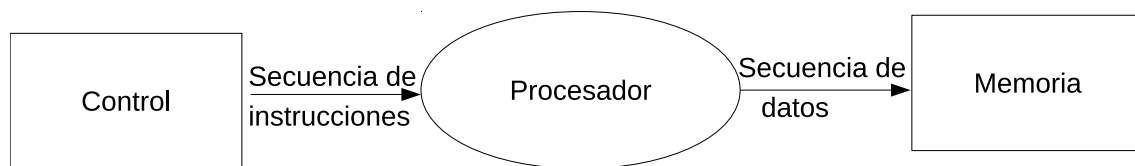


Figura 2.2: Arquitectura SISD.

2.1.2. Multiprocesadores

Estas arquitecturas consisten en una computadora que se compone de un arreglo de procesadores y un banco de memoria común, conectados a través de un bus rápido o uno de los varios tipos de red de interconexión dinámica multietapas. Existen dos tipos de sistemas multiprocesador:

- **Simétrico.** En estos los microprocesadores del sistema comparten los recursos del sistema, como la entrada/salida de memoria y disco. Se distribuye uniformemente la carga de trabajo a los procesadores disponibles.
- **Asimétrico.** En estos, diferentes procesadores gobiernan las distintas tareas y recursos del sistema, El enfoque principal que adoptan, es incrementar el rendimiento del sistema para múltiples usuarios de red, en múltiples segmentos de red de área local. Cada CPU se dedica a una función específica entrada/salida de la tarjeta de interfaz de red u operaciones con manejo de archivos.

La mejor manera de programar dentro de los multiprocesadores es a través de los hilos ya que estos se comunican por medio de la memoria compartida ya que estando dentro de un mismo proceso de ejecución comparten el mismo espacio de memoria, en este tipo de programación hay que tener cuidado a la hora de manejar las variables globales ya que se pueden generar errores no deseados a la hora de leer/escribir en ellas.

2.1.3. Multicomputadoras

Estas arquitecturas se componen de una arquitectura de memoria distribuida, es decir, no compartida. En esta arquitectura se tiene un arreglo de microprocesadores, cada uno con su propia memoria local. Son sistemas donde cada unidad de proceso es autónoma e independiente de las demás pero trabajan de manera conjunta. La interconexión de unidades de proceso se hace a través de redes de computadora de preferencia de altas velocidades.

La mejor manera de programar a través de las multicomputadoras es por medio del paso de mensajes, ya que cada proceso se ejecuta en una máquina diferente, si estos deben de enviar los resultados de algún calculo a otro proceso que lo necesite, tiene que ser a través de un mensaje. La directiva más conocida para paso de mensajes es la de MPI, la cual, ha hecho posible el desarrollo de un grado aceptable de portabilidad del software paralelo. MPI proporciona una interfaz de programación basada en envío y recepción de mensajes que es independiente de cualquier arquitectura paralela particular, además de que puede enlazarse con C++ o Fortran.