



UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE CIENCIAS
Maestría y Doctorado en Ciencias e Ingeniería

Sistema de Reconocimiento de Patrones CW-SIFT

Tesis que para obtener el grado de Doctor en Ciencias presenta

Carolina Barajas García

Aprobada por el siguiente comité

Dra. Selene Solorza Calderón
Director del Comité

Dr. Josué Álvarez Borrego
Miembro del Comité

Dr. Everardo Gutiérrez López
Miembro del Comité

Dr. Jesús Lerma Aragón
Miembro del Comité

Dr. Antonio González Fernández
Miembro del Comité

Ensenada, Baja California, 9 de agosto de 2019

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE CIENCIAS



MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA
ÁREA DE COMPUTACIÓN

SISTEMA DE RECONOCIMIENTO DE PATRONES CW-SIFT

TESIS

que para cubrir los requisitos necesarios para obtener el grado de

DOCTOR EN CIENCIAS

Presenta:

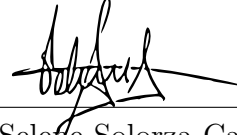
CAROLINA BARAJAS GARCÍA

Ensenada, Baja California, México, agosto de 2019

RESUMEN de la tesis de **CAROLINA BARAJAS GARCÍA**, presentada como requisito para la obtención del grado de DOCTOR EN CIENCIAS. Ensenada, Baja California, agosto de 2019.

SISTEMA DE RECONOCIMIENTO DE PATRONES CW-SIFT

Resumen aprobado por:



Dra. Selene Solorza Calderón

Director de Tesis

En este trabajo se presenta una nueva metodología para generar un descriptor de puntos clave en imágenes digitales invariante a traslación, rotación y escala, llamado descriptor W-SIFT (Wavelet - Scale Invariant Feature Transform). La metodología para construir el descriptor W-SIFT ubica puntos clave en tres diferentes pirámides de escala para extraer características representativas de la imagen. Estas pirámides se construyen utilizando la transformada wavelet de Haar en dos dimensiones. Por lo tanto, la invarianza a traslación se logra por el uso de puntos clave y la invarianza a escala se obtiene a través de las pirámides de escala. La invarianza a rotación se resuelve sumando los valores de intensidad de la imagen dada filtrada con una máscara binaria con una circunferencia de radio r centrada en el punto clave, el resultado se almacena en el descriptor. Para cada punto clave, se usaron 40 máscaras binarias ($r = 1, 2, \dots, 40$ píxeles); por lo tanto, el descriptor W-SIFT es un vector de 40 entradas. La robustez de W-SIFT se probó utilizando imágenes con diferentes escalas y rotaciones y comparando los resultados con las metodologías SIFT (Scale Invariant Feature Transform) y SURF (Speeded-Up Robust Features). W-SIFT muestra un rendimiento similar a SIFT y mucho mayor que SURF. Otra ventaja que W-SIFT muestra es ser más rápido que SIFT, construyendo los descriptores hasta cuatro veces más rápido. Por lo tanto, W-SIFT es una metodología rápida, eficiente y fácil de implementar.

Palabras Clave: W-SIFT, wavelets, características locales, puntos clave, SIFT, SURF.

ABSTRACT of the thesis presented by **CAROLINA BARAJAS GARCÍA**, in fulfillment of the requirements of the degree of DOCTORATE IN SCIENCES. Ensenada, Baja California, august 2019.

CW-SIFT PATTERN RECOGNITION SYSTEM

Approved by:



Dra. Selene Solorza Calderón

Thesis Advisor

In this paper is presented a novel scale, translation and rotation invariant keypoint descriptor of digital images, named W-SIFT (Wavelet - Scale Invariant Feature Transform). The methodology to construct the W-SIFT descriptor locates in three different scale pyramids keypoints to extract representative features of the image. These scale pyramids are built using the 2D multi-resolution representation by Haar wavelet. Hence, the translation invariance was achieved using keypoints, and the scale invariance was obtained via the scale pyramids. The rotation invariance is worked out by adding the intensity values of the given image filtered by a binary mask with a circumference of radius r centered on the keypoint; the result was stored in the descriptor. For each keypoint, 40 binary masks were used ($r = 1, 2, \dots, 40$ pixels); thus the STR-invariant 1D-W-SIFT descriptor is a vector of 40 entries. The robustness of the STR-invariant 1D-W-SIFT descriptor was tested using images with different scales and rotations and comparing the results with the SIFT (Scale Invariant Feature Transform) and SURF (Speeded-Up Robust Features) descriptors. The W-SIFT methodology shows results similar to SIFT and much higher than SURF. Another advantage that W-SIFT shows is being faster than SIFT, building the descriptors up to four times faster. Therefore, WLFD is a fast, efficient and easy-to-implement methodology.

Keywords: W-SIFT, wavelets, local features, keypoints, SIFT, SURF.

A mi pequeña y feliz familia,

Alicia y Erik

Agradecimientos

En primer lugar quiero agradecer a la Dra. Selene Solorza Calderón por su apoyo, tiempo, dedicación y paciencia invertida en este trabajo y en mí. Gracias por todo lo que me has enseñado y lo bien que me has preparado a lo largo de estos años.

A mis sinodales Dr. Josué Álvarez Borrego, Dr. Everardo Gutiérrez López, Dr. Jesús Lerma Aragón y Dr. Antonio González Fernández, por sus correcciones, sugerencias y aportaciones para mejorar este trabajo, así como a la Dra. Eloísa del Carmen García Canceco, coordinadora del posgrado en la Facultad de Ciencias.

Al CONACYT, por la beca otorgada para realizar mis estudios de doctorado, así como a PRODEP y UABC por el apoyo brindado mediante el proyecto interno con número 400/1755.

A mi esposo Erik y mi hija Alicia, por siempre recibir su apoyo incondicional, por hacerme ser una mejor persona y superarme cada día, los amo.

A mis padres Rafael y Alma, a mi hermana Alma y a mis sobrinas Julieta y Anya, gracias a ustedes soy lo que soy, somos una gran familia.

A mis amigos, Tarek y Andrei, por nunca decirme que no cuando los necesité, por escucharme y estar ahí para mí.

Contenido

	Página
Resumen en español	i
Resumen en inglés	ii
Dedicatoria	iii
Agradecimientos	iv
Contenido	v
Lista de Figuras	vii
Lista de Tablas	x
I. Introducción	1
I.1 Antecedentes	1
I.2 Justificación	5
I.3 Planteamiento del problema	7
I.4 Preguntas de investigación	7
I.5 Objetivos	8
I.5.1 Objetivo general	8
I.5.2 Objetivos específicos	8
I.6 Metodología	9
II. Transformada Wavelet	11
II.1 Fundamentos matemáticos	12
II.2 Propiedades	13
II.2.1 Linealidad	13
II.2.2 Simetría	14
II.2.3 Traslación	14
II.2.4 Dilatación	15
II.2.5 Antilinealidad	15
II.3 Ejemplos de wavelets	16
II.3.1 Wavelet Sombrero Mexicano	16
II.3.2 Wavelet de Morlet	17
II.3.3 Wavelet de Meyer	18
II.3.4 Wavelet de Haar	19
II.4 Transformada wavelet en dos dimensiones	19
III. Descriptores de rasgos locales	21
III.1 SIFT	22

Contenido (continuación)

	Página
III.1.1 Generación del espacio de diferencias de gaussianas	23
III.1.2 Localización de los puntos clave	25
III.1.3 Asignación de una orientación a cada punto clave	27
III.1.4 Asignación del descriptor a cada punto clave	29
III.2 SURF	30
III.2.1 Imagen integral	31
III.2.2 Generación del espacio escala	31
III.2.3 Localización de los puntos clave	33
III.2.4 Asignación de una orientación a cada punto clave	33
III.2.5 Asignación del descriptor a cada punto clave	35
IV. W-SIFT	37
IV.1 Metodología	37
IV.1.1 Generación de la pirámide wavelet	37
IV.1.2 Localización de los puntos clave	40
IV.1.3 Asignación del descriptor a cada punto clave	42
IV.2 Emparejamiento de descriptores	44
V. Base de datos de imágenes digitales	46
VI. Pruebas y análisis de resultados	48
VI.1 Análisis de tiempo de cómputo	48
VI.2 Análisis de rendimiento	49
VI.2.1 Pruebas de rotación y escala	50
VI.2.2 Detección de puntos clave en imágenes con textura sin bordes definidos	55
VI.3 Detección de puntos clave en imágenes con iluminación no homogénea	60
VI.4 Detección de puntos clave en imágenes con ruido moteado	62
VI.5 Rendimiento del método de emparejamiento	64
VII. Clasificación de imágenes	67
VIII. CW-SIFT	70
IX. Conclusiones	75
X. Trabajo a futuro	78
REFERENCIAS	79

Lista de Figuras

Figura		Página
1	Gráfica de la wavelet Sombrero Mexicano.	16
2	Gráfica de la wavelet de Morlet.	17
3	Gráfica de la wavelet de Meyer.	18
4	Gráfica de la wavelet de Haar.	19
5	(a) Imagen de referencia. (b) Boceto de la transformada wavelet aplicada en dos dimensiones.	20
6	Sub-imágenes resultantes de la transformada wavelet en dos dimensiones utilizando la wavelet de Haar. (a) Imagen de referencia. (b) Frecuencias AA. (c) Frecuencias BA. (d) Frecuencias AB.	20
7	Ejemplo del espacio escala.	23
8	Ejemplo del espacio de diferencias de gaussianas.	25
9	Bosquejo de la selección de candidatos a puntos clave.	26
10	Ejemplo de un histograma de gradientes orientados de 36 entradas. . .	28
11	Esquema del proceso para generar el descriptor de un punto clave. . .	29
12	Diagrama de bloques para generar el descriptor SIFT.	30
13	(a) Imagen de referencia. (b) Imagen integral de Fig. 13a.	31
14	(a) Ejemplo de un filtro gaussiano utilizado por SIFT. (b) Ejemplo de un filtro de caja utilizado por SURF.	32
15	(a) Filtro wavelet de Haar en dirección x . (b) Filtro wavelet de Haar en dirección y . El color negro representa el valor escalar -1 y el blanco 1	33
16	Boceto donde se aplica el filtro wavelet de Haar dentro de la ventana circular (color azul) alrededor del punto clave (píxel rojo). (a) Filtro wavelet de Haar en dirección x . (b) Filtro wavelet de Haar en dirección y	34
17	Boceto de la asignación de orientación a un punto clave.	34
18	Ejemplo del procedimiento para generar el descriptor SURF.	36
19	Diagrama de bloques de para generar el descriptor SURF.	36
20	Imagen de referencia escalada de -50% a $+50\%$ con un tamaño de paso de 1%	38

Lista de Figuras (continuación)

Figura	Página
21 Pirámides wavelet. (a) Pirámide P_{AA} . (b) Pirámide P_{BA} . (c) Pirámide P_{AB}	39
22 (a) P_{BA}^1 . (b) Máscara binaria de borde M^1 . (c) Imagen filtrada F_{BA}^1	40
23 Boceto para analizar si un píxel es máximo o mínimo.	41
24 (a) Puntos clave detectados con P_{AA} . (b) Puntos clave detectados con P_{BA} . (c) Puntos clave detectados con P_{AB}	42
25 Ejemplo de un filtro de circunferencia de radio r centrado en un punto clave (punto rojo).	42
26 Ejemplo de un descriptor W-SIFT.	43
27 Diagrama de bloques para generar el descriptor W-SIFT.	44
28 Ejemplo de los emparejamientos entre dos imágenes.	45
29 Base de datos de imágenes de referencia.	47
30 Análisis del tiempo de cómputo para W-SIFT, SIFT y SURF.	49
31 Pruebas de rendimiento para imágenes rotadas. (a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.	53
32 Pruebas de rendimiento para imágenes escaladas. (a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.	54
33 Ejemplos de lesiones de piel (imágenes tomadas de la librería de imágenes Dermofit [1]. (a), (b) y (c): Queratosis actínica. (d) y (e): Carcinoma de células basales.	55
34 Máscaras binarias que cubren solo la lesión de las imágenes de la Fig. 33.	55
35 Puntos clave emparejados para la imagen I_A , Fig. 33a. (a) Zona de lesión de piel. (b) Imagen de la lesión de piel. Arriba: W-SIFT. En medio: SIFT. Abajo: SURF.	57
36 Pruebas de rendimiento para imágenes con textura sin bordes definidos. (a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.	58
37 Ejemplo de imágenes con iluminación no homogénea. (a) I_2 . (b) I_3	60
38 Pruebas de rendimiento para imágenes con iluminación no homogénea. (a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.	61

Lista de Figuras (continuación)

Figura		Página
39	Ejemplo de imágenes con ruido moteado para I_{12} . (a) Varianza = 0.01. (b) Varianza = 0.05. (c) Varianza = 0.1. (d) Varianza = 0.15. (e) Varianza = 0.2.	62
40	Pruebas de rendimiento para imágenes con ruido moteado. (a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.	63
41	(a) Imagen de referencia. (b) Imagen de prueba, versión rotada 30° de la IR. (c) Regresión de la IP a la IR.	64
42	Curva ROC para la clasificación de imágenes. FPR (False Positive Rate), TPR (True Positive Rate), se muestra en asterisco rojo el punto de inflexión de la curva.	68
43	Pruebas de rendimiento para la clasificación de imágenes.	68
44	Pruebas de rendimiento para la clasificación de imágenes con un umbral de 302 emparejamientos.	69
45	Ejemplos de lesiones de piel a color (imágenes tomadas de la librería de imágenes Dermofit [1]). (a), (b) y (c): Queratosis actínica. (d) y (e): Carcinoma de células basales.	70
46	Descomposición RGB para I_B (Fig. 45b). (a) Canal rojo I_B^r . (b) Canal verde I_B^g . (c) Canal azul I_B^b	71
47	Detección de puntos clave. (a) Canal rojo I_B^r . (b) Canal verde I_B^g . (c) Canal azul I_B^b	71
48	Emparejamiento de imágenes. (a) Canal rojo. (b) Canal verde. (c) Canal azul. (d) Combinación de los tres canales en la imagen a color. .	72

Lista de Tablas

Tabla		Página
I	Resultados de rotación para I_2 , utilizando W-SIFT.	51
II	Resultados de rendimiento para imágenes con textura sin bordes definidos.	59
III	Distancia Euclidiana entre IR y sus correspondientes IP escaladas. . . .	65
IV	Distancia Euclidiana entre IR y sus correspondientes IP rotadas.	66
V	Resultados imágenes a color.	74

Capítulo I

Introducción

I.1 Antecedentes

Desde mediados del siglo pasado, la extracción de características a una imagen ha sido un área activa en el campo del reconocimiento de patrones. Una característica es una pieza de información que es relevante para la resolución, mediante imágenes digitales, de alguna tarea computacional específica. Las características pueden ser estructuras en la imagen como puntos, bordes u objetos, y las cuales se engloban en un descriptor, el cual especifica propiedades elementales del objeto, tales como la forma, el color, la textura, entre otros. Un ejemplo de estos son los descriptores de rasgos locales SIFT (por sus siglas en inglés: Scale Invariant Feature Transform), introducidos por Lowe en 2004 [2, 3].

SIFT es ampliamente utilizado, debido a la idea innovadora de utilizar puntos clave para representar características relevantes de una imagen. El uso de puntos clave permite trabajar con imágenes que contienen varios objetos, escenarios e incluso fragmentos de objetos, a diferencia de otras metodologías como *Ánalysis de Componentes Principales* (PCA, por sus siglas en inglés) [4] o los sistemas de reconocimiento de patrones basado de máscaras binarias de firmas unidimensionales [5, 6, 7]. Además, SIFT se utiliza en diferentes áreas, como seguimiento de objetos, identificación de rostros, detección de falsificaciones y el reconocimiento de iris, entre otros [8, 9, 10, 11]. Sin embargo, SIFT tiene varias desventajas, el tiempo de cómputo es considerable y tiene un bajo rendimiento con imágenes que presentan ruido o iluminación no homogénea. Además,

cuando la imagen tiene textura sin bordes definidos, SIFT detecta muy pocos puntos clave. Por otra parte, SIFT es una metodología que no es fácil de implementar, consta de pasos complicados como el cálculo de matrices Hessianas, polinomios de Taylor en 2D de segundo orden y la localización de valores extremos en funciones 2D; estas operaciones implican el uso de matrices de 2×2 compuestas por operadores diferenciales parciales de primer y segundo orden.

Existen en la literatura variantes de SIFT que intentan mejorar su rendimiento y deficiencias o resolver un problema específico modificándolo de alguna manera. Por ejemplo, ORB (por sus siglas en inglés: Oriented and Rotated Brief) que es resistente al ruido [12], *bi*-SIFT el cual genera descriptores de características más abstractas [13, 14], PCA-SIFT que en lugar de usar histogramas ponderados suavizados lo reemplaza por PCA [15]. La variante más utilizada de SIFT es SURF (por sus siglas en inglés: Speeded-Up Robust Features), presentada por Herbert Bay en 2006 [16]. SURF es una versión rápida de SIFT, disminuye el tiempo de cómputo pero a su vez reduce considerablemente su eficacia.

Otro tipo de variantes de SIFT son aquellas que utilizan la transformada wavelet. Por ejemplo, Aravindan y Anzar [17] aplican la transformada wavelet biortogonal para el reconocimiento parcial de huellas dactilares. Esta metodología trabaja con dos sub-ímagenes de la transformada wavelet, la sub-imagen de las bajas-bajas (BB) frecuencias para preservar la información global de la imagen y la sub-imagen de las altas-altas (AA) frecuencias para resaltar la información de los bordes. Después, aplican SIFT en BB y AA para obtener los descriptores. En 2009, Lim y colaboradores propusieron WSIFT [18], este trabajo muestra un análisis comparativo del desempeño de SIFT empleando las transformadas wavelet Haar, Daubechies y Gabor de la imagen dada, en

lugar de usar la imagen en coordenadas espaciales. El sistema WSIFT no es una mejora de SIFT, ya que se utiliza sin alguna modificación. Otra metodología que maneja una transformada wavelet y SIFT es WLF [19], en este trabajo la transformada wavelet de Haar se aplica a la imagen dada para generar tres pirámides usando las sub-imágenes de las frecuencias altas-altas (AA), bajas-altas (BA) y altas-bajas (AB) y cinco niveles para cada sub-imagen. Después, se aplican los filtros tipo Haar para crear los descriptores [20]. La metodología WLF es invariante a traslación y escala, con un rango de escala de $\pm 15\%$, pero no es invariante a rotación.

Una característica que las metodologías anteriormente mencionados comparten, es que solo trabajan con imágenes en escala de grises. Sin embargo, el color es una característica esencial para la clasificación de objetos, si esta información se desestima se perderá una propiedad muy importante. Por ejemplo, en Boese y colaboradores [21] el color se utiliza para estudiar lesiones de hojas de la *Zostera marina*, pero el procesamiento de la imagen se realiza a mano a través de múltiples programas: Adobe PhotoShop, Canon, Photostitch y ERMapper. En la industria de la pesca el color es determinante para clasificar y contar la producción cosechada de organismos acuáticos y así tener una apropiada explotación de los recursos marinos. Además, dicho atributo es útil para detectar alguna enfermedad en los organismos acuáticos, por lo que la automatización de dichos procesos es de importancia tanto industrial como de preservación de recursos naturales. Por lo tanto, surge la necesidad de desarrollar técnicas más robustas que aporten a la resolución de problemas de clasificación de imágenes a color.

Se han publicado varias técnicas variantes de SIFT para trabajar con imágenes a color. En 2006, Abdel-Hakim y Farag [22] propusieron el algoritmo CSIFT (Color SIFT), el

cual utiliza el modelo de invariancia a color desarrollado por Geusebroek [23], pero como solo toma en cuenta variaciones fotométricas y no transformaciones geométricas puede llegar a no ser eficiente en casos prácticos. Otra metodología propuesta es SIFT-CCH (SIFT-Color Co-occurrence Histograms) de Ancuti y Bekaert [24], donde se combina la técnica SIFT con el histograma co-ocurrente a color (CCH) [25]. El propósito de SIFT-CCH es analizar de una manera sencilla la información del espacio de color. Sin embargo, esta técnica no trabaja con imágenes que presentan ruido, iluminación no homogénea o variaciones azimutales, además la mejora del rendimiento del sistema con respecto a SIFT no es considerable. En 2012, Alitappeh y colaboradores [26] crearon un descriptor para imágenes a color basado en SIFT y que utiliza variaciones en la iluminación. Este algoritmo usa una técnica de agrupación de puntos de intensidades similares con la finalidad de eliminar los que no sean necesarios. Sin embargo, debido a que el espacio de color utilizado tiene muchos canales, el número de puntos extraídos es considerable, lo cual genera un excesivo número de cálculos. Aunque se han propuesto varias técnicas variantes de SIFT para trabajar con imágenes a color, éstas no han sido tan populares como en el caso de las imágenes en escala de grises debido a que la complejidad y los cálculos a realizar aumentan considerablemente, generando un tiempo de cómputo alto.

En este trabajo, desarrollamos un nuevo procedimiento para construir tres pirámides basadas en la transformada wavelet de Haar, con el objetivo de obtener la invarianza a rotación y un rango de invariancia a escala de $\pm 50\%$. Se propone una nueva metodología de características locales, la cual es hasta cuatro veces más rápida que SIFT preservando la precisión (a diferencia de SURF, que es rápida pero con baja precisión). La detección de puntos clave y la construcción de los descriptores invariantes a traslación, rotación y escala, se basan en la extracción de características locales y en la transfor-

mada wavelet de Haar, ya que esta wavelet tiene varias propiedades, como el análisis multiresolución en 2D, manejo de variables de traslación y escala, así como un excelente rendimiento para detectar bordes. Por último, se generalizó la metodología a imágenes a color, descomponiendo la imagen en sus tres canales de color (rojo, verde y azul) y aplicando la metodología en cada uno de ellos, obteniendo una metodología rápida, robusta y aplicable a imágenes a color.

I.2 Justificación

El ser humano requiere poco esfuerzo para reconocer múltiples objetos en imágenes digitales, a pesar de que dichos objetos presenten distorsiones geométricas como cambio de tamaño, rotación o cambio de posición. Sin embargo esta tarea es un gran desafío para un sistema digital de reconocimiento de patrones. Uno de los objetivos del reconocimiento de patrones es obtener una metodología que sea invariante a la mayoría de las transformaciones geométricas, que trabaje con imágenes a color y que el tiempo sea lo más cercano al tiempo real. Estos sistemas pueden beneficiar a personas con discapacidad visual ofreciendo una mejor calidad de vida; así como a construir robots eficientes que ayuden en situaciones de emergencia y no poner en riesgo la vida de una persona.

En la literatura existen varios sistemas de reconocimiento de patrones, la mayoría basados en las técnicas de rasgos locales (SIFT), los cuales presentan buen rendimiento al clasificar imágenes con diferentes transformaciones geométricas como rotación angular, escala, traslación y pequeñas variaciones en el ángulo azimutal. Estos sistemas tienen varias desventajas, el tiempo de cómputo es considerable y tienen bajo desempeño

ante imágenes con poca cantidad de ruido o iluminación no homogénea. Además, su generalización para la clasificación de imágenes a color es complicada y emplea un tiempo de cómputo considerable, y cuando las variaciones azimutales aumentan, los sistemas pierden su eficiencia.

Se han publicado variantes de SIFT basadas en transformadas wavelets. Algunas de ellas mejoran el problema del costo computacional de SIFT, pero no lo suficiente para emplearlos en problemas de tiempo real, además ninguna de ellas extiende su aplicabilidad a imágenes a color. Una posible solución a los problemas presentados por SIFT es usar las ventajas que han mostrado las transformadas wavelets en el reconocimiento de patrones: la detección de bordes, que podría ser de gran ayuda al momento de identificar los puntos clave que utiliza SIFT y obtener una mayor certeza al momento de hacer la clasificación de las imágenes; la eliminación de ruido y de datos irrelevantes; el análisis de multiresolución, con la cual se podría generar la pirámide de escalas y obtener una mayor eficiencia que el utilizar la pirámide Gaussiana como lo hace SIFT.

Por otro lado, una parte muy importante en toda investigación, es la apropiada validación de los resultados, ya que ésta nos indicará el nivel de confianza ante la toma de decisiones. El sistema SIFT se valida utilizando el índice de repetibilidad, el cual es una medida que cuantifica la concordancia de las mediciones de una misma característica [3]. Por lo regular, en los sistemas variantes a SIFT se realiza la validación comparando los resultados directamente con aquellos que se obtendrían mediante SIFT, sin llevar a cabo un análisis más profundo del nivel de confianza obtenido por dichos sistemas. En este trabajo se propone, además de comparar los resultados con diferentes sistemas de rasgos locales, utilizar alguna técnica de validación para medir el nivel de confianza del sistema de reconocimiento de patrones propuesto.

I.3 Planteamiento del problema

Las metodologías existentes de reconocimiento de patrones basados en la teoría de rasgos locales SIFT presentan varias desventajas, el tiempo de cómputo es considerable, su generalización a imágenes a color es complicada, cuando se presentan variaciones azimutales los sistemas pierden su eficiencia, además no existe un sistema que sea invariante a todas las transformaciones geométricas al mismo tiempo (escala, traslación, rotación angular y azimutal) y que a su vez clasifique imágenes a color. Estos requerimientos son necesarios para identificar un objeto en una imagen compuesta, es decir, una imagen que contiene multiples objetos en alguna escena. En este trabajo se propone diseñar una metodología de reconocimiento de patrones para la clasificación de imágenes a color a partir de las técnicas SIFT y las transformadas wavelets que sea eficiente para clasificar imágenes que presenten objetos con las transformaciones geométricas mencionadas anteriormente. Además, con base en la estadística paramétrica, se establecerá una técnica para medir el nivel de confianza del sistema.

I.4 Preguntas de investigación

1. ¿Cuál transformada wavelet es la óptima para la selección de los puntos clave de las imágenes digitales que se utilizarán en la metodología SIFT?
2. ¿Cuál transformada wavelet es la más apropiada para llevar a cabo la pirámide de escalas que se utilizará en la metodología SIFT?
3. ¿Cuál es la eficiencia del sistema propuesto W-SIFT para imágenes en escala de grises?

4. ¿De qué manera se puede generalizar la metodología propuesta para que clasifique imágenes a color?
5. ¿Cuál es el nivel de confianza del sistema de reconocimiento de patrones propuesto CW-SIFT?

I.5 Objetivos

I.5.1 Objetivo general

Diseñar una metodología de reconocimiento de patrones para la clasificación de imágenes digitales a color, mediante una combinación de las técnicas de rasgos locales SIFT y las transformadas wavelets. Esta metodología será eficiente para clasificar imágenes que presenten objetos con las transformaciones geométricas: traslación, rotación angular, escala y variaciones en el ángulo azimutal. Además, con base en la estadística paramétrica, se establecerá una técnica para medir el nivel de confianza del sistema.

I.5.2 Objetivos específicos

1. Determinar cuál transformada wavelet ofrece mejores resultados para la selección de los puntos clave en imágenes digitales.
2. Determinar cuál transformada wavelet es la más apropiada para generar el espacio escala.
3. Diseñar e implementar en lenguaje de programación MatLab el sistema de reconocimiento de patrones W-SIFT para que clasifique imágenes digitales en escala de grises.

4. Evaluar la eficiencia de la metodología propuesta W-SIFT y optimizar el tiempo de cómputo.
5. Generalizar el sistema de reconocimiento de patrones propuesto para que clasifique imágenes a color mediante la implementación de W-SIFT en el espacio de color RGB.
6. Evaluar el nivel de confianza del sistema digital CW-SIFT para validar su funcionamiento.

I.6 Metodología

1. Revisión bibliográfica.
2. Estudiar las diferentes técnicas de extracción de características en imágenes digitales mediante las transformadas wavelets.
3. Estudiar los filtros utilizados en las técnicas de rasgos locales: Laplaciano, Gaussiano y Gabor.
4. Implementar la metodología clásica SIFT propuesta por G. Lowe.
5. Analizar la manera en que SIFT genera la pirámide de escalas y cómo sustituirla por una generada con una transformada wavelet.
6. Analizar la forma en la que en SIFT se seleccionan los puntos clave y cómo estos, a partir de una transformada wavelet, se pueden reducir en número y a la vez incrementar la eficacia de los mismos.

7. Diseñar e implementar en lenguaje de programación MatLab un sistema de reconocimiento de patrones W-SIFT, para que clasifique imágenes digitales en escala de grises.
8. Calcular el índice de repetibilidad del sistema W-SIFT.
9. Optimizar el tiempo de cómputo del sistema W-SIFT.
10. Diseñar e implementar en lenguaje de programación MatLab el sistema de reconocimiento de patrones para que clasifique imágenes digitales a color en el espacio RGB.
11. Estudiar técnicas de validación de la estadística paramétrica y determinar cuál es la más apropiada para medir el nivel de confianza del sistema CW-SIFT.

Capítulo II

Transformada Wavelet

La transformada wavelet (o transformada de ondícula) es un tipo especial de transformada matemática que representa a una señal en términos de versiones trasladadas y escaladas de una onda finita (denominada wavelet madre). Su principal ventaja es que permite llevar a cabo un análisis a diferentes escalas o resoluciones (análisis de multiresolución), esto conlleva a que se pueden estudiar las señales en el dominio del tiempo y la frecuencia simultáneamente sin perder resolución.

Las wavelets deben su nombre a los trabajos de Alfred Haar [27], Jean Morlet [28], Alexander Grossman [29], Ingrid Daubechies [30], Yves Meyer [31] entre otros. El interés por las wavelets ha crecido en las últimas dos décadas por diferentes razones. Por un lado, el concepto de wavelet puede verse como una síntesis de ideas que se originaron en los últimos treinta años en el área de la ingeniería, de la física y de las matemáticas puras, por otro lado, las wavelets proveen una herramienta matemática con una gran variedad de aplicaciones. Por ejemplo, en el análisis de señales de sonido e imágenes, o en análisis numérico; la transformada wavelet aparece como una necesidad de analizar funciones no estacionarias en su escala de frecuencia-tiempo.

En la transformada wavelet, la escala juega un papel muy importante, dicho proceso analiza los datos en diferentes escalas o resoluciones. Es decir, si se toma una ventana de los datos, notaríamos el mismo comportamiento que si se toma una ventana más pequeña, y así sucesivamente. La transformada wavelet provee una representación en

frecuencia-tiempo simultáneamente, corta la señal de interés en varias partes y analiza cada una de ellas por separado. Las frecuencias altas pueden ser mejor localizadas en el tiempo y las frecuencias bajas en la frecuencia. La transformada wavelet, al igual que la transformada de Fourier, es una transformación reversible, que permite ir de la señal original a la transformada y viceversa [32].

II.1 Fundamentos matemáticos

La wavelet (de ahora en adelante denotada como ψ), vista como un proceso matemático, es una operación lineal que descompone una señal compleja en bloques elementales que aparecen en diferentes escalas o resoluciones. La selección de la mejor wavelet depende de las características propias de la señal de estudio; entonces, el objetivo es encontrar la mejor función base. A esta función se le conoce como wavelet madre. Sólo aquellas funciones que satisfagan las siguientes condiciones pueden ser consideradas como wavelet madre [33]:

1. Tener energía finita

$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty. \quad (1)$$

2. Si $\hat{\psi}(w)$ es la transformada de Fourier de $\psi(t)$, es decir

$$\hat{\psi}(w) = \int_{-\infty}^{\infty} \psi(t) e^{-i(2\pi w)t} dt, \quad (2)$$

entonces la siguiente condición debe cumplirse

$$C_g = \int_0^{\infty} \frac{|\hat{\psi}(w)|^2}{|w|} dw < \infty. \quad (3)$$

A la ec. 3 se le conoce como la condición de admisibilidad y a C_g como la constante de admisibilidad [32, 33].

Definición.

Dada una función $f(t) \in L^2(\mathbb{R})$, es decir, f es una función cuadrado integrable, la **transformada wavelet** se define como

$$W_{\psi_{a,b}}\{f(t)\} = \int_{-\infty}^{\infty} f(t) \overline{\psi_{a,b}(t)} dt, \quad (4)$$

donde

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \quad (5)$$

y ψ es la wavelet de análisis y a, b son parámetros de escala y traslación, respectivamente [32].

II.2 Propiedades

II.2.1 Linealidad

Sean $f, g \in L^2(\mathbb{R})$ y $\alpha, \beta \in \mathbb{R}$, entonces

$$W_{\psi_{a,b}}\{\alpha f + \beta g\} = \alpha W_{\psi_{a,b}}\{f(t)\} + \beta W_{\psi_{a,b}}\{g(t)\}. \quad (6)$$

Demostración.

$$\begin{aligned} W_{\psi_{a,b}}\{\alpha f + \beta g\} &= \int_{-\infty}^{\infty} (\alpha f + \beta g) \overline{\psi_{a,b}(t)} dt = \alpha \int_{-\infty}^{\infty} f(t) \overline{\psi_{a,b}(t)} dt + \beta \int_{-\infty}^{\infty} g(t) \overline{\psi_{a,b}(t)} dt \\ &= \alpha W_{\psi_{a,b}}\{f(t)\} + \beta W_{\psi_{a,b}}\{g(t)\}. \end{aligned}$$

■

II.2.2 Simetría

Sean ψ, φ ondículas, entonces

$$W_{\psi_{a,b}}\{\varphi(t)\} = |a|W_{\varphi_{\frac{1}{a}, -\frac{b}{a}}}^*\{\psi(t)\}. \quad (7)$$

Demostración.

$$W_{\psi_{a,b}}\{\varphi(t)\} = \int_{-\infty}^{\infty} \varphi(t)\psi_{a,b}^*(t)dt = \int_{-\infty}^{\infty} \varphi(t)\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right)dt,$$

haciendo el siguiente cambio de variable: $u = \frac{t-b}{a}$, $du = \frac{dt}{a}$,

$$\begin{aligned} W_{\psi_{a,b}}\{\varphi(t)\} &= \int_{-\infty}^{\infty} \varphi(ua+b)\frac{1}{\sqrt{|a|}}\bar{\psi}(u)|a|du = |a| \int_{-\infty}^{\infty} \frac{1}{\sqrt{|a|}}\varphi\left(\frac{t-(-b)}{\frac{1}{a}}\right)\psi^*(t)dt \\ &= |a| \int_{-\infty}^{\infty} \overline{\varphi_{\frac{1}{a}, -\frac{b}{a}}(t)}\psi(t)dt = |a| \int_{-\infty}^{\infty} \psi(t)\overline{\varphi_{\frac{1}{a}, -\frac{b}{a}}^*(t)}dt = |a|W_{\varphi_{\frac{1}{a}, -\frac{b}{a}}}^*\{\psi(t)\}. \end{aligned}$$

■

II.2.3 Traslación

Sea $f \in L^2(\mathbb{R})$, $c \in \mathbb{R}$, entonces

$$W_{\psi_{a,b}}\{f(t-c)\} = W_{\psi_{a,b-c}}\{f(t)\}.$$

Demostración.

$$W_{\psi_{a,b}}\{f(t-c)\} = \int_{-\infty}^{\infty} f(t-c)\psi_{a,b}^*(t)dt,$$

haciendo el siguiente cambio de variable, $u = t - c$, $du = dt$,

$$\begin{aligned} W_{\psi_{a,b}}\{f(t-c)\} &= \int_{-\infty}^{\infty} f(u)\psi_{a,b}^*(u+c)du = \int_{-\infty}^{\infty} f(t)\psi_{a,b}^*(t-(-c))dt \\ &= \int_{-\infty}^{\infty} f(t)\psi_{a,b-c}^*(t)dt = W_{\psi_{a,b-c}}\{f(t)\}. \end{aligned}$$

■

II.2.4 Dilatación

Sea $f \in L^2(\mathbb{R})$, $c \in \mathbb{R}^+$, entonces

$$W_{\psi_{a,b}} \left\{ \frac{1}{c} f\left(\frac{t}{c}\right) \right\} = W_{\psi_{\frac{a}{c}, \frac{b}{c}}} \{f(t)\}.$$

Demostración.

$$W_{\psi_{a,b}} \left\{ \frac{1}{c} f\left(\frac{t}{c}\right) \right\} = \int_{-\infty}^{\infty} \frac{1}{c} f\left(\frac{t}{c}\right) \psi_{a,b}^*(t) dt,$$

haciendo el siguiente cambio de variable, $u = \frac{t}{c}$, $du = \frac{dt}{c}$,

$$\begin{aligned} W_{\psi_{a,b}} \left\{ \frac{1}{c} f\left(\frac{t}{c}\right) \right\} &= \frac{1}{c} \int_{-\infty}^{\infty} f(u) \psi_{a,b}^*(uc) du = \int_{-\infty}^{\infty} f(t) \psi_{a,b}^*(tc) dt \\ &= \int_{-\infty}^{\infty} f(t) \overline{\frac{1}{\sqrt{|a|}} \psi\left(\frac{ct-b}{a}\right)} dt = \int_{-\infty}^{\infty} f(t) \overline{\frac{1}{\sqrt{|a|}} \psi\left(\frac{t-\frac{b}{c}}{\frac{a}{c}}\right)} dt \\ &= \int_{-\infty}^{\infty} f(t) \psi_{\frac{a}{c}, \frac{b}{c}}^*(t) dt = W_{\psi_{\frac{a}{c}, \frac{b}{c}}} \{f(t)\}. \end{aligned}$$

■

II.2.5 Antilinealidad

Sean ψ, φ ondículas, $f \in L^2(\mathbb{R})$, $\alpha, \beta \in \mathbb{C}$, entonces

$$W_{\alpha\psi_{a,b} + \beta\varphi_{a,b}} \{f(t)\} = \bar{\alpha} W_{\psi_{a,b}} \{f(t)\} + \bar{\beta} W_{\varphi_{a,b}} \{f(t)\}. \quad (8)$$

Demostración.

$$\begin{aligned} W_{\alpha\psi_{a,b} + \beta\varphi_{a,b}} \{f(t)\} &= \int_{-\infty}^{\infty} f(t) (\overline{\alpha\psi_{a,b} + \beta\varphi_{a,b}}) dt = \int_{-\infty}^{\infty} f(t) (\overline{\alpha\psi_{a,b}} + \overline{\beta\varphi_{a,b}}) dt \\ &= \int_{-\infty}^{\infty} f(t) \overline{\alpha\psi_{a,b}(t)} dt + \int_{-\infty}^{\infty} f(t) \overline{\beta\varphi_{a,b}(t)} dt = \bar{\alpha} W_{\psi_{a,b}} \{f(t)\} + \bar{\beta} W_{\varphi_{a,b}} \{f(t)\}. \end{aligned}$$

■

II.3 Ejemplos de wavelets

El término ψ es el que define a la wavelet que se está utilizando. Existen muchas ya definidas, a continuación se mencionan las más utilizadas.

II.3.1 Wavelet Sombrero Mexicano

La wavelet Sombrero Mexicano (o wavelet de Ricker) se emplea con frecuencia para modelar datos sísmicos y para la electrodinámica computacional. Esta wavelet se define utilizando la segunda derivada de la función gaussiana, como

$$\psi(t) = (1 - 2t^2)e^{-t^2}. \quad (9)$$

La gráfica de la ec.(9) se muestra en la Fig. 1.

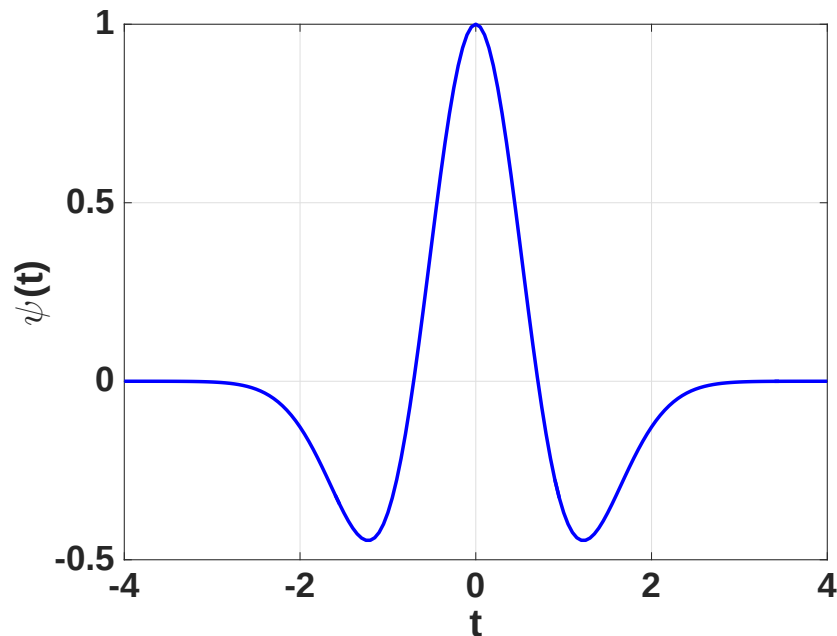


Figura 1: Gráfica de la wavelet Sombrero Mexicano.

II.3.2 Wavelet de Morlet

La wavelet de Morlet está compuesta de una exponencial compleja multiplicada por una ventana gaussiana. Esta wavelet está estrechamente relacionada con la percepción humana, tanto en audiencia como en la visión. Sus principales aplicaciones son en la medicina, donde sirve para clarificar las interpretaciones de espectros de trauma; y en la transcripción de la música. Matemáticamente está definida como

$$\psi(t) = e^{i\omega_0 t - \frac{t^2}{2}}. \quad (10)$$

La gráfica de la ec.(10) se muestra en la Fig. 2.

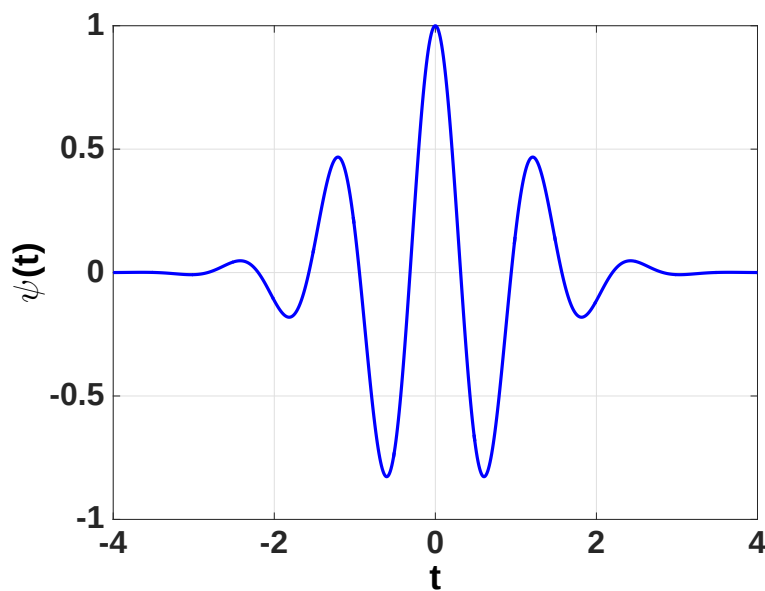


Figura 2: Gráfica de la wavelet de Morlet.

II.3.3 Wavelet de Meyer

La wavelet de Meyer es una wavelet ortogonal propuesta por Yves Meyer. Esta definida en el dominio de la frecuencia en términos de la función ν como:

$$\psi(\omega) = \begin{cases} \frac{1}{\sqrt{2\pi}} \sin\left(\frac{\pi}{2}\nu\left(\frac{3|\omega|}{2\pi} - 1\right)\right) e^{j\omega/2} & \text{si } 2\pi/3 < |\omega| < 4\pi/3, \\ \frac{1}{\sqrt{2\pi}} \cos\left(\frac{\pi}{2}\nu\left(\frac{3|\omega|}{4\pi} - 1\right)\right) e^{j\omega/2} & \text{si } 4\pi/3 < |\omega| < 8\pi/3, \\ 0 & \text{de otra manera,} \end{cases} \quad (11)$$

donde

$$\nu(x) = \begin{cases} 0 & \text{si } x < 0, \\ x & \text{si } 0 < x < 1, \\ 1 & \text{si } x > 1. \end{cases} \quad (12)$$

La gráfica de la wavelet de Meyer se muestra en la Fig. 3.

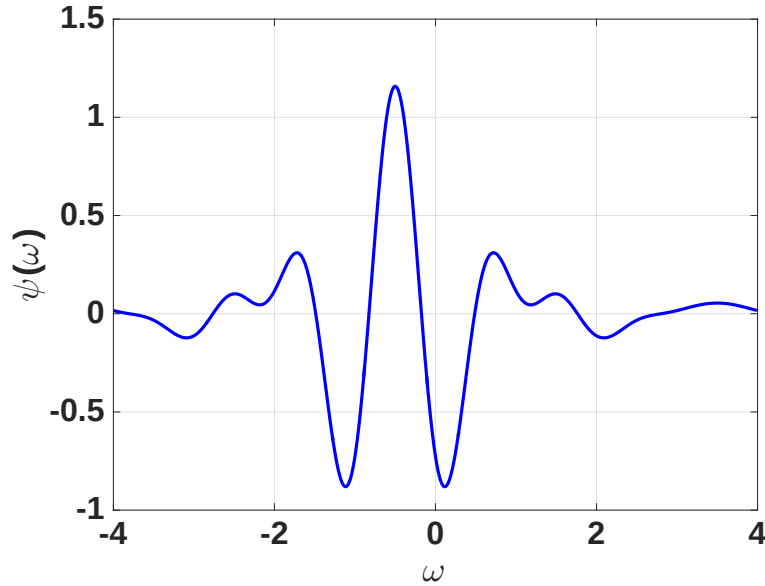


Figura 3: Gráfica de la wavelet de Meyer.

II.3.4 Wavelet de Haar

La wavelet de Haar es la más sencilla de implementar y es utilizada mayormente para la compresión de imágenes. En esta wavelet el término ψ esta definido como,

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2}, \\ -1, & \frac{1}{2} \leq t < 1, \\ 0, & \text{de otra manera.} \end{cases} \quad (13)$$

La gráfica de la ec.(13) se observa en la Fig. 4.

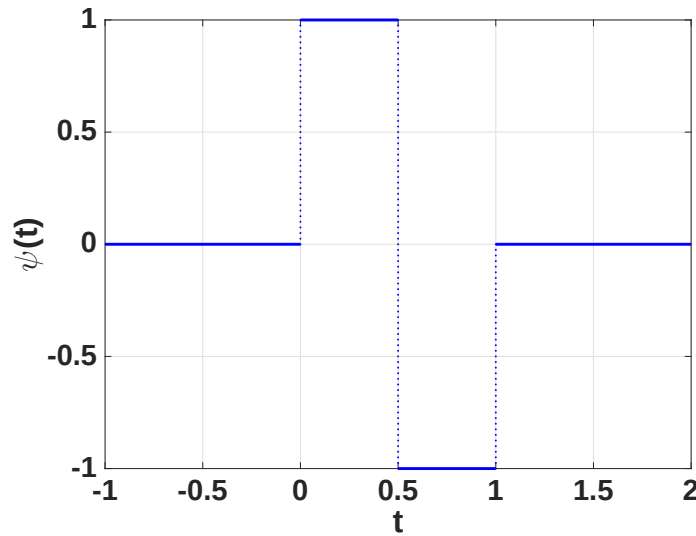


Figura 4: Gráfica de la wavelet de Haar.

II.4 Transformada wavelet en dos dimensiones

Cuando la transformada wavelet se aplica en dos dimensiones, se divide la imagen en cuatro sub-imágenes (cada una de la mitad de tamaño de la imagen de referencia), con diferentes frecuencias: altas-altas (AA), altas-bajas (AB), bajas-altas (BA) y bajas-

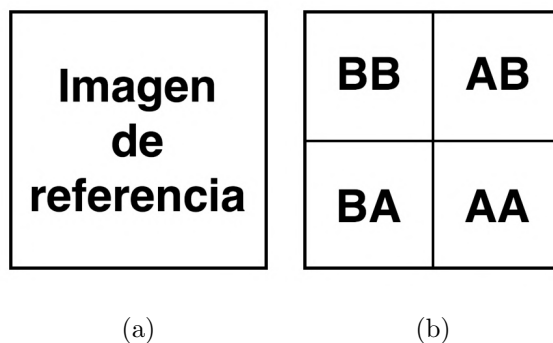


Figura 5: (a) Imagen de referencia. (b) Boceto de la transformada wavelet aplicada en dos dimensiones.

bajas (BB). En la Figura 5 se muestra un boceto de este procedimiento.

En este trabajo se utilizó la wavelet de Haar con las sub-imágenes AA, BA y AB, ya que estas frecuencias dan información relevante de la imagen, resaltando los bordes diagonales, verticales y horizontales. La Fig. 6 muestra un ejemplo de estas frecuencias. La frecuencia BB es descartada porque extrae los detalles burdos de la imagen y no agrega información relevante al estudio.

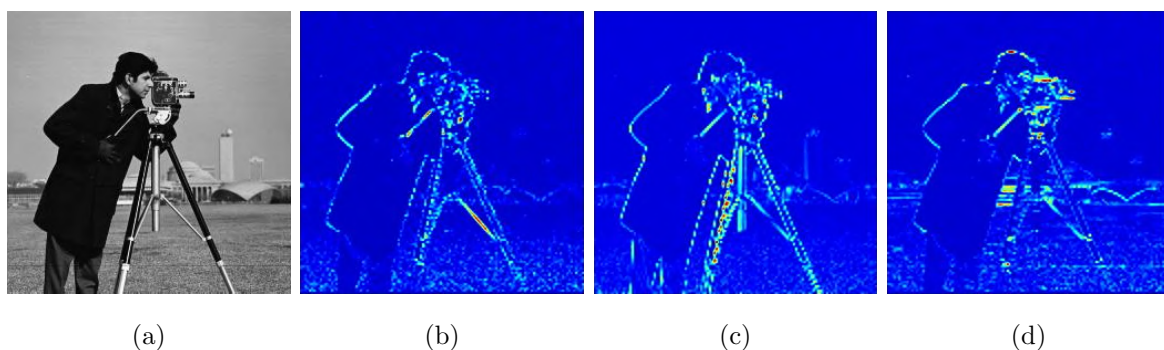


Figura 6: Sub-imágenes resultantes de la transformada wavelet en dos dimensiones utilizando la wavelet de Haar. (a) Imagen de referencia. (b) Frecuencias AA. (c) Frecuencias BA. (d) Frecuencias AB.

Capítulo III

Descriptores de rasgos locales

Las metodologías de reconocimiento de patrones y clasificación basados en rasgos locales tienen como propósito encontrar en una imagen puntos clave que representen información relevante de tal manera que se pueda distinguir de una imagen diferente. Estos rasgos o características pueden ser puntos, bordes o una pequeña porción de la imagen. Ya que se tienen los puntos clave localizados, a cada uno de estos se le asocia un descriptor, el cual es un vector que contiene las características del punto clave. Utilizar puntos clave y descriptores reduce los costos de cómputo y procesamiento ya que se trabaja con una dimensión menor que el de las imágenes. En general las metodologías para generar descriptores de rasgos locales siguen las siguientes etapas:

1. **Generación del espacio escala:** Se generan imágenes con diferentes escalas con el propósito de obtener la invariancia a escala y realzar las características de interés para poder localizar los puntos clave.
2. **Localización de los puntos clave:** El objetivo de este paso es localizar en el espacio escala los puntos que cumplan los criterios establecidos que los caracterizan como puntos clave, dichos criterios varían de acuerdo a la metodología.
3. **Generación del descriptor del punto clave:** Se busca obtener un vector que contenga información suficiente para describir al punto clave y poder distinguirlo del resto, usualmente se analiza una vecindad centrada en el punto clave.

4. **Emparejamiento de descriptores:** En la imagen de referencia y en la imagen problema, se empareja al par de descriptores que tenga la distancia mínima y además, cumpla criterios adicionales establecidos de acuerdo a la metodología utilizada.

En este capítulo se estudiarán las dos metodologías de localización de rasgos locales más utilizados: SIFT (por sus siglas en inglés: *Scale Invariant Feature Transform*) y SURF (por sus siglas en inglés: *Speeded-Up Robust Features*).

III.1 SIFT

La metodología SIFT fue creada por David G. Lowe, la primera versión se publicó en 1999 [2], y la versión completa y patentada aparece en 2004 [3]. Esta metodología consta de cuatro etapas principales:

1. **Generación del espacio de diferencias de gaussianas.**
2. **Localización de los puntos clave.**
3. **Asignación de una orientación a cada punto clave.**
4. **Asignación del descriptor a cada punto clave.**

III.1.1 Generación del espacio de diferencias de gaussianas

Con el propósito de lograr la invariancia a escala, la primera etapa de la metodología es generar el espacio escala. Para crear este espacio se toma la imagen de referencia y se generan imágenes con diferente nivel de desenfoque, a este conjunto de imágenes se le llama octave (primer renglón en la Fig. 7). Después, la imagen de referencia se escala a la mitad de su tamaño y con esta nueva imagen se generan las imágenes con diferente nivel de desenfoque para generar el segundo octave (segundo renglón en la Fig. 7) y así sucesivamente. Al conjunto de todas las imágenes generadas en los diferentes octaves se le denomina espacio escala, por ejemplo todas las imágenes de la Fig. 7 forman un espacio escala.

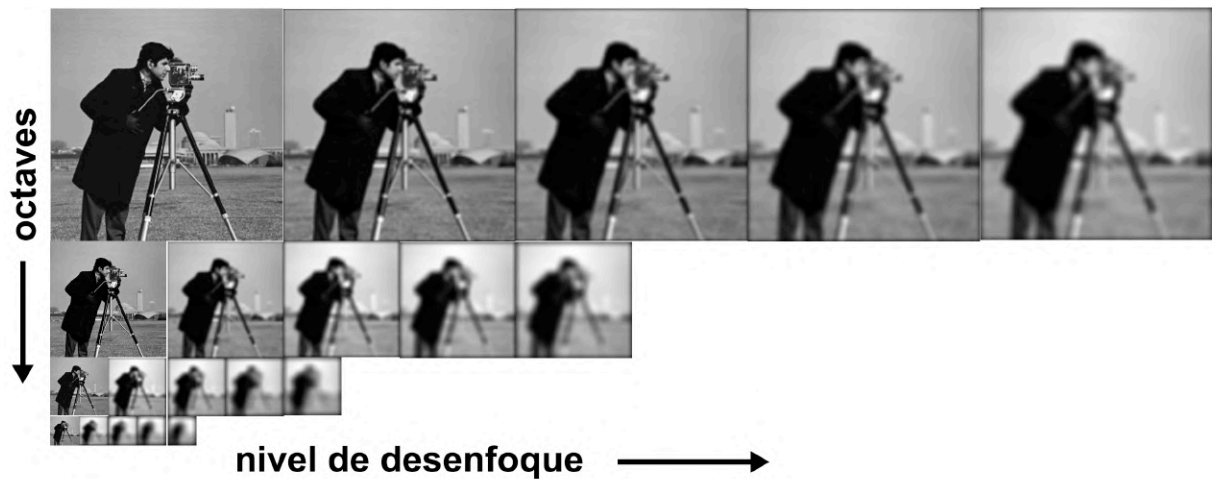


Figura 7: Ejemplo del espacio escala.

El número de octaves y de niveles de desenfoque depende del tamaño de la imagen de referencia, Lowe [3] sugiere que cuatro octaves y cinco niveles de desenfoque son suficientes para obtener buenos resultados. El desenfoque que se utiliza en SIFT es la convolución de una función gaussiana, $G(x, y, \sigma)$, con la imagen de referencia $I(x, y)$, matemáticamente esto es

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (14)$$

donde $*$ es la operación de convolución, $x = 1, \dots, M$, $y = 1, \dots, N$, $M, N \in \mathbb{R}$, $M \times N$ es el tamaño de la imagen de referencia, σ es el parámetro de desviación estándar y

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \quad (15)$$

Para detectar eficientemente puntos clave estables en el espacio escala, Lowe utiliza una función generada a partir de una diferencia de gaussianas, ésta se calcula mediante la diferencia de dos imágenes consecutivas del mismo octave las cuales difieren en el nivel de desenfoque, matemáticamente esto es

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma), \quad (16)$$

donde k es el factor de desenfoque. En la Fig. 8 se observan los resultados de aplicar la ec.(16) a las imágenes del espacio escala de la Fig. 7. Al conjunto de imágenes de la Fig.8 se le conoce como el espacio de diferencias de gaussianas.

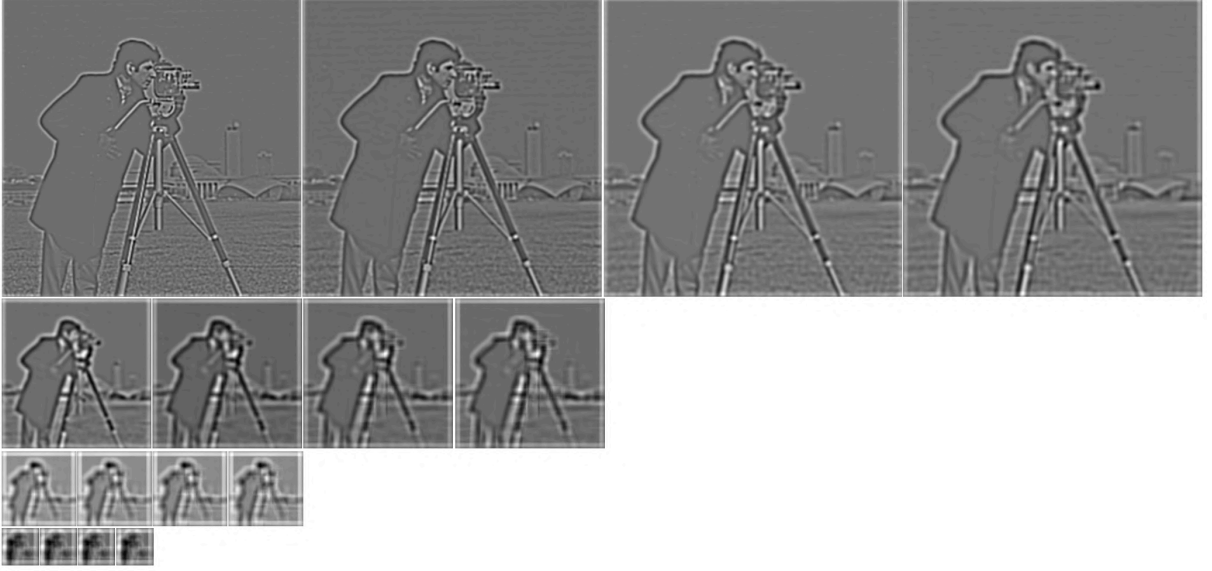


Figura 8: Ejemplo del espacio de diferencias de gaussianas.

III.1.2 Localización de los puntos clave

Para detectar los primeros candidatos a puntos clave en el espacio de diferencias de gaussianas, se analiza el nivel de intensidad de los píxeles en el espacio de diferencias de gaussianas, por ejemplo el nivel de intensidad del píxel marcado en rojo en la Fig. 9, y se compara con los niveles de intensidad de los píxeles de sus ocho vecinos en el mismo nivel (píxeles marcados en verde en la Fig. 9), y también se compara con los niveles de intensidad de los dieciocho píxeles en la vecindad equivalente a la posición del posible candidato a punto clave en los dos niveles adyacentes (píxeles marcados en azul en la Fig. 9). Si el valor de intensidad del píxel marcado en rojo en la Fig. 9 es el máximo o el mínimo del conjunto de los veintisiete valores de intensidad se toma como candidato a punto clave. En esta etapa se tiene el primer conjunto de candidatos a puntos clave $\mathbf{x}_j = (x_j, y_j) \in D(x, y, \sigma)$ donde $j = 0, 1, 2, \dots, p$ con $p \leq M \times N$ el número de candidatos a puntos clave.

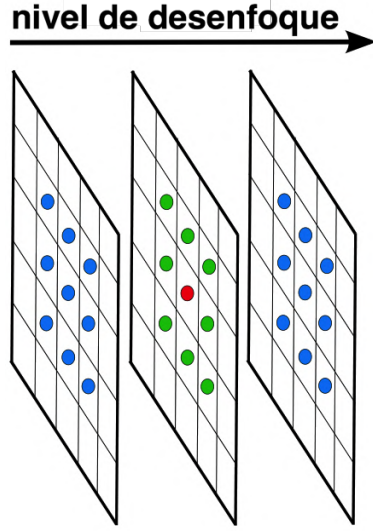


Figura 9: Bosquejo de la selección de candidatos a puntos clave.

Una vez que se seleccionaron todos los candidatos a puntos clave, $\mathbf{x}_j = (x_j, y_j)$, estos tienen que pasar por un siguiente filtro, que sirve para excluir a los puntos de bajo contraste. Esto se realiza utilizando la expansión de Taylor de segundo orden a la función del espacio de diferencias de gaussianas, $D(\mathbf{x}, \sigma)$,

$$\tilde{D}(\mathbf{x}, \sigma) = D(\mathbf{x}_j, \sigma) + (\mathbf{x} - \mathbf{x}_j)^T \frac{\partial D(\mathbf{x}_j, \sigma)}{\partial \mathbf{x}} + \frac{1}{2} (\mathbf{x} - \mathbf{x}_j)^T \frac{\partial^2 D(\mathbf{x}_j, \sigma)}{\partial \mathbf{x}^2} (\mathbf{x} - \mathbf{x}_j), \quad (17)$$

donde $\mathbf{x} = (x, y)$ pertenece a una vecindad Ω centrada en $\mathbf{x}_j$, Lowe no especifica el tamaño de dicha vecindad en ninguno de los dos artículos publicados por él [2, 3]. Después se localiza el valor extremo $\hat{\mathbf{x}}$, el cual servirá para rechazar puntos que tienen bajo contraste, esto es

$$\nabla \tilde{D}(\hat{\mathbf{x}}, \sigma) = 0. \quad (18)$$

Mediante experimentos computacionales, Lowe [3] llegó a la conclusión de que todos los valores extremos que cumplan que $|\tilde{D}(\hat{\mathbf{x}}, \sigma)| < 0.03$ deben de ser eliminados, porque presentan bajo contraste y no representan información relevante de la imagen.

El siguiente paso es eliminar los puntos que son falsos bordes, esto se hace verificando si la curvatura del candidato a punto clave es o no menor a un umbral. La curvatura se calcula utilizando la matriz Hessiana:

$$H(\mathbf{x}_j, \sigma) = \begin{bmatrix} \frac{\partial^2 D(\mathbf{x}_j, \sigma)}{\partial x^2} & \frac{\partial^2 D(\mathbf{x}_j, \sigma)}{\partial x \partial y} \\ \frac{\partial^2 D(\mathbf{x}_j, \sigma)}{\partial y \partial x} & \frac{\partial^2 D(\mathbf{x}_j, \sigma)}{\partial y^2} \end{bmatrix}, \quad (19)$$

de la cual Lowe considera que se satisfaga

$$\frac{tr^2(H(\mathbf{x}_j, \sigma))}{Det(H(\mathbf{x}_j, \sigma))} < \frac{(\gamma + 1)^2}{\gamma}, \quad (20)$$

para no descartar a $\mathbf{x}_j$ como punto clave, usando el umbral $\gamma = 10$, $tr(H(\mathbf{x}_j), \sigma)$ es la traza de $H(\mathbf{x}_j, \sigma)$ y $Det(H(\mathbf{x}_j, \sigma))$ es el determinante de $H(\mathbf{x}_j, \sigma)$.

III.1.3 Asignación de una orientación a cada punto clave

Ya que se seleccionaron los puntos clave, el siguiente paso es asignarle a cada uno de ellos una orientación, esto con el propósito de obtener la invariancia a rotación. El procedimiento consiste en seleccionar el punto clave y una vecindad alrededor de él. A cada píxel en esa vecindad se le calcula la magnitud y orientación, ec.(21) y ec.(22), respectivamente, utilizando diferencias de píxeles [34].

$$m(x, y) = \sqrt{\left(L(x+1, y) - L(x-1, y)\right)^2 + \left(L(x, y+1) - L(x, y-1)\right)^2}, \quad (21)$$

$$\theta(x, y) = \tan^{-1} \left[\left(L(x, y+1) - L(x, y-1)\right) / \left(L(x+1, y) - L(x-1, y)\right) \right]. \quad (22)$$

El siguiente paso es generar un histograma de gradientes orientados de 36 *bins*, donde cada entrada muestrea 10° , para cubrir de esta manera los 360° . El histograma se construye al localizar la orientación de cada píxel en uno de los 36 *bins* y acumular las magnitudes correspondientes. Ya que se realiza este procedimiento con todos los píxeles de la vecindad, el histograma tendrá un *bin* en el cual la acumulación será la máxima, esta orientación será la que se le asignará al punto clave. La Fig. 10 muestra un ejemplo donde se resalta que la orientación asignada será el *bin* correspondiente al rango de $10^\circ - 19^\circ$.

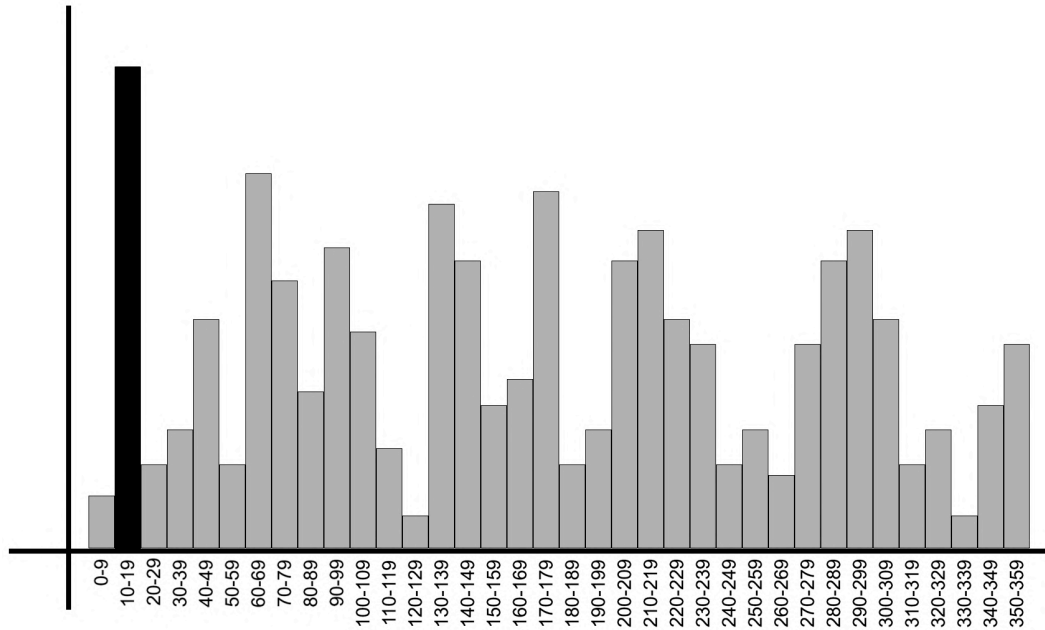


Figura 10: Ejemplo de un histograma de gradientes orientados de 36 entradas.

III.1.4 Asignación del descriptor a cada punto clave

Ya que cada punto clave tiene una orientación, lo siguiente es asignarle a cada uno su descriptor. El primer paso para generar el descriptor es tomar una ventana de 16×16 píxeles centrada en el punto clave y girarla la orientación asignada. El siguiente paso es dividir esta ventana en 16 ventanas de tamaño 4×4 píxeles (Fig. 11a). Después, se genera para cada una de las 16 ventanas un histograma de gradientes orientados de 8 *bins*, cada entrada muestrea 45° (Fig. 11b). Finalmente, se construye un vector de 128 elementos con el valor de cada uno de los *bins* de los 16 histogramas, como lo muestra la Fig. 11c. Este vector de 128 entradas es el descriptor del punto clave.

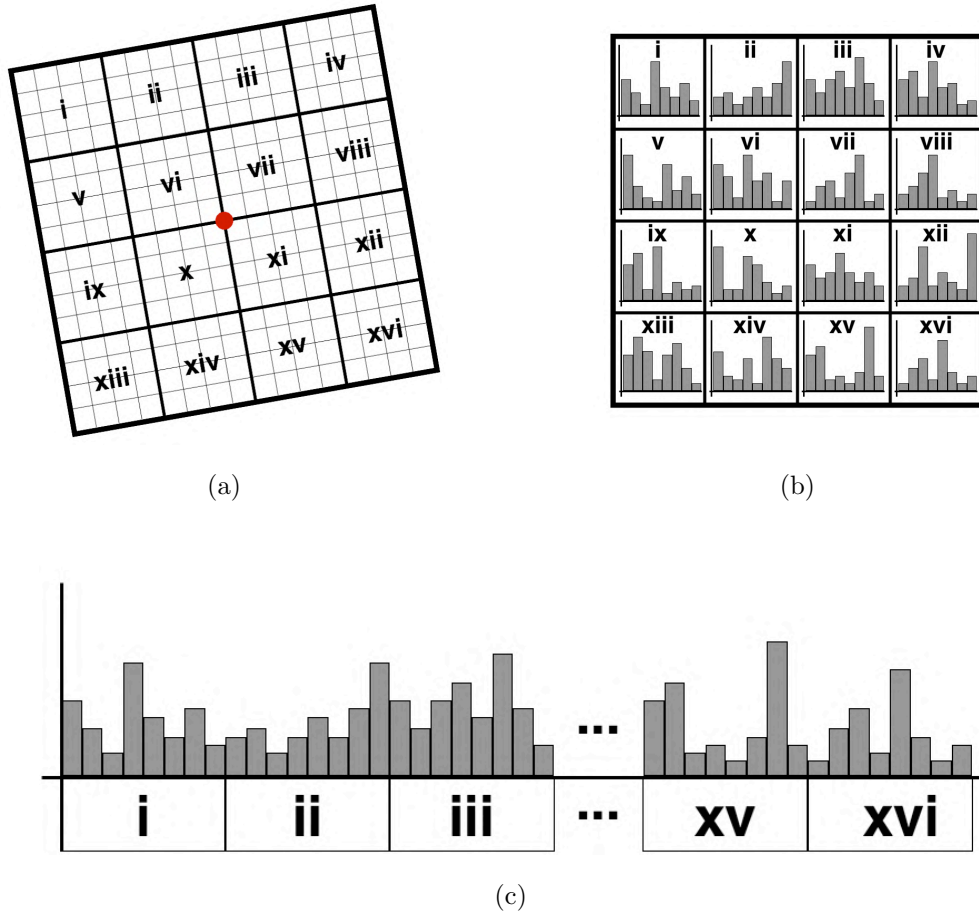


Figura 11: Esquema del proceso para generar el descriptor de un punto clave.

La Fig. 12 muestra un diagrama de bloques de los pasos para construir el descriptor SIFT.

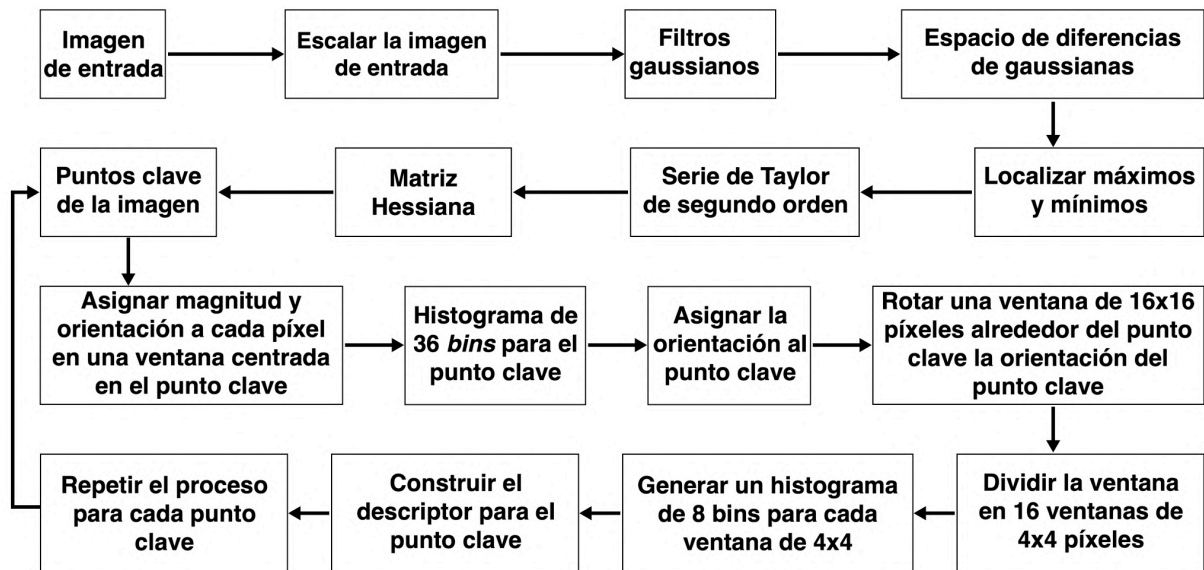


Figura 12: Diagrama de bloques para generar el descriptor SIFT.

III.2 SURF

Una de las grandes desventajas de SIFT es su considerable tiempo de cómputo, por lo cual, en 2006 se publicó SURF [16] el cual promete ser una versión rápida de SIFT. Esta metodología consta de las siguientes etapas:

1. Imagen integral.
2. Generación del espacio escala.
3. Localización de los puntos clave.
4. Asignación de una orientación a cada punto clave.
5. Asignación del descriptor a cada punto clave.

III.2.1 Imagen integral

En lugar de utilizar la imagen de referencia SURF utiliza las imágenes integrales, ya que permiten hacer el proceso de convolución con mayor rapidez. Las imágenes integrales se definen de la siguiente manera,

$$I_{\Sigma}(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(x, y), \quad (23)$$

donde $I(x, y)$ es la imagen de referencia, $x = 1, \dots, M$, $y = 1, \dots, N$, $M, N \in \mathbb{R}$, $M \times N$ es el tamaño de $I(x, y)$. La Fig. 13 muestra un ejemplo de una imagen integral.

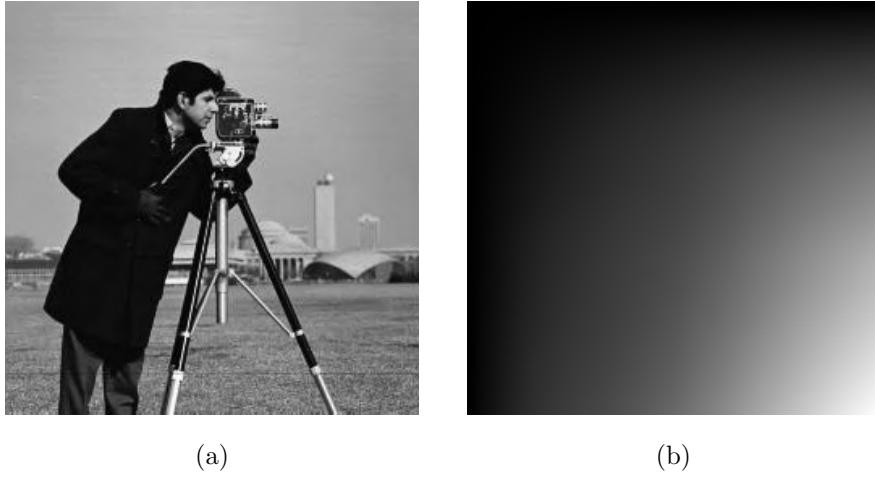


Figura 13: (a) Imagen de referencia. (b) Imagen integral de Fig. 13a.

III.2.2 Generación del espacio escala

El espacio escala de SIFT se representa con una pirámide de imágenes, donde éstas se desenfocan utilizando un filtro gaussiano y luego se obtienen imágenes de menor tamaño para construir los diferentes octaves de la pirámide (Fig. 7). En lugar de utilizar los filtros gaussianos de SIFT (Fig. 14a), SURF utiliza filtros de caja (Fig. 14b).

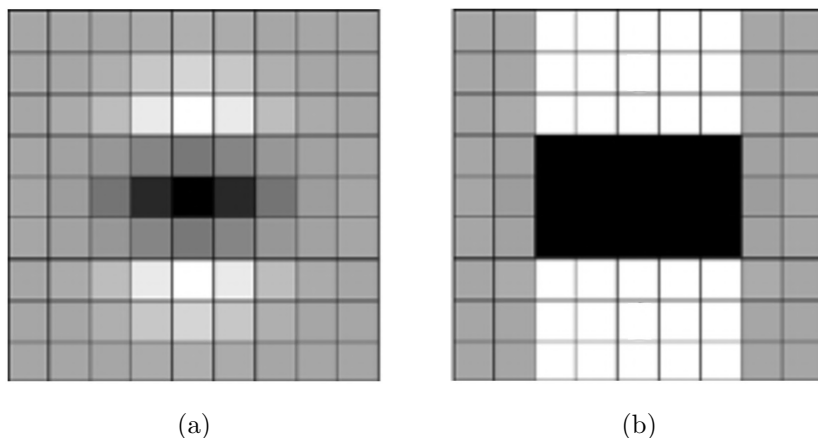


Figura 14: (a) Ejemplo de un filtro gaussiano utilizado por SIFT. (b) Ejemplo de un filtro de caja utilizado por SURF.

Entonces, para generar el espacio escala, se usan filtros de caja de diferentes tamaños, manteniendo el tamaño de la imagen de referencia siempre igual.

Para generar el primer octave de la pirámide, el primer filtro es de tamaño 9×9 píxeles con el cual se obtiene la primera imagen del octave. Después se aumenta el tamaño del filtro a 15×15 , 21×21 y 27×27 píxeles, con estos cuatro filtros se genera el resto de las imágenes del primer octave de la pirámide de escalas. Para el segundo octave se toman filtros de tamaño 15×15 , 27×27 , 39×39 y 51×51 píxeles, recordemos que el tamaño de la imagen de referencia se mantiene igual. En el tercer octave se utilizan filtros de tamaño 27×27 , 51×51 , 75×75 y 99×99 píxeles. Si la imagen de referencia es muy grande y se necesita un cuarto octave, se usan los filtros de tamaño 51×51 , 99×99 , 147×147 y 195×195 píxeles; este octave usualmente no se usa, ya que la cantidad de puntos clave detectados es casi nula.

III.2.3 Localización de los puntos clave

Para localizar los puntos clave en la imagen y en sus escalas, SURF utiliza la técnica de Neubeck y Van Gool [35] de supresión no máxima aplicada en una vecindad de $3 \times 3 \times 3$. Después, interpola los máximos del determinante de la matriz Hessiana en escala y en el espacio de la imagen con el método propuesto por Brown y Lowe [36].

III.2.4 Asignación de una orientación a cada punto clave

Ya que se tienen localizados los puntos clave, se tiene la invariancia a traslación y escala, por lo que hace falta la invariancia a rotación. El siguiente paso es asignarle una orientación a cada punto clave. Para esto, primeramente se aplican los filtros wavelet de Haar en dirección x y y de tamaño $4s$ (Fig. 15a y Fig. 15b, el color negro representa el valor escalar -1 y el blanco 1), dentro de una ventana circular de tamaño $6s$ alrededor del punto clave, donde s es la escala donde se localizó el punto clave. Las Figuras 16a y 16b muestran un bosquejo de este procedimiento, en la Fig. 16a se está aplicando el filtro wavelet de Haar en dirección x y en la Fig. 16b en dirección y , al sumar los resultados de aplicar estos filtros se obtienen los valores α_x y α_y , con los cuales se forma la coordenada (α_x, α_y) .

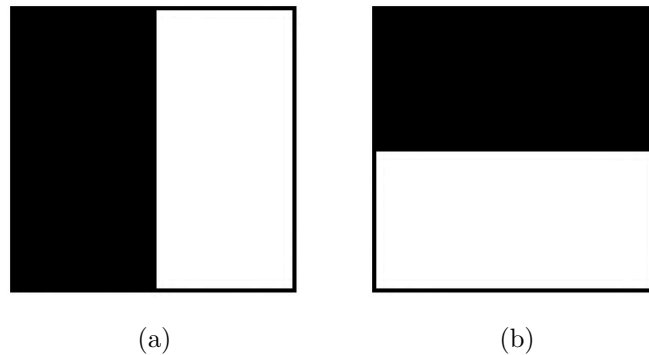


Figura 15: (a) Filtro wavelet de Haar en dirección x . (b) Filtro wavelet de Haar en dirección y . El color negro representa el valor escalar -1 y el blanco 1 .

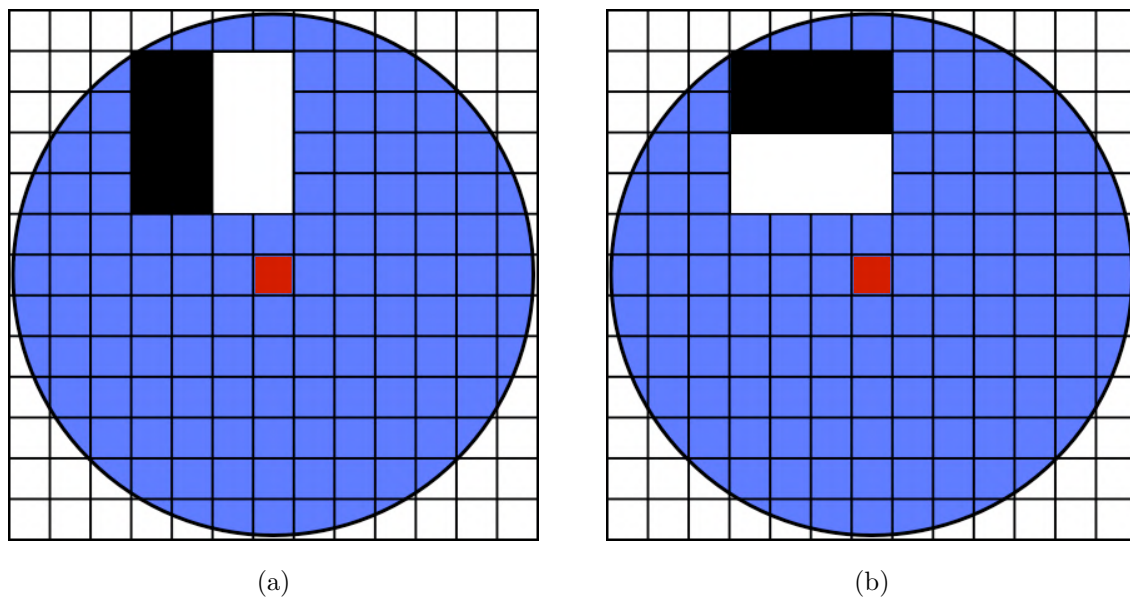


Figura 16: Boceto donde se aplica el filtro wavelet de Haar dentro de la ventana circular (color azul) alrededor del punto clave (píxel rojo). (a) Filtro wavelet de Haar en dirección x . (b) Filtro wavelet de Haar en dirección y .

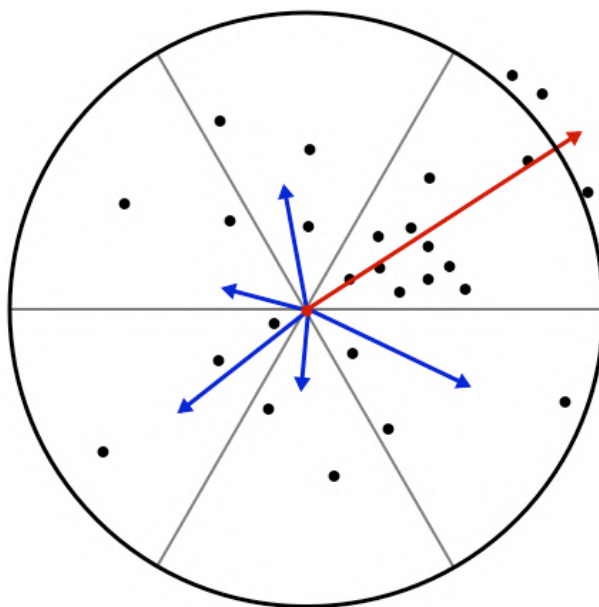


Figura 17: Boceto de la asignación de orientación a un punto clave.

Después de aplicar los filtros Haar dentro de toda la vecindad circular se obtienen varias coordenadas (α_x, α_y) , en la Fig. 17 se grafican como puntos negros. El siguiente paso es dividir la ventana circular en seis porciones del mismo tamaño y sumar todas las respuestas en la dirección horizontal ($\Sigma\alpha_x$) y vertical ($\Sigma\alpha_y$) dentro de cada porción, el resultado de este proceso son los vectores en la Fig. 17. Por último, el vector de magnitud mayor determinará la orientación del punto clave, vector en color rojo en la Fig. 17.

III.2.5 Asignación del descriptor a cada punto clave

Para construir el descriptor de un punto clave se toma una vecindad cuadrada de tamaño $20s$ centrada en el punto clave y rotada la orientación que se le asignó al punto clave previamente, como se muestra en la Fig. 18a. Después, esta vecindad se divide en 16 subregiones, y para cada una se calcula la respuesta de aplicar la wavelet de Haar en dirección x y y . Del filtro en dirección x se extraen dos escalares, el primero es la suma de las α_x (obtenidas de igual manera que en la subsección III.2.4) y el otro escalar es la suma de los valores absolutos de cada una de las multiplicaciones realizadas al aplicar el filtro, esto es $|\alpha_x|$. Análogamente se hace lo mismo con el filtro en dirección y . Por lo tanto se obtienen 4 valores para cada subregión $(\Sigma\alpha_x, \Sigma\alpha_y, \Sigma|\alpha_x|, \Sigma|\alpha_y|)$ (Fig. 18b) teniendo un total de $16 \times 4 = 64$ valores para cada punto clave. Estos valores se acomodan en forma de vector, como lo muestra la Fig 18c, y se obtiene el descriptor del punto clave.

La Fig. 19 muestra un diagrama de bloques de los pasos seguidos para construir el descriptor SURF.

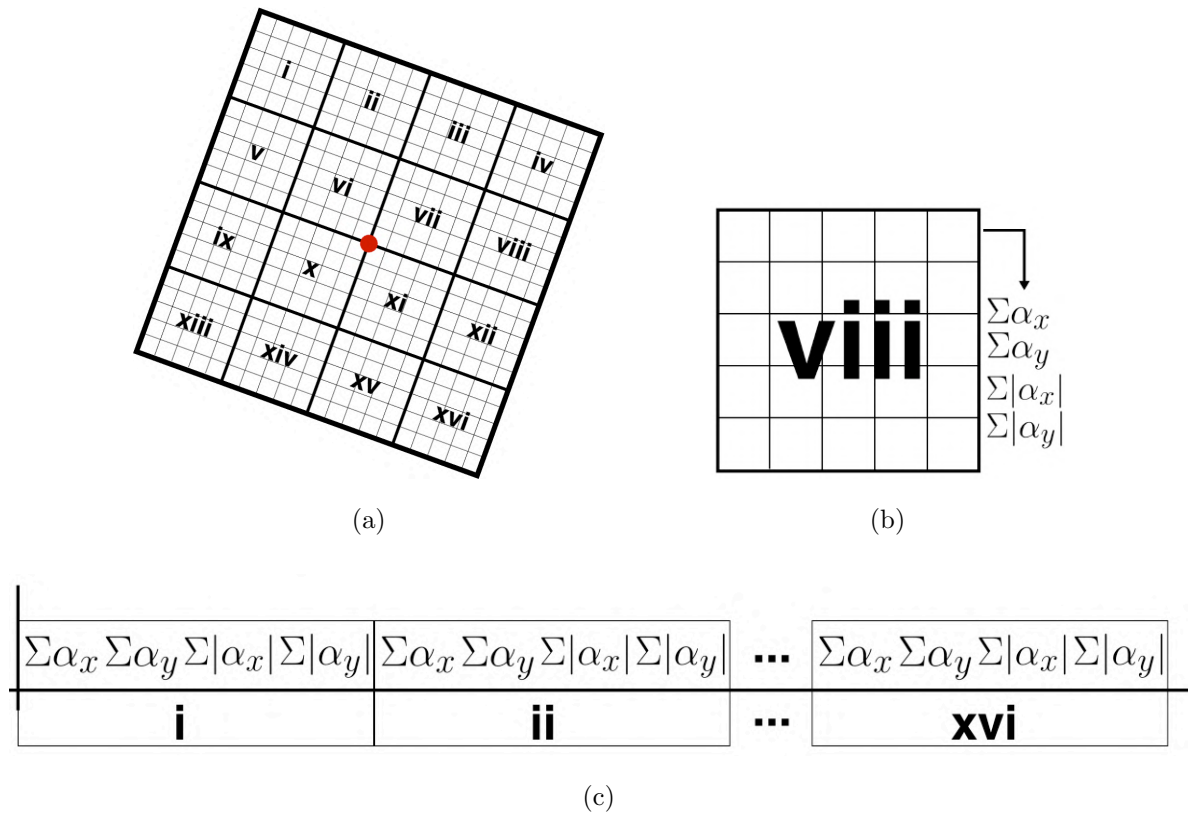


Figura 18: Ejemplo del procedimiento para generar el descriptor SURF.

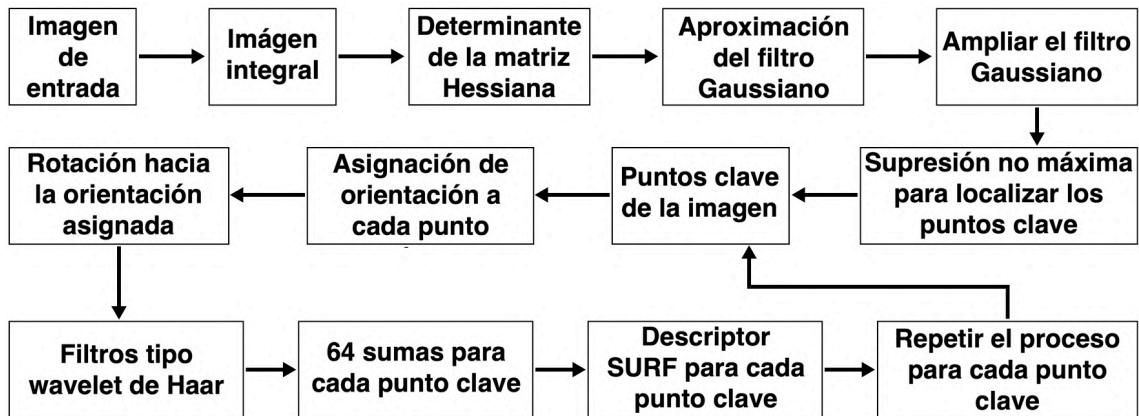


Figura 19: Diagrama de bloques de para generar el descriptor SURF.

Capítulo IV

W-SIFT

En esta sección se presenta la metodología W-SIFT, la cual es el trabajo realizado para este trabajo de tesis.

IV.1 Metodología

La metodología para generar el descriptor Wavelet SIFT (W-SIFT) consta de tres etapas principales:

1. **Generación de la pirámide wavelet.**
2. **Localización de los puntos clave.**
3. **Asignación del descriptor a cada punto clave.**

IV.1.1 Generación de la pirámide wavelet

Uno de los propósitos de construir una pirámide wavelet es crear un espacio donde se resalten las características más significativas de una imagen. Utilizando la transformada wavelet esto es posible, ya que una propiedad de ésta es la detección de bordes en una imagen. Otro propósito de la pirámide es obtener la invariancia a escala, esto se logra tomando imágenes de diferentes tamaños para crear la pirámide wavelet. En total, se generarán tres pirámides wavelet, usando las tres sub-imágenes obtenidas de la transformada wavelet de Haar (Fig. 6) como se hizo en el capítulo II.

El procedimiento para generar la primera pirámide, P_{AA} , consiste en tomar la imagen de referencia y escalarla de -50% a $+50\%$ con un tamaño de paso de 1% , obteniendo 101 imágenes (Fig. 20). Luego, se aplica la transformada wavelet de Haar a cada una de las imágenes escaladas tomando solamente las frecuencias AA, de esta manera se resaltan los detalles diagonales. Dado que la transformada wavelet reduce la imagen a la mitad, se tiene que agrandar al doble de tamaño las imágenes resultantes. La Fig. 21a muestra la pirámide P_{AA} . La segunda pirámide, P_{BA} , se genera tomando solo las frecuencias BA, realzando los detalles verticales, Fig. 21b. Finalmente, la tercera pirámide, P_{AB} , es creada usando las frecuencias AB, donde se destacan los detalles horizontales, Fig. 21c.

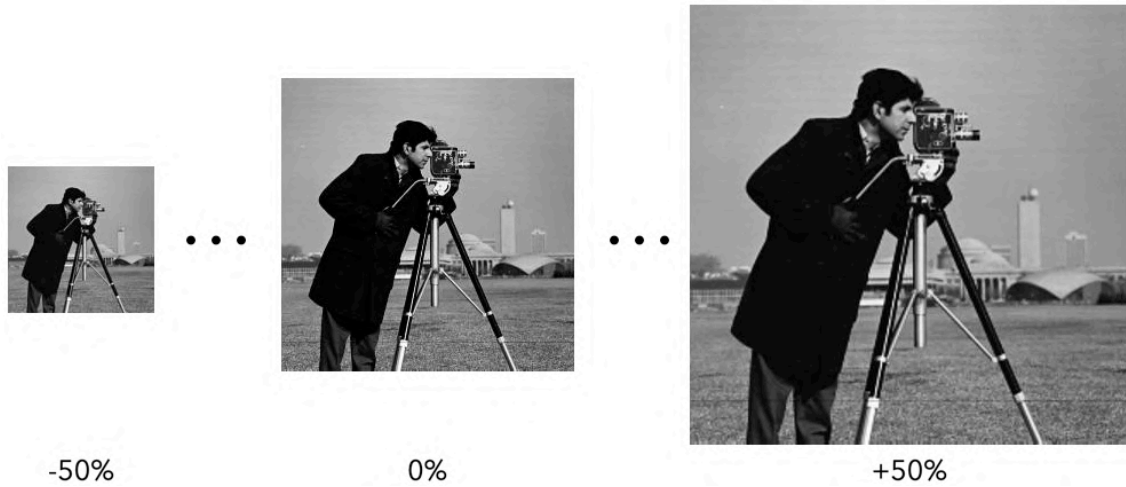


Figura 20: Imagen de referencia escalada de -50% a $+50\%$ con un tamaño de paso de 1% .

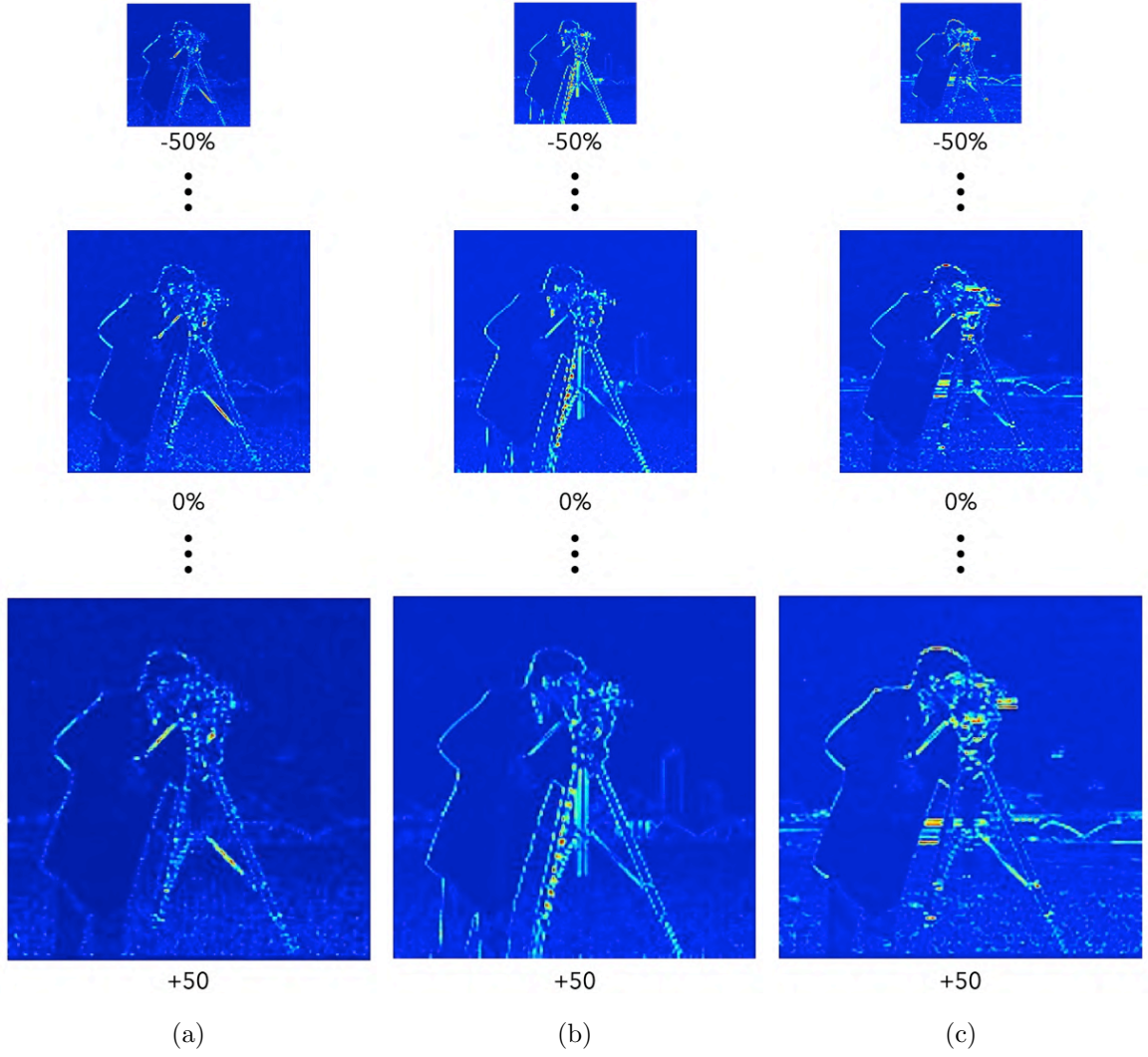


Figura 21: Pirámides wavelet. (a) Pirámide P_{AA} . (b) Pirámide P_{BA} . (c) Pirámide P_{AB} .

IV.1.2 Localización de los puntos clave

El primer paso para localizar los puntos clave es filtrar cada elemento de la pirámide con una máscara binaria, la cual tiene unos donde la imagen tenga un borde y ceros en cualquier otro lado. Esta máscara es construida utilizando el detector de bordes Sobel aplicado a la imagen de referencia y después escalado al mismo tamaño del nivel de la pirámide que se esta trabajando [37]. Por ejemplo, el primer nivel de P_{BA} es filtrado utilizando la máscara binaria de bordes M^1 . Matemáticamente esto es,

$$F_{BA}^1 = P_{BA}^1 \circ M^1, \quad (24)$$

donde el super índice indica el nivel de la pirámide y $\circ$ representa el producto Hadamard [38]. En la Fig. 22 se observa un ejemplo de este proceso.

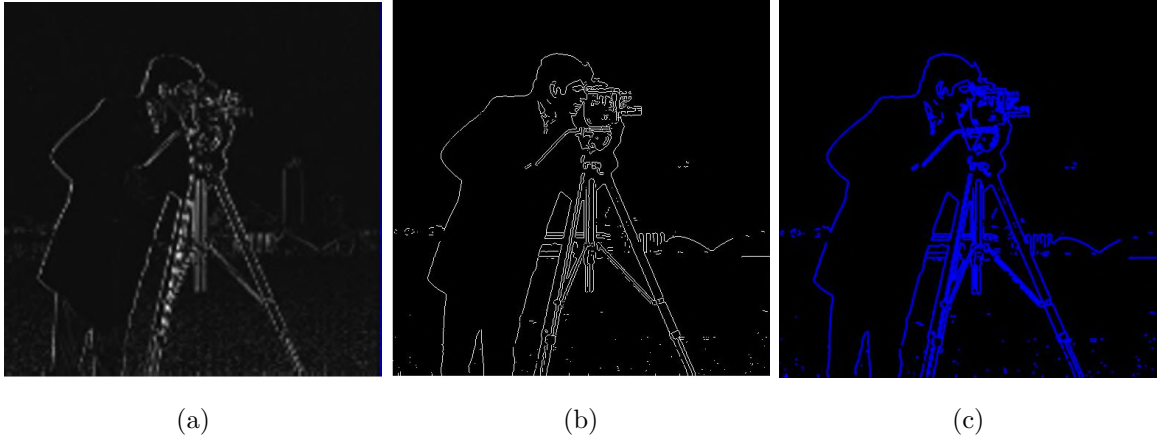


Figura 22: (a) P_{BA}^1 . (b) Máscara binaria de borde M^1 . (c) Imagen filtrada F_{BA}^1 .

El proceso de filtrado se realiza en las tres pirámides de la siguiente manera

$$F_{AA}^n = P_{AA}^n \circ M^n, \quad (25)$$

$$F_{BA}^n = P_{BA}^n \circ M^n, \quad (26)$$

$$F_{AB}^n = P_{AB}^n \circ M^n, \quad (27)$$

donde $n = 1, \dots, 101$.

Una vez que las tres pirámides son filtradas con las máscaras binarias de bordes, el siguiente paso es analizar si los valores de intensidad de los puntos que son diferentes de cero, llamadas puntos borde, son máximos o mínimos. Este proceso se realiza comparando cada punto borde, por ejemplo el punto rojo en la Fig. 23, con sus ocho vecinos del mismo nivel (puntos verdes en Fig. 23) y con sus dieciocho vecinos en las escalas adyacentes (puntos azules en Fig. 23). Si el valor de intensidad del píxel seleccionado es el máximo o mínimo del conjunto de veintisiete píxeles, entonces se dice que es punto clave. La Fig. 24 muestra los puntos clave obtenidos con las tres pirámides wavelets.

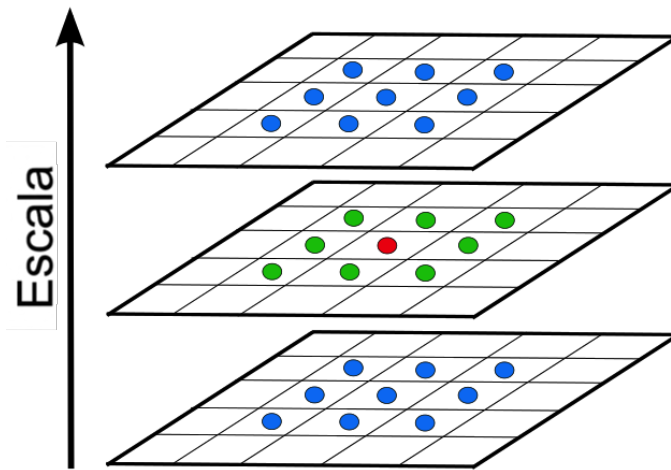


Figura 23: Boceto para analizar si un píxel es máximo o mínimo.



Figura 24: (a) Puntos clave detectados con P_{AA} . (b) Puntos clave detectados con P_{BA} . (c) Puntos clave detectados con P_{AB} .

IV.1.3 Asignación del descriptor a cada punto clave

Hasta ahora, se tiene una metodología que es invariante a escala (mediante el uso de pirámides wavelet) y a traslación (utilizando puntos clave). Ahora, para obtener la invariancia a rotación, se generan descriptores utilizando máscaras binarias de circunferencia de diferentes radios. El primer paso para crear el descriptor es tomar una vecindad de 81×81 píxeles alrededor del punto clave. Luego, se utiliza un filtro binario

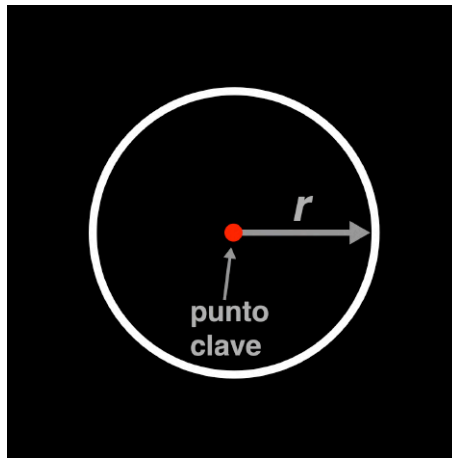


Figura 25: Ejemplo de un filtro de circunferencia de radio r centrado en un punto clave (punto rojo).

de circunferencia de radio r píxeles centrado alrededor del punto clave, Fig. 25 muestra un ejemplo de un filtro de circunferencia de radio r . El siguiente paso es sumar los valores de intensidad de los píxeles que pasaron el filtro, el resultado de esta suma es la entrada r del descriptor. De esta manera, independientemente del ángulo de rotación que presente la imagen, siempre se sumarán los mismos datos, logrando la invarianza a rotación [39]. Este procedimiento se realiza para $r = 1, 2, 3, \dots, 40$, lo que da como resultado un descriptor de 40 entradas el cual es invariante a traslación, escala y rotación, en la Fig. 26 se muestra un ejemplo de un descriptor W-SIFT. Este proceso se realiza para cada uno de los puntos clave detectados.

La Fig. 27 muestra un diagrama de bloques de los pasos de la metodología W-SIFT.

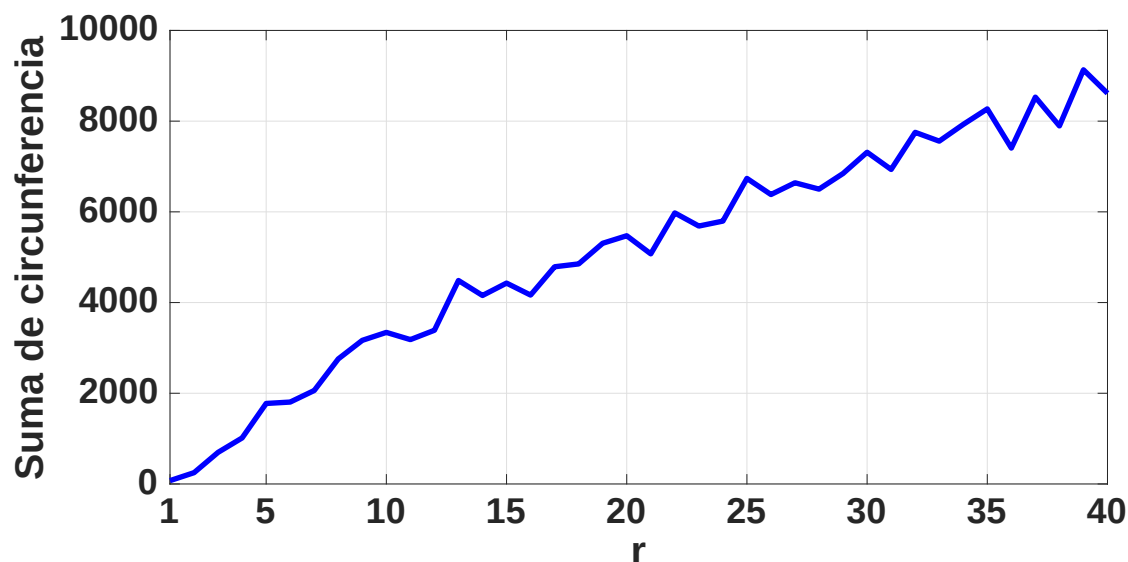


Figura 26: Ejemplo de un descriptor W-SIFT.

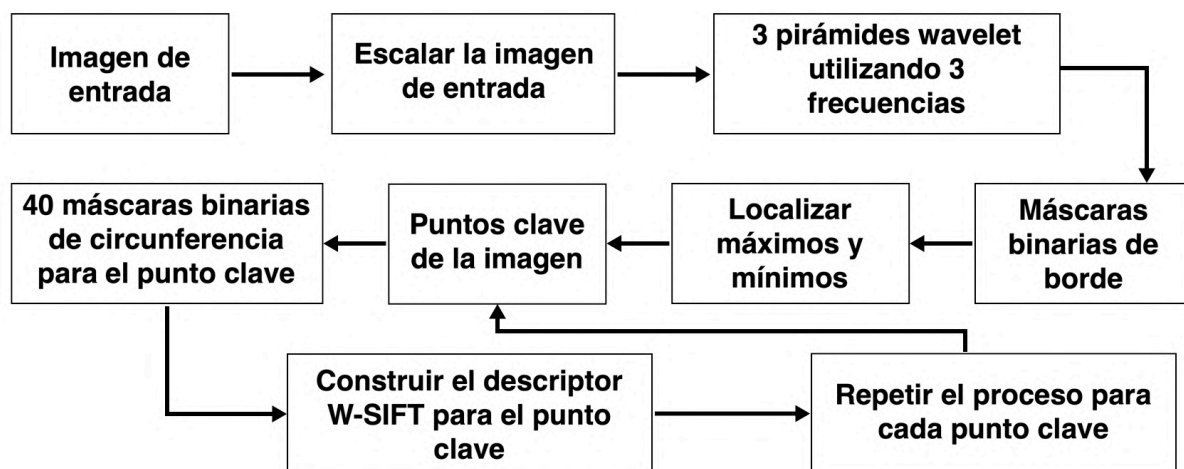


Figura 27: Diagrama de bloques para generar el descriptor W-SIFT.

IV.2 Emparejamiento de descriptores

Para poder comparar dos imágenes y saber si tienen características similares o si se trata de la misma imagen, se necesita ubicar los puntos clave en una imagen de referencia (IR) y en una imagen de prueba (IP). Luego, se generan los descriptores para todos los puntos clave de las dos imágenes. Después, se calcula la distancia Euclidiana de un descriptor de la IP con todos los descriptores de la IR. Entonces, decimos que un par de descriptores serán emparejados si se cumplen los siguientes criterios:

1. El par tiene la distancia mínima, de todas las distancias calculadas.
2. La distancia mínima satisface ser menor que el umbral de 0.025.

Se realiza el mismo procedimiento para el resto de los descriptores de la imagen de prueba para obtener todos los emparejamientos entre IR e IP. La Fig. 28 muestra un ejemplo de los emparejamientos entre una imagen (IR) y una versión reducida 10% de la misma (IP).



Figura 28: Ejemplo de los emparejamientos entre dos imágenes.

Capítulo V

Base de datos de imágenes digitales

Para probar la robustez de la metodología W-SIFT al reconocer objetos que presentan rotación o escalamiento, se utilizaron como imágenes de referencia una base de datos de 35 imágenes digitales en escala de grises de objetos aleatorios (Fig. 29), las cuales son parte de la biblioteca de imágenes digitales de Amsterdam [40]. Estas imágenes fueron seleccionadas debido a la diferencia que tienen en sus formas geométricas, bordes y la cantidad de detalle. Además, estas imágenes presentan una iluminación no homogénea y variaciones en los niveles de intensidad.

Para generar la base de datos de imágenes de prueba, cada objeto se rotó de 0° a 360° con $\Delta\theta = 10^\circ$. Además, las imágenes se escalaron $\pm 50\%$ con un tamaño de paso de 1% , lo que da un total de 4795 imágenes de prueba.



Figura 29: Base de datos de imágenes de referencia.

Capítulo VI

Pruebas y análisis de resultados

VI.1 Análisis de tiempo de cómputo

Para analizar el tiempo de cómputo de W-SIFT, SIFT y SURF, se realizó una prueba para calcular el tiempo de ejecución que emplea cada metodología al detectar los puntos clave y construir para cada uno su descriptor. La prueba consiste en tomar las imágenes escaladas de la base de datos de imágenes de prueba y calcular la media del tiempo de ejecución de 30 repeticiones por imagen. Para realizar las pruebas a la metodología SIFT se utilizaron los programas compilados provistos en el sitio web de David Lowe los cuales están elaborados en Matlab y C [41], para SURF se utilizaron las herramientas de Sistema de Visión por Computadora de Matlab y la metodología W-SIFT se implementó en Matlab. Todas las pruebas se realizaron en una computadora DELL G5 de 64 bits con Windows 10, Intel Core i7-8750H, 2.21GHz, 8GB RAM.

La Fig. 30 muestra el resultado del análisis del tiempo de cómputo, donde se observa que el mayor tiempo requerido por SURF está por debajo de 0.02 segundos, 0.89 segundos para SIFT y 0.36 segundos para W-SIFT. Aunque el mejor tiempo de cómputo se obtuvo con SURF, lo cual era de esperarse ya que es la versión rápida de SIFT, en las siguientes secciones se verá que el rendimiento de esta metodología es deficiente. Con los datos de la Fig. 30 se deduce que el tiempo de ejecución de W-SIFT es 2.4 veces más rápido que el tiempo de ejecución de SIFT, para todas las imágenes de prueba, y para algunas imágenes SIFT es hasta 5 veces más lento que W-SIFT. Por lo tanto, se

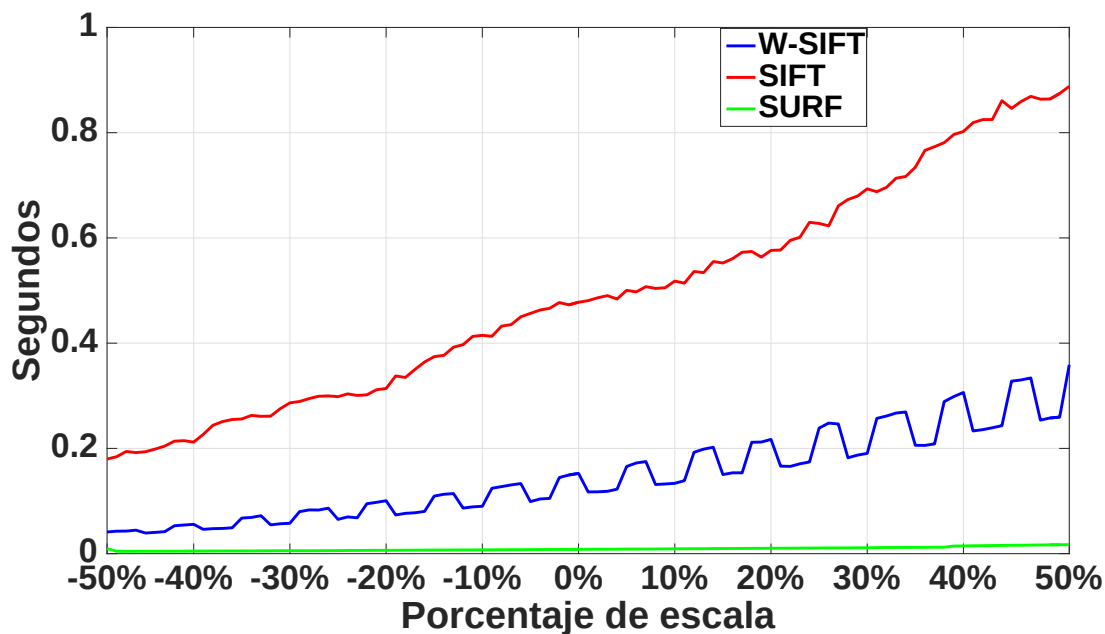


Figura 30: Análisis del tiempo de cómputo para W-SIFT, SIFT y SURF.

logró el objetivo de obtener una metodología más rápida que SIFT.

VI.2 Análisis de rendimiento

Las pruebas de rendimiento se realizaron haciendo la comparación entre una imagen de prueba (IP) (la cual es una imagen rotada o escalada) y una imagen de referencia (IR). Dado que se conoce la rotación o la escala que tiene cada imagen de prueba, se puede calcular cuántos emparejamientos son verdaderos positivos (VP), falsos positivos (FP), verdaderos negativos (VN) o falsos negativos (FN). Utilizando estos valores se determinan las siguientes métricas:

- **Precisión:** Proporción de emparejamientos positivos correctamente identificados como positivos.

$$Precisión = \frac{VP}{VP + FP}, \quad (28)$$

- **Especificidad:** Proporción de emparejamientos identificados como negativos del total de emparejamientos realmente negativos.

$$Especificidad = \frac{VN}{VN + FP}, \quad (29)$$

- **Sensitividad:** Proporción de emparejamientos identificados como positivos del total de emparejamientos realmente positivos.

$$Sensitividad = \frac{VP}{VP + FN}, \quad (30)$$

- **Exactitud:** Proporción de emparejamientos correctamente identificados del total de emparejamientos realizados.

$$Exactitud = \frac{VP + VN}{VP + VN + FP + FN}. \quad (31)$$

VI.2.1 Pruebas de rotación y escala

El primer paso de la prueba es calcular cuántos emparejamientos son VP, FP, VN, o FN entre una imagen de prueba y su correspondiente imagen de referencia. Por ejemplo, la Tabla I muestra los resultados, utilizando W-SIFT, entre la imagen I_2 y sus 36 versiones rotadas. El último renglón de la Tabla I muestra la media de todos los valores calculados de VP, FP, VN y FN. Estos valores promedio serán utilizados en las ecs. (28), (29), (30) y (31). Por lo tanto, para la imagen I_2 obtenemos los siguientes resultados: $Precisión = 0.95586$, $Sensitividad = 0.87759$, $Especificidad = 0.99235$ y $Exactitud = 0.97413$. La Figura 31 presenta las gráficas de los resultados de precisión, sensibilidad, especificidad y exactitud de las pruebas de rotación para todas las

imágenes de prueba de la base de datos para las metodologías W-SIFT, SIFT y SURF.

En la Fig. 32 se muestra los resultados análogos para las pruebas de escala.

Tabla I: Resultados de rotación para I_2 , utilizando W-SIFT.

Grados	VP	FP	FN	VN	Grados	VP	FP	FN	VN
0°	572	0	0	0	180°	52	1	9	404
10°	121	4	11	318	190°	39	2	5	433
20°	87	2	12	377	200°	39	2	5	420
30°	57	1	11	368	210°	15	4	2	432
40°	42	4	15	376	220°	17	2	1	435
50°	43	4	16	363	230°	8	3	0	431
60°	40	4	14	364	240°	20	2	0	434
70°	52	6	9	348	250°	28	3	11	408
80°	78	6	8	405	260°	50	2	9	411
90°	106	5	7	385	270°	66	1	5	444
100°	71	4	17	371	280°	36	4	10	440
110°	42	5	6	405	290°	42	3	10	419
120°	32	4	6	385	300°	31	5	11	384
130°	41	1	10	387	310°	47	4	8	400
140°	27	2	6	411	320°	46	4	10	420
150°	32	1	10	393	330°	67	2	23	352
160°	32	3	9	428	340°	64	3	15	362
170°	35	3	7	416	350°	138	1	15	344
MEDIAS:		VP = 64.31		FP = 2.97		FN = 8.97		VN = 385.36	

En las Figs. 31 y 32 se observan resultados comparables entre W-SIFT y SIFT, teniendo valores similares de sensibilidad y exactitud para la rotación, y valores similares de precisión para la escala. En el resto de las métricas W-SIFT muestra valores un poco más bajos que SIFT. W-SIFT y SIFT tienen valores altos de sensibilidad, lo que significa que la mayoría de los emparejamientos que se están haciendo son correctos. Por lo tanto, W-SIFT con una sensibilidad equivalente a la de SIFT implica que se identificó un porcentaje equivalente de emparejamientos positivos del total de emparejamientos realmente positivos. Por otro lado, para las pruebas de rotación SURF muestra resultados por debajo de W-SIFT para todas las métricas. Para las pruebas de escala SURF tiene valores de precisión muy por debajo de W-SIFT, para la especificidad y exactitud muestra resultados similares a W-SIFT y solo en la sensibilidad presenta un valor superior a W-SIFT. Por lo tanto, el desempeño de W-SIFT con imágenes que presentan rotación o escala es comparable con SIFT y mejor que SURF.

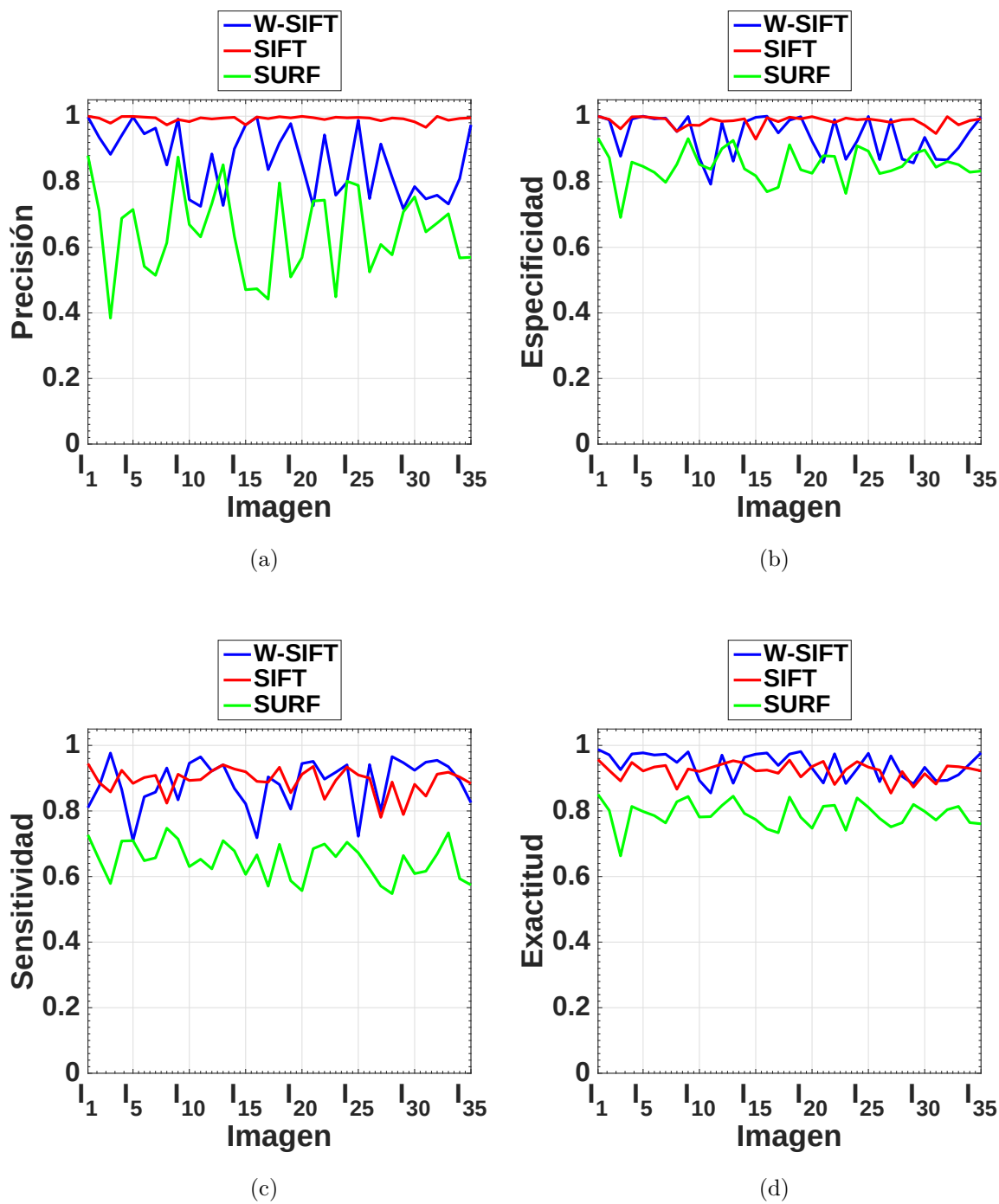


Figura 31: Pruebas de rendimiento para imágenes rotadas. (a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.

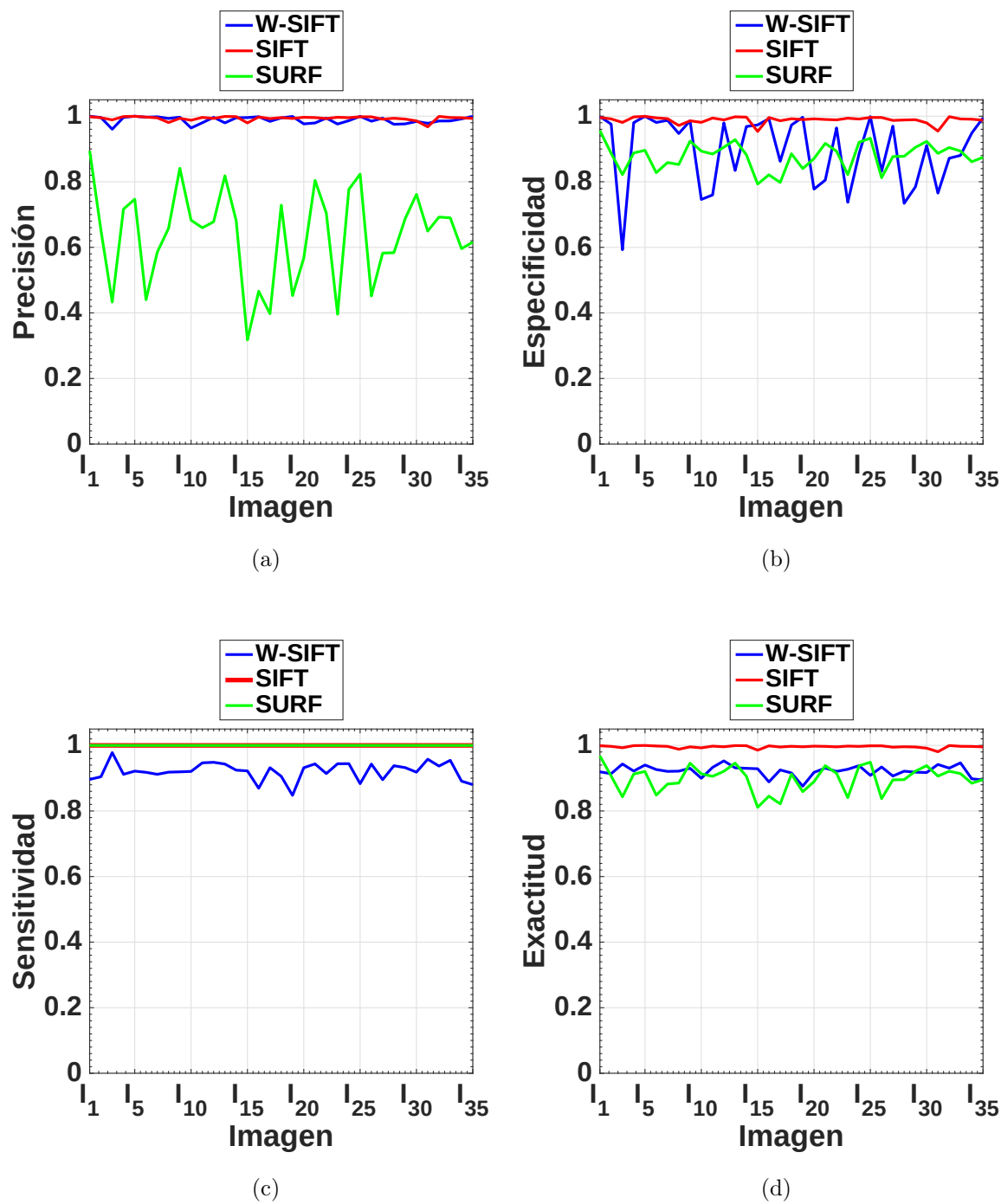


Figura 32: Pruebas de rendimiento para imágenes escaladas. (a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.

VI.2.2 Detección de puntos clave en imágenes con textura sin bordes definidos

Como se ha mencionado anteriormente, el número de puntos clave que se emparejan utilizando SIFT y SURF disminuye cuando se trabaja con imágenes con textura sin bordes definidos. Para probar esto, se realizó una prueba utilizando cinco imágenes de lesiones de piel, Fig. 33, las cuales son parte de la biblioteca de imágenes Dermofit [1]. Estas imágenes fueron diagnosticadas por expertos (dermatólogos y dermatopatólogos). Además, para cada imagen, se proporcionó una máscara binaria que cubre solamente el área donde se encuentra la lesión, Fig. 34.

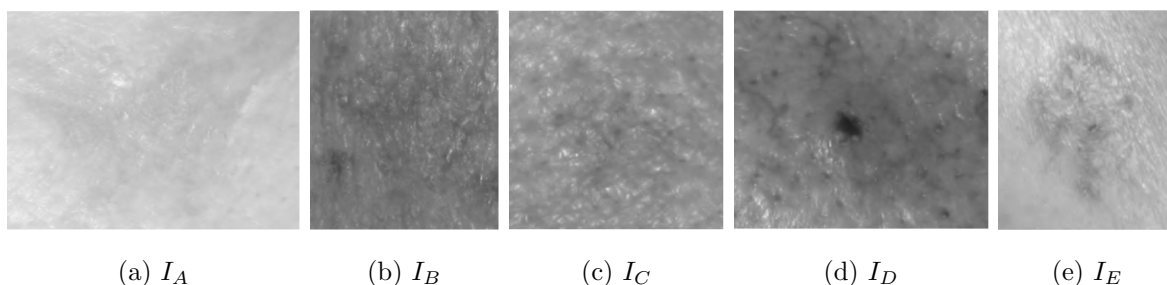


Figura 33: Ejemplos de lesiones de piel (imágenes tomadas de la librería de imágenes Dermofit [1]. (a), (b) y (c): Queratosis actínica. (d) y (e): Carcinoma de células basales.

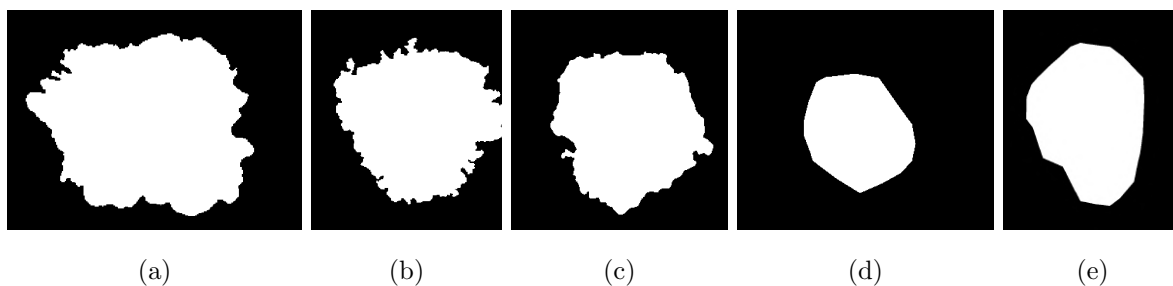


Figura 34: Máscaras binarias que cubren solo la lesión de las imágenes de la Fig. 33.

La Fig. 35 muestra un ejemplo de la respuesta de las metodologías W-SIFT, SIFT y SURF para la imagen Fig. 33a. Las Figuras 35b muestran los puntos clave emparejados entre la zona que solo muestra la lesión de piel y la imagen problema. La metodología W-SIFT emparejó 434 puntos clave, cubriendo una porción significativa de la lesión, mientras que SIFT detectó solo 6 puntos y SURF ninguno.

La Fig. 36 presenta las gráficas de los resultados para W-SIFT y SIFT; en estas gráficas no se presenta SURF ya que, como en algunas imágenes no hay emparejamientos, no se puede calcular la precisión y la sensibilidad. Los valores resultantes de las pruebas para las imágenes con textura sin bordes definidos de la Fig. 33, para las tres metodologías, se muestran en la Tabla II. Las tres metodologías tienen una especificidad de 1, W-SIFT y SIFT tienen un valor de 1 en precisión, pero al utilizar SIFT y SURF se detectan y emparejan muy pocos puntos, lo cual dificulta analizar y reconocer si la imagen presenta una lesión de piel o no. Como se observa en la Tabla II, utilizando SIFT, el número máximo de puntos clave emparejados fueron 64 para la Fig. 33b y el mínimo fue de solo 6 puntos para la Fig. 33a. Con SURF solo se emparejaron 3 puntos para la Fig. 33b, 10 para Fig. 33e y 0 en el resto de las imágenes. En cambio, usando W-SIFT, el número más bajo de puntos clave emparejados fue de 90 para la imagen Fig. 33d y el máximo fue 1177 para Fig. 33e. Las imágenes de la Fig. 33 son lesiones reales de piel, las cuales contienen información médica que pueden ser de ayuda para detectar cáncer de piel. Por lo tanto, es necesario contar con una metodología que detecte y empareje un número considerable de puntos clave que cubran la porción dañada y ayude a diagnosticar si la piel tiene una lesión o no, y W-SIFT muestra ser una opción más adecuada que SIFT y SURF para lograr este propósito.

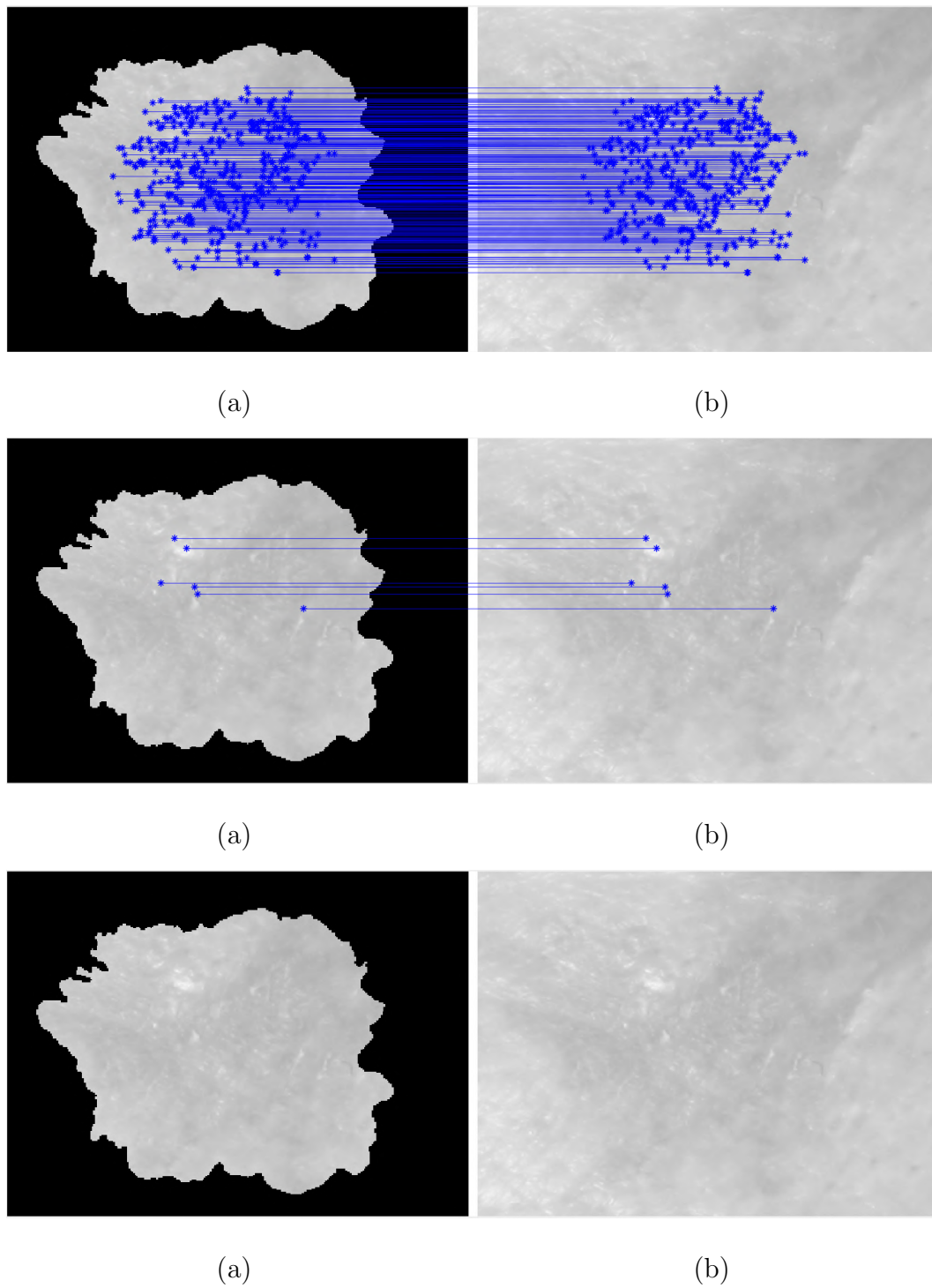


Figura 35: Puntos clave emparejados para la imagen I_A , Fig. 33a. (a) Zona de lesión de piel. (b) Imagen de la lesión de piel. Arriba: W-SIFT. En medio: SIFT. Abajo: SURF.

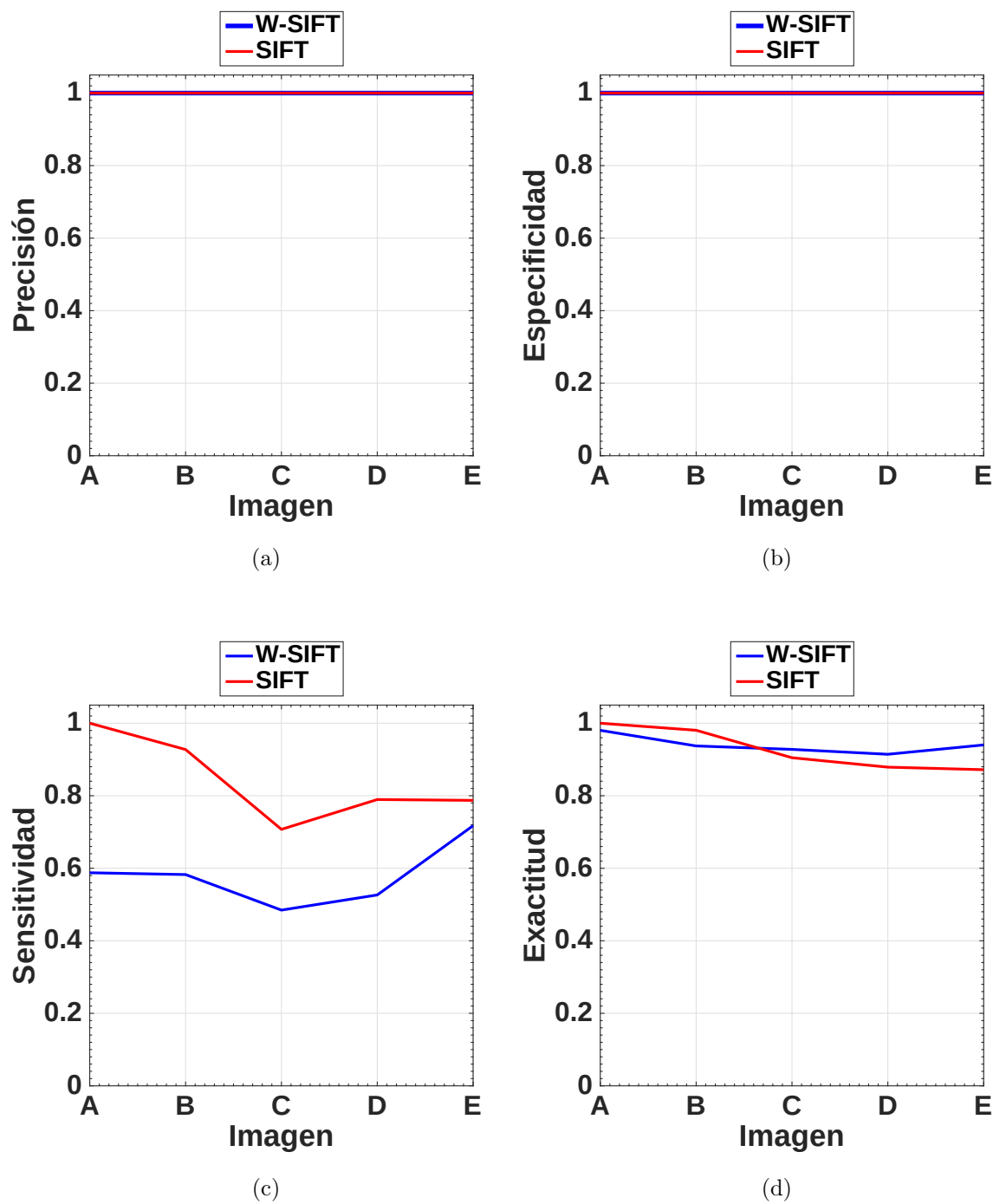


Figura 36: Pruebas de rendimiento para imágenes con textura sin bordes definidos.

(a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.

Tabla II: Resultados de rendimiento para imágenes con textura sin bordes definidos.

W-SIFT					
Imagen	Emparejamientos	Sensitividad	Especificidad	Precisión	Exactitud
I_A	434	0.5876	1	1	0.9803
I_B	351	0.5826	1	1	0.9373
I_C	116	0.4846	1	1	0.9278
I_D	90	0.5264	1	1	0.9143
I_E	1177	0.7174	1	1	0.9401
SIFT					
Imagen	Emparejamientos	Sensitividad	Especificidad	Precisión	Exactitud
I_A	6	1	1	1	1
I_B	64	0.9274	1	1	0.9805
I_C	29	0.7073	1	1	0.9048
I_D	15	0.7895	1	1	0.8788
I_E	37	0.7872	1	1	0.8718
SURF					
Imagen	Emparejamientos	Sensitividad	Especificidad	Precisión	Exactitud
I_A	0	-	1	-	1
I_B	3	0.7500	1	1	0.9944
I_C	0	-	1	-	1
I_D	0	-	1	-	0.9545
I_E	10	0.8333	1	1	0.9860

VI.3 Detección de puntos clave en imágenes con iluminación no homogénea

Para probar el rendimiento de las metodologías W-SIFT, SIFT y SURF en imágenes con iluminación no homogénea, utilizamos siete imágenes con diferentes ángulos de iluminación para cada imagen de referencia de la base de datos, Fig. 29. Estas imágenes también son parte de la biblioteca de imágenes digitales de Amsterdam [40]. La Fig. 37 muestra un ejemplo de los siete ángulos de iluminación para I_2 , e I_3 .

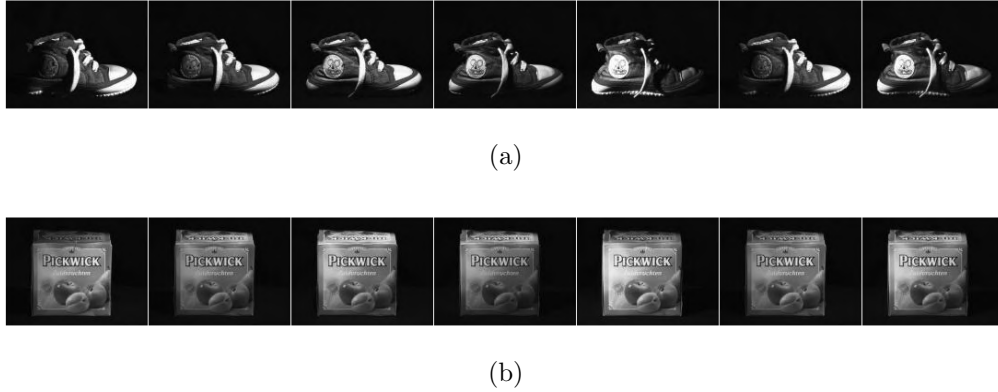


Figura 37: Ejemplo de imágenes con iluminación no homogénea. (a) I_2 . (b) I_3 .

Los resultados de las pruebas de rendimiento se muestran en la Fig. 38. En estas figuras se observa que la sensibilidad de W-SIFT es mayor que la de SIFT y SURF, mientras que para el resto de las métricas los valores de W-SIFT están por debajo, esto sucede ya que la metodología W-SIFT no está diseñada para trabajar con este tipo de imágenes. Para mejorar la robustez de W-SIFT ante imágenes con iluminación no homogénea, es necesario analizar los descriptores de los puntos clave donde los resultados fueron exitosos, como en las imágenes I_{16} , I_{25} e I_{30} , y determinar qué características deben extraerse para aumentar los valores generales de precisión y exactitud para imágenes con iluminación no homogénea.

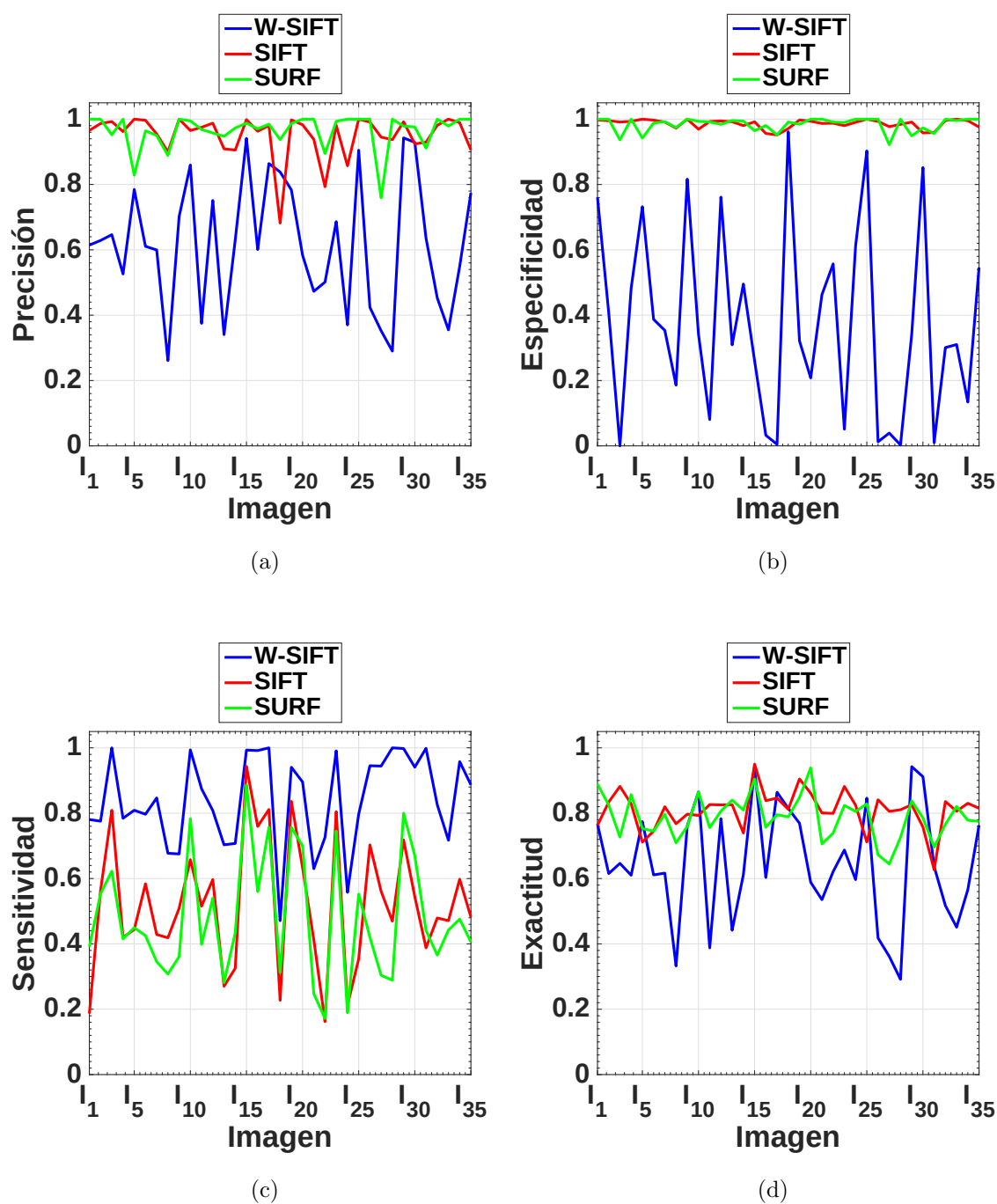


Figura 38: Pruebas de rendimiento para imágenes con iluminación no homogénea.

(a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.

VI.4 Detección de puntos clave en imágenes con ruido moteado

Otra prueba que se realizó a las metodologías fue calcular los resultados ante imágenes con ruido moteado, para esto, las imágenes de la Fig. 29 fueron alteradas con ruido moteado con media cero y varianza desde cero hasta 0.2, con un tamaño de paso de 0.01, utilizando 30 imágenes por muestra. La Figura 39 muestra un ejemplo de imágenes con ruido moteado para I_{12} . En la Fig. 40 observamos que la especificidad y precisión de W-SIFT y SIFT es similar y superior a SURF. La exactitud de W-SIFT es mayor que la de SIFT y SURF, para todas las imágenes, y la sensibilidad de las tres metodologías está por debajo de 0.5. Aunque la metodología W-SIFT no está diseñada para trabajar con imágenes con ruido moteado, los resultados obtenidos son similares a los de SIFT y mucho mayor que los de SURF.

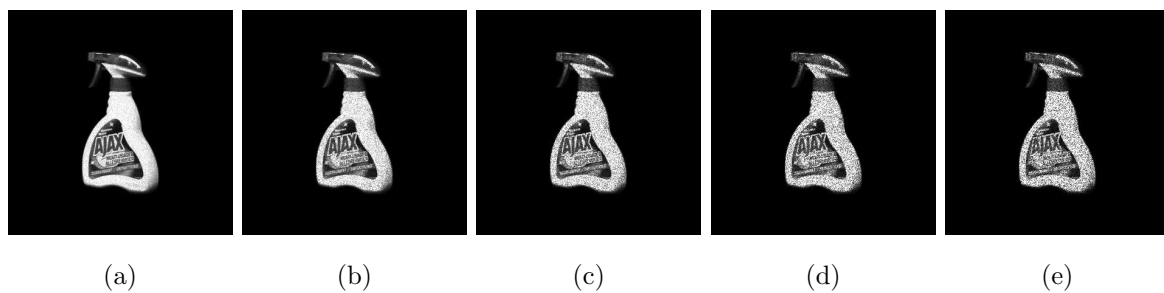


Figura 39: Ejemplo de imágenes con ruido moteado para I_{12} . (a) Varianza = 0.01. (b) Varianza = 0.05. (c) Varianza = 0.1. (d) Varianza = 0.15. (e) Varianza = 0.2.

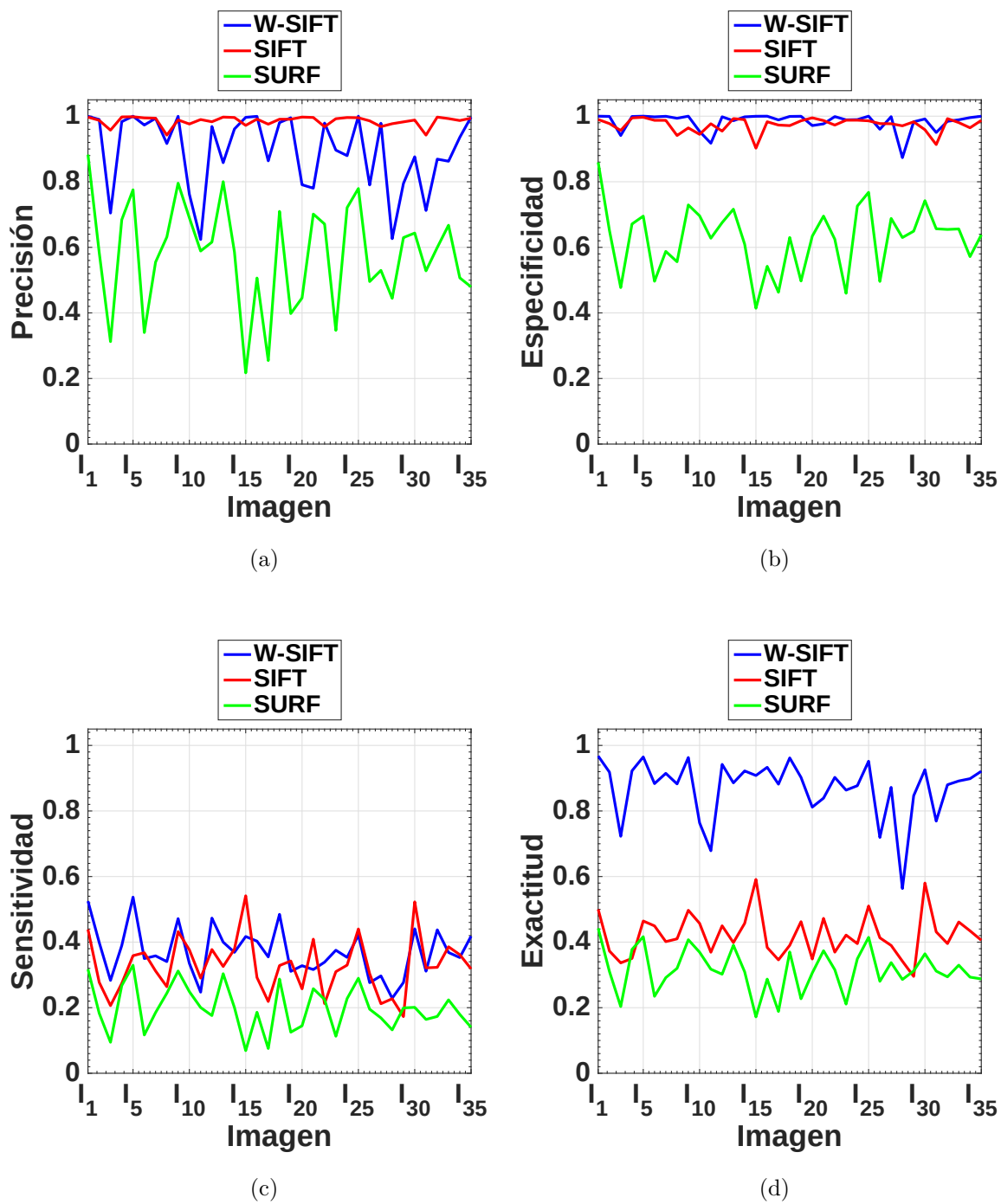


Figura 40: Pruebas de rendimiento para imágenes con ruido moteado. (a) Precisión. (b) Especificidad. (c) Sensitividad. (d) Exactitud.

VI.5 Rendimiento del método de emparejamiento

Para verificar el rendimiento del método de emparejamiento propuesto se calculó la matriz homógrafa, $\mathbf{H}$, utilizando los puntos clave que hicieron emparejamiento entre una IP y la IR. La matriz homógrafa debe satisfacer que

$$p = \mathbf{H}q, \quad (32)$$

donde $p = (x_1, y_1) \in \text{IR}$ y $q = (x_2, y_2) \in \text{IP}$, son un par de puntos clave que hicieron emparejamiento [42]. Después, se utilizó esta matriz para regresar la imagen distorsionada IP a la IR [43]. La Fig. 41 presenta un ejemplo de este procedimiento. En las Figs. 41a y 41b se muestra la imagen de referencia y la imagen de prueba, respectivamente, siendo la IP una versión rotada 30° de la IR. La imagen en la Fig. 41c es la regresión de la IP a la IR utilizando la matriz homógrafa.

Para medir los resultados, se calculó la distancia Euclidiana entre dos puntos clave. Primero, se toma un par de puntos clave que sean un emparejamiento, por ejemplo, $p = (x_1, y_1) \in \text{IR}$ y $q = (x_2, y_2) \in \text{IP}$. Después, se multiplica q por la matriz homógrafa

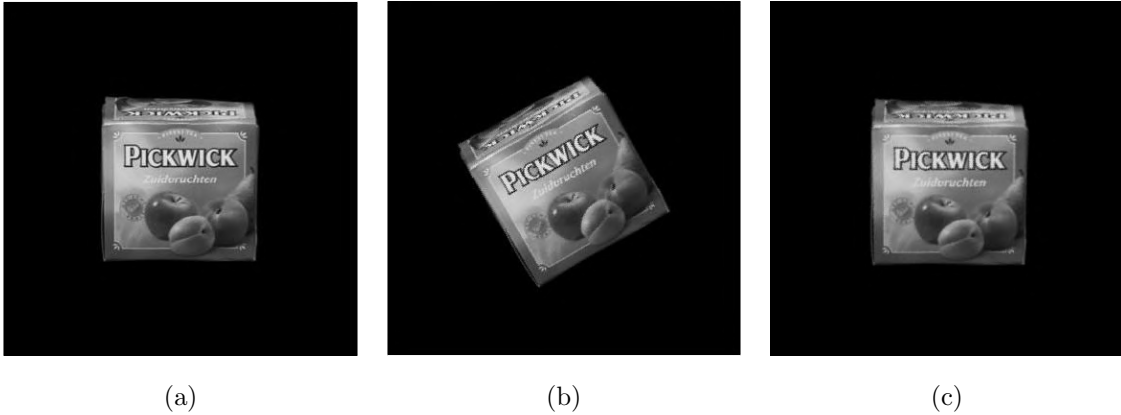


Figura 41: (a) Imagen de referencia. (b) Imagen de prueba, versión rotada 30° de la IR. (c) Regresión de la IP a la IR.

$\mathbf{H}$, esto es $w = \mathbf{H}q$, y se calcula la distancia Euclidiana entre los puntos p y w , si el emparejamiento es correcto la distancia debe ser cercana a cero [43]. Finalmente, se obtiene la distancia entre todos los puntos clave que hicieron emparejamiento entre la IP y la IR. Este proceso se realizó para todas las imágenes de referencia de la base de datos con sus correspondientes imágenes de prueba escaladas. La Tabla III muestra los resultados de este análisis. El valor de la distancia Euclidiana que se reporta en la tabla representa el promedio de las 101 distancias calculadas entre las imágenes escaladas con su correspondiente IR. La Tabla IV presenta el resultado análogo para las imágenes de rotación, donde las 36 imágenes de prueba rotadas se utilizan para calcular la distancia con cada IR. Como se observa en las Tablas III y IV, las distancias son menores a un píxel para 34 de las 35 imágenes para la prueba de escala y para 28 de las 35 imágenes

Tabla III: Distancia Euclidiana entre IR y sus correspondientes IP escaladas.

Imagen de referencia	Distancia Euclidiana	Imagen de referencia	Distancia Euclidiana	Imagen de referencia	Distancia Euclidiana
I₁	0.290	I₁₃	0.320	I₂₅	0.425
I₂	0.154	I₁₄	0.208	I₂₆	0.195
I₃	0.178	I₁₅	0.388	I₂₇	0.094
I₄	0.245	I₁₆	0.102	I₂₈	0.223
I₅	0.227	I₁₇	0.189	I₂₉	0.197
I₆	0.241	I₁₈	1.080	I₃₀	0.139
I₇	0.222	I₁₉	0.122	I₃₁	0.129
I₈	0.193	I₂₀	0.212	I₃₂	0.243
I₉	0.540	I₂₁	0.204	I₃₃	0.284
I₁₀	0.192	I₂₂	0.362	I₃₄	0.127
I₁₁	0.137	I₂₃	0.235	I₃₅	0.198
I₁₂	0.291	I₂₄	0.204		

para la prueba de rotación, mientras que los valores restantes se encuentran entre uno y tres píxeles para las dos pruebas. Estos resultados significan que el método de emparejamiento propuesto está dando excelentes resultados.

Tabla IV: Distancia Euclidiana entre IR y sus correspondientes IP rotadas.

Imagen de referencia	Distancia Euclidiana	Imagen de referencia	Distancia Euclidiana	Imagen de referencia	Distancia Euclidiana
I₁	1.327	I₁₃	0.076	I₂₅	2.609
I₂	0.184	I₁₄	0.119	I₂₆	0.080
I₃	0.050	I₁₅	2.559	I₂₇	0.414
I₄	0.164	I₁₆	2.284	I₂₈	0.057
I₅	2.807	I₁₇	0.169	I₂₉	0.090
I₆	0.272	I₁₈	1.961	I₃₀	0.323
I₇	0.288	I₁₉	0.694	I₃₁	0.079
I₈	0.134	I₂₀	0.066	I₃₂	0.096
I₉	2.316	I₂₁	0.111	I₃₃	0.264
I₁₀	0.095	I₂₂	0.194	I₃₄	0.128
I₁₁	0.052	I₂₃	0.121	I₃₅	0.604
I₁₂	0.613	I₂₄	0.114		

Capítulo VII

Clasificación de imágenes

El primer paso para la clasificación de imágenes es obtener el umbral de emparejamientos para decidir si dos imágenes son la misma o no. Para calcular este umbral se generó la curva ROC (*Receiver Operating Characteristic*) variando el número de emparejamientos desde 0 hasta 3000. Para cada una de las 35 IR de la Fig. 29 se realizó la comparación con las 4795 IP utilizando un umbral de x emparejamientos, dando un total de 167825 comparaciones, después, se obtuvieron los valores VP, FP, VN y FN y se calcularon el FPR (False Positive Rate) y el TPR (True Positive Rate). Este proceso se repite variando el umbral x desde 0 hasta 3000 con un tamaño de paso de 1. La Fig. 42 muestra la curva ROC cuya área bajo la curva es de 0.9179 el cual es un excelente resultado [44]. Además, en la Fig. 42 se presenta, con un punto rojo, el punto de inflexión de la curva ROC, el cual se obtiene con un umbral de 78 emparejamientos. En la Fig. 43 se muestran las gráficas de la precisión, sensibilidad, especificidad, exactitud y medida F, obtenidas para cada umbral x desde 0 hasta 3000 con un tamaño de paso de 1. Además, en líneas verticales se resaltan dos umbrales: 78 y 302, donde 78 es, como se mencionó anteriormente, el punto de inflexión de la curva ROC, y 302 es la intersección de las gráficas de la precisión, sensibilidad y medida F.

Para realizar la clasificación de imágenes, se eligió utilizar el umbral de 302 emparejamientos. La clasificación se realiza tomando una de las imágenes de referencia y haciendo el emparejamiento con cada una de las 4795 imágenes de prueba, si dos imágenes tienen 302 o mas emparejamientos se clasificarán como la misma imagen.

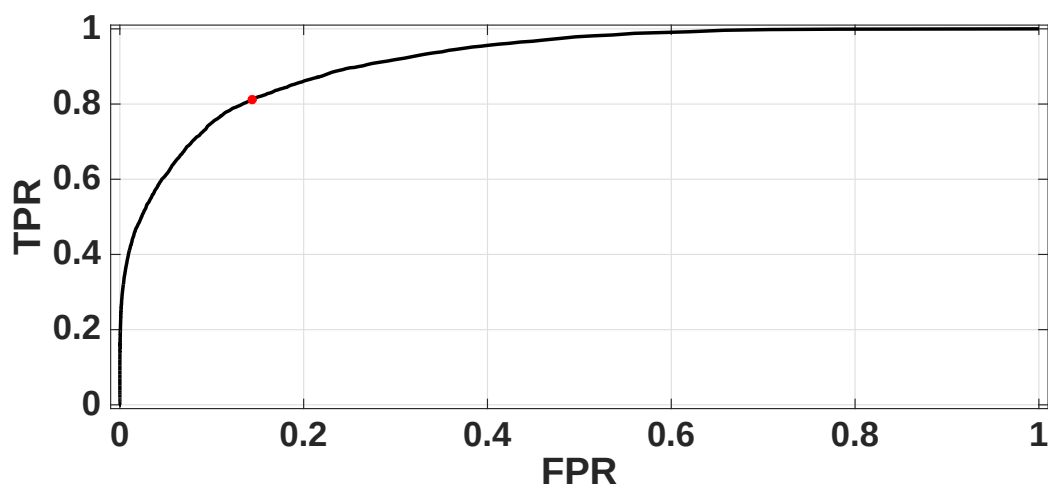


Figura 42: Curva ROC para la clasificación de imágenes. FPR (False Positive Rate), TPR (True Positive Rate), se muestra en asterisco rojo el punto de inflexión de la curva.

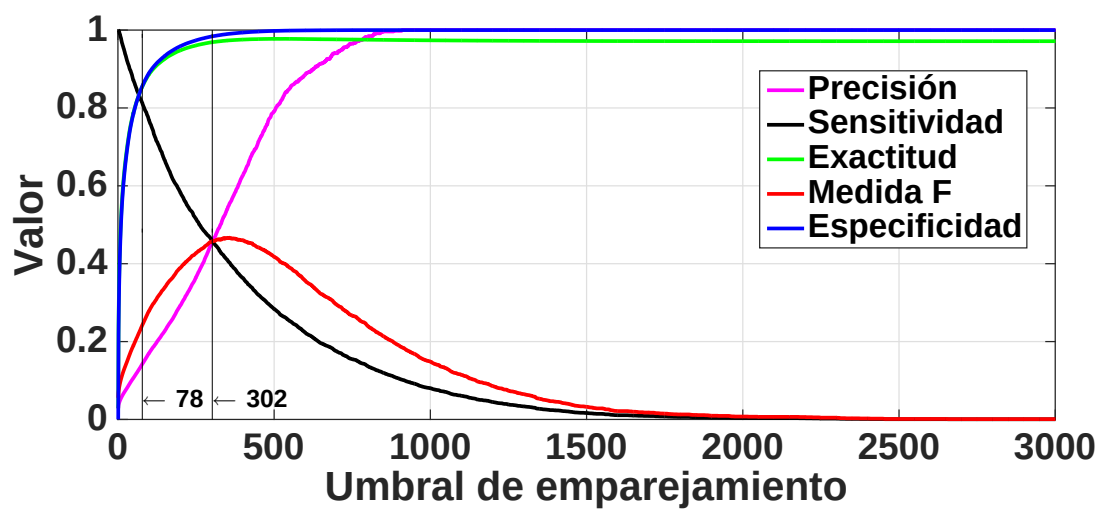


Figura 43: Pruebas de rendimiento para la clasificación de imágenes.

El rendimiento para la clasificación de imágenes se presenta en la Fig. 44, mostrando buenos resultados con valores promedio de 0.7 de precisión, 0.98 de especificidad, 0.97 de exactitud y 0.51 de medida F.

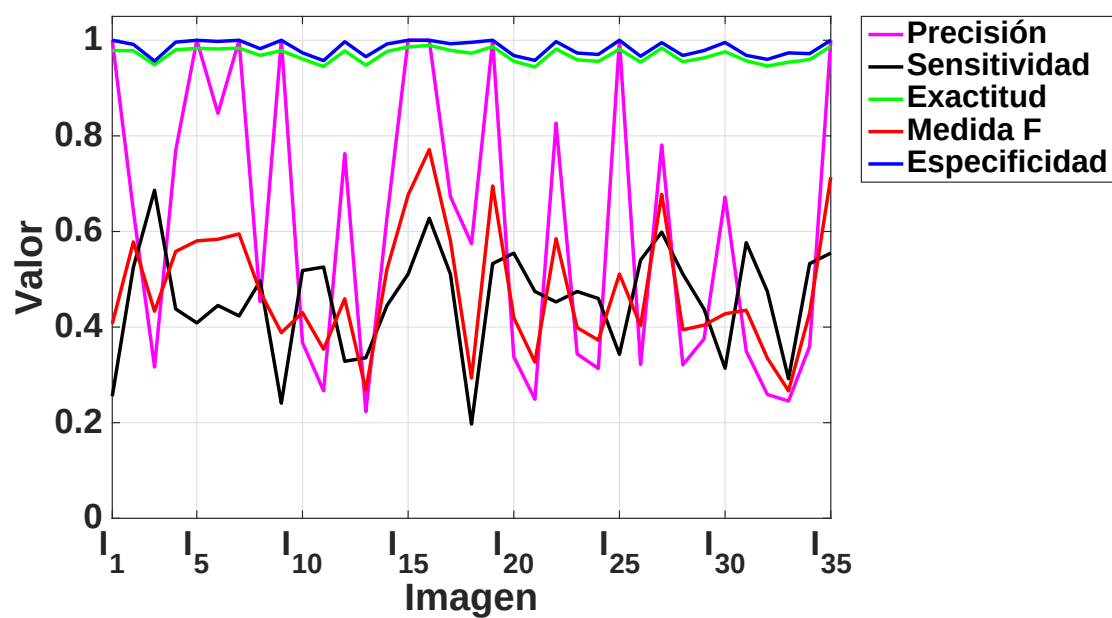


Figura 44: Pruebas de rendimiento para la clasificación de imágenes con un umbral de 302 emparejamientos.

Capítulo VIII

CW-SIFT

La última parte de este trabajo de tesis es generalizar la metodología W-SIFT para la detección de objetos a color. Esta generalización se realizó implementando W-SIFT en el espacio de color RGB, a esta metodología se le llamará Color Wavelet SIFT (CW-SIFT). Para analizar el rendimiento de CW-SIFT se utilizaron las imágenes de la Fig. 33 a color, las cuales se muestran en la Fig. 45.

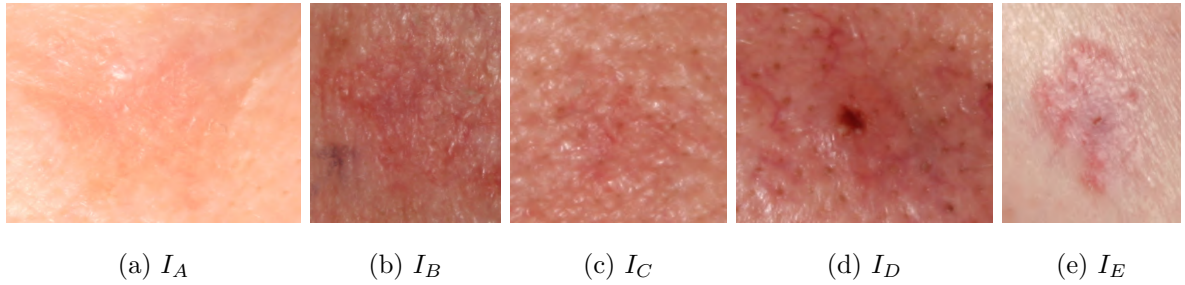


Figura 45: Ejemplos de lesiones de piel a color (imágenes tomadas de la librería de imágenes Dermofit [1]). (a), (b) y (c): Queratosis actínica. (d) y (e): Carcinoma de células basales.

El primer paso de la metodología CW-SIFT es descomponer la imagen de referencia en sus tres canales RGB. Por ejemplo, la Fig. 46 muestra los tres canales rojo, verde y azul para la imagen I_B (Fig. 45b). A estas imágenes les llamaremos I_B^r para el canal rojo, I_B^g para el canal verde e I_B^b para el canal azul.

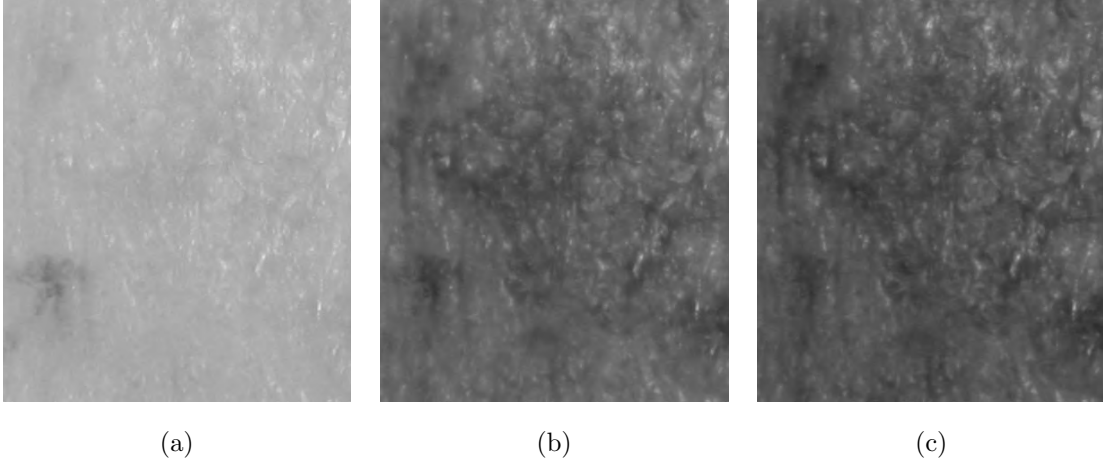


Figura 46: Descomposición RGB para I_B (Fig. 45b). (a) Canal rojo I_B^r . (b) Canal verde I_B^g . (c) Canal azul I_B^b .

El siguiente paso de la metodología CW-SIFT es aplicar a cada uno de los canales la metodología W-SIFT para detectar los puntos clave y asignarles un descriptor, esto es posible ya que cada una de las imágenes I_B^r , I_B^g e I_B^b es una imagen en escala de grises. La Figura 47 muestra los puntos clave detectados en cada uno de los canales utilizando W-SIFT.

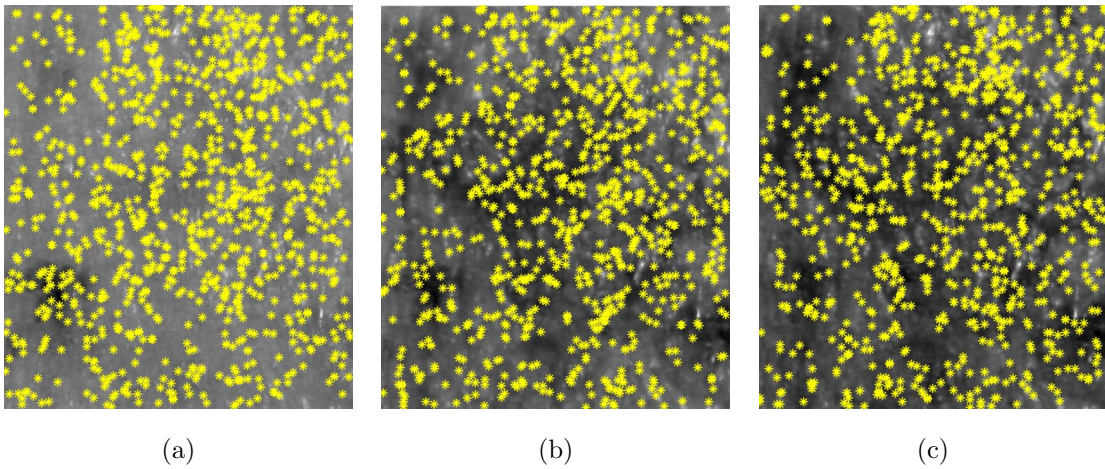


Figura 47: Detección de puntos clave. (a) Canal rojo I_B^r . (b) Canal verde I_B^g . (c) Canal azul I_B^b .

Una vez obtenidos los puntos clave se calcula el descriptor de cada uno de ellos. Después se realiza el emparejamiento con los descriptores de los puntos clave de la imagen que solo contiene la lesión de piel. Este proceso se realiza tres veces, uno para cada canal, utilizando la metodología de emparejamiento de descriptores descrita en la sección IV.2. En las Fig. 48a, Fig. 48b y Fig. 48c se muestran los emparejamientos para los canales rojo, verde y azul, respectivamente. El último paso de la metodología CW-SIFT es combinar los resultados de los tres canales de color en la imagen de referencia a color, como lo muestra la Fig. 48d.

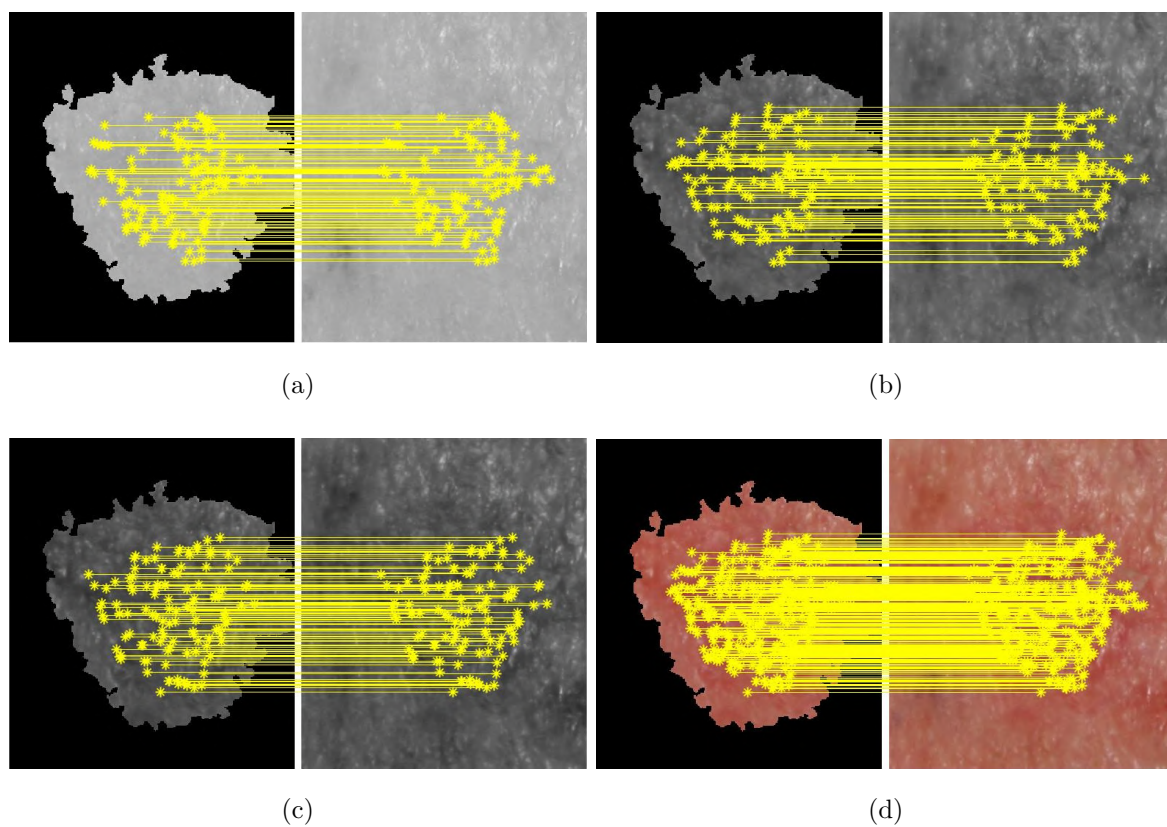


Figura 48: Emparejamiento de imágenes. (a) Canal rojo. (b) Canal verde. (c) Canal azul. (d) Combinación de los tres canales en la imagen a color.

La Tabla V contiene los resultados para el número de emparejamientos y los valores de sensibilidad, especificidad, precisión y exactitud para las imágenes a color de la Fig. 45. En esta tabla se observa que los resultados son similares a los obtenidos en la sección VI.2.2 con W-SIFT, teniendo valores de precisión y especificidad de 1, valores de sensibilidad oscilando entre 0.3 y 0.7, y valores de exactitud altos entre 0.7 y 1. Además, el número de puntos clave emparejados se mantiene elevado, teniendo en conjunto de los tres canales, 447 emparejamientos para la imagen I_A , 462 para I_B , 106 para I_C , 97 para I_D y 1669 para I_E .

Tabla V: Resultados imágenes a color.

Canal rojo					
Imagen	Emparejamientos	Sensitividad	Especificidad	Precisión	Exactitud
I_A	0	-	1	-	1
I_B	126	0.7778	1	1	0.9816
I_C	31	0.3333	1	1	0.9788
I_D	30	0.3488	1	1	0.9560
I_E	480	0.7108	1	1	0.9538
Canal verde					
Imagen	Emparejamientos	Sensitividad	Especificidad	Precisión	Exactitud
I_A	246	0.5681	1	1	0.7435
I_B	142	0.5504	1	1	0.9141
I_C	44	0.4314	1	1	0.8886
I_D	31	0.3457	1	1	0.9119
I_E	586	0.7494	1	1	0.9441
Canal azul					
Imagen	Emparejamientos	Sensitividad	Especificidad	Precisión	Exactitud
I_A	201	0.6484	1	1	0.9400
I_B	194	0.5843	1	1	0.9034
I_C	31	0.3563	1	1	0.8889
I_D	36	0.3214	1	1	0.9034
I_E	603	0.7354	1	1	0.9387

Capítulo IX

Conclusiones

La metodología de reconocimiento de patrones basada en extracción de características locales invariante a escala, rotación y traslación propuesta en este trabajo de tesis, cumple con los objetivos planteados en la misma y se le da respuesta a las preguntas de investigación.

La primera parte de este trabajo es la metodología para generar el descriptor Wavelet SIFT (W-SIFT) el cual muestra ser invariante a escala, traslación y rotación. La metodología W-SIFT tiene un excelente rendimiento, mostrando altos valores de precisión y exactitud en el emparejamiento de puntos clave entre imágenes que presentan objetos con escala, traslación y rotación. Para obtener la invarianza a escala, la metodología W-SIFT usa tres pirámides de escala construidas a partir de la transformada wavelet de Haar, por su propiedad de detectar bordes. La invarianza a rotación se introduce utilizando filtros de circunferencia y la invarianza a traslación se logra utilizando puntos clave.

La metodología W-SIFT propuesta se comparó con SIFT y SURF, ya que son las metodologías más utilizadas para la extracción de características. Los resultados mostraron que SIFT es 2.4 veces más lento que W-SIFT, y para algunas imágenes W-SIFT es hasta 5 veces más rápido. Ambas metodologías, W-SIFT y SIFT, tienen resultados altos similares de sensibilidad y exactitud, para las pruebas de rotación, y de precisión para la escala, al emparejar los descriptores de los puntos clave de las

imágenes de prueba con los descriptores de las imágenes de referencia. Por otro lado, SURF muestra muy bajos resultados, estando por debajo de W-SIFT para la mayoría de las métricas, tanto en las pruebas de rotación como en la escala. Además, el rendimiento de las tres metodologías se comparó al detectar y emparejar puntos clave en imágenes con textura sin bordes definidos, teniendo como resultado que W-SIFT y SIFT tienen valor de precisión de 1, no se compararon los resultados con SURF ya que en algunas imágenes no se detectó ningún punto clave. Para imágenes con textura sin bordes definidos, W-SIFT detecta y empareja una cantidad de puntos clave mayor que SIFT y ciertamente que SURF.

Otra prueba que se realizó para verificar la robustez de W-SIFT fue la de analizar el rendimiento del método de emparejamiento propuesto. Esta prueba consistió en calcular la distancia Euclidiana entre los puntos clave que hicieron emparejamiento de la imagen de referencia y los puntos clave de la imagen de regresión, la cual es la imagen de prueba multiplicada por la matriz homógrafa. Los resultados mostraron que el método de emparejamiento propuesto da excelentes resultados, teniendo distancias promedio menores a un píxel para la mayoría de las imágenes de prueba.

La metodología W-SIFT se utilizó para la clasificación de imágenes. Para medir el desempeño en esta tarea, se realizó un análisis utilizando la curva ROC, obteniendo un área bajo la curva de 0.9179, lo cual indica un excelente resultado para el clasificador. Después, se buscó el umbral para el cual la eficiencia de la clasificación se maximizara, escogiendo un umbral de 302 emparejamientos, el cual es la intersección de las curvas de la medida-F, la precisión y la sensibilidad. El rendimiento de la metodología para la clasificación de imágenes muestra buenos resultados con valores promedio altos de especificidad, exactitud y precisión.

La última parte de este trabajo es la metodología para generar el descriptor CW-SIFT, el cual es una generalización de W-SIFT para el reconocimiento de objetos en imágenes a color. Esta generalización se logró implementando W-SIFT en el espacio de color RGB. Se descompone la imagen en sus tres canales RGB y a cada uno de ellos se le aplica W-SIFT. Después, se realiza el emparejamiento de los descriptores con una imagen de prueba y por último se combinan los resultados de los tres canales para obtener un conjunto de emparejamientos. Los resultados mostraron que CW-SIFT empareja un número considerable de puntos clave en imágenes a color, manteniendo valores altos de precisión, especificidad, sensibilidad y exactitud.

Capítulo X

Trabajo a futuro

El descriptor W-SIFT no está diseñado para reconocer objetos en imágenes que tienen ruido o una iluminación no homogénea. Aun así, los resultados obtenidos para las pruebas de ruido moteado son similares a los de SIFT e incluso para la exactitud es mayor W-SIFT, y en todas las métricas W-SIFT es superior a SURF. Como trabajo a futuro, tenemos que mejorar la robustez de W-SIFT ante imágenes con iluminación no homogénea y ruido moteado. Además, se deben estudiar otras técnicas para mejorar la clasificación, como el agrupamiento k -means, SVM (Support-Vector Machines) o BoW (bag of words) [45, 46, 47].

Como trabajo a futuro para el descriptor CW-SIFT es necesario analizar el desempeño de esta metodología al realizar la clasificación de imágenes, para esto se requiere de una base de datos de imágenes a color más amplia y llevar a cabo una apropiada comparación con metodologías para generar descriptores en imágenes a color.

Referencias

- [1] Lucia Ballerini, Robert B. Fisher, Ben Aldridge, and Jonathan Rees. A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. In *Color Medical Image Analysis*, pages 63–86. Springer, 2013.
- [2] David G. Lowe. Object recognition from local scale-invariant features. *Computer Vision. The proceedings of the seventh IEEE international conference*, 2:1150–1157, 1999.
- [3] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [4] Ian Jolliffe. *Principal component analysis*. Springer, 2011.
- [5] Carolina Barajas-García, Selene Solorza-Calderón, and Josué Álvarez-Borrego. Classification of fragments of objects by the Fourier masks pattern recognition system. *Optics Communications*, 367:335–345, 2016.
- [6] Alfredo Solís Ventura, Josué Álvarez Borrego, and Selene Solorza. Adaptive non-linear correlation with a binary mask invariant to rotation and scale. *Optics Communications*, 339:185–193, 2015.
- [7] Esbanyely Garza-Flores and Josué Álvarez-Borrego. Pattern recognition using binary masks based on the fractional Fourier transform. *Journal of Modern optics*, 65(14):1634–1657, 2018.
- [8] Huiyu Zhou, Yuan Yuan, and Chunmei Shi. Object tracking using SIFT features and mean shift. *Computer vision and image understanding*, 113(3):345–352, 2009.

- [9] Manuele Bicego, Andrea Lagorio, Enrico Grosso, and Massimo Tistarelli. On the use of SIFT features for face authentication. In *Computer Vision and Pattern Recognition Workshop, 2006*, pages 35–35, 2006.
- [10] Hailing Huang, Weiqiang Guo, and Yu Zhang. Detection of copy-move forgery in digital images using SIFT algorithm. In *Computational Intelligence and Industrial Application, 2008*, volume 2, pages 272–276, 2008.
- [11] Fernando Alonso-Fernandez, Pedro Tome-Gonzalez, Virginia Ruiz-Albacete, and Javier Ortega-Garcia. Iris recognition based on SIFT features. In *Biometrics, Identity and Security (BIoS), 2009*, pages 1–8, 2009.
- [12] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: An efficient alternative to SIFT or SURF. In *Computer Vision (ICCV), 2011*, volume 11, pages 2564–2571, 2011.
- [13] Ignazio Infantino, Filippo Vella, Giovanni Spoto, and Salvatore Gaglio. Image Representation with Bag of bi-SIFT. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2009*, pages 287–293, 2009.
- [14] Kazim Şekeroğlu and Ömer Muhammet Soysal. Comparison of SIFT, bi-SIFT, and tri-SIFT and their frequency spectrum analysis. *Machine Vision and Applications*, 28(8):875–902, 2017.
- [15] Yan Ke and Rahul Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. *Computer Vision and Pattern Recognition (CVPR), 2004.*, 2:506–513, 2004.
- [16] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF: Speeded Up Robust Features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

- [17] Ashok Aravindan and S.M. Anzar. Robust partial fingerprint recognition using wavelet SIFT descriptors. *Pattern Analysis and Applications*, 20(4):963–979, 2017.
- [18] Joohyun Lim, Youngouk Kim, and Joonki Paik. Comparative analysis of wavelet-based scale-invariant feature extraction using different wavelet bases. In *Signal Processing, Image Processing and Pattern Recognition*, pages 297–303. Springer, 2009.
- [19] Carolina Barajas-García and Selene Solorza-Calderón. Wavelet Local Feature (WLF) pattern recognition system. *Research in Computing Science*, 147(4):125–133, 2018.
- [20] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages 511–518. IEEE, 2001.
- [21] Bruce L. Boese, Patrick J. Clinton, Danielle Dennis, Robert C. Golden, and Bryan Kim. Digital image analysis of *Zostera marina* leaf injury. *Aquatic Botany*, 88(1):87–90, 2008.
- [22] Alaa E. Abdel-Hakim and Aly A. Farag. CSIFT: a SIFT descriptor with color invariant characteristics. In *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 2006*, volume 2, pages 1978–1983. IEEE, 2006.
- [23] Jan-Mark Geusebroek, Rein Van den Boomgaard, Arnold W.M. Smeulders, and Hugo Geerts. Color invariance. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 23(12):1338–1350, 2001.
- [24] Cosmin Ancuti and Philippe Bekaert. SIFT-CCH: Increasing the SIFT distinctness

- by Color Co-occurrence Histograms. In *Image and Signal Processing and Analysis (ISPA) 2007*, pages 130–135. IEEE, 2007.
- [25] Peng Chang and John Krumm. Object recognition with Color Co-occurrence Histograms. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference*, volume 2. IEEE, 1999.
- [26] Reza J. Alitappeh, Kossar J. Saravi, and Fariborz Mahmoudi. A new illumination invariant feature based on SIFT descriptor in color space. *Procedia Engineering*, 41:305–311, 2012.
- [27] Alfred Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910.
- [28] J. Morlet, G. Arens, E. Fourceau, and D. Glard. Wave propagation and sampling theory - Part I: Complex signal and scattering in multilayered media. *Geophysics*, 47(2):203–221, 1982.
- [29] Alexander Grossmann and Jean Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, 15(4):723–736, 1984.
- [30] Ingrid Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE transactions on information theory*, 36(5):961–1005, 1990.
- [31] Yves Meyer. *Wavelets and operators*, volume 1. Cambridge university press, 1995.
- [32] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. Siam, 1992.
- [33] Paul S Addison. *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC press, 2017.

- [34] Rafael C. Gonzalez, Richard E. Woods, and Steven L. Eddins. *Digital Image Processing Using Matlab*. Gatesmark, Publishing, 2009.
- [35] Alexander Neubeck and Luc Van Gool. Efficient non-maximum suppression. In *Pattern Recognition (ICPR) 2006*, volume 3, pages 850–855. IEEE, 2006.
- [36] Matthew Brown and David G. Lowe. Invariant features from interest point groups. In *British Machine Vision Conference (BMVC) 2002*, volume 4, 2002.
- [37] John Canny. A computational approach to edge detection. In *Readings in Computer Vision*, pages 184–203. Elsevier, 1987.
- [38] Gene H. Golub and Charles F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.
- [39] Selene Solorza-Calderón. A position, rotation, and scale invariant image descriptor based on rays and circular paths. In *Applications of Digital Image Processing XXXVIII*, volume 9599, page 95991V. International Society for Optics and Photonics, 2015.
- [40] Jan-Mark Geusebroek, Gertjan J. Burghouts, and Arnold W.M. Smeulders. The Amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112, 2005.
- [41] David G. Lowe. Demo software: SIFT keypoint detector. <https://www.cs.ubc.ca/~lowe/keypoints/>, 2005.
- [42] Simon Baker, Ankur Datta, and Takeo Kanade. Parameterizing homographies. *Technical Report CMU-RI-TR-06-11*, 2006.
- [43] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

- [44] Tom Fawcett. An introduction to ROC analysis. *Pattern recognition letters*, 27(8):861–874, 2006.
- [45] Christopher M. Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [46] Simon Rogers and Mark Girolami. *A first course in machine learning*. Chapman and Hall/CRC, 2016.
- [47] Yin Zhang, Rong Jin, and Zhi-Hua Zhou. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(4):43–52, 2010.