



UNIVERSIDAD DE GUANAJUATO

CAMPUS IRAPUATO-SALAMANCA,  
DIVISIÓN DE INGENIERÍAS

Análisis de Imágenes Utilizando Atributos de Textura una  
Aproximación en 2D y 3D

TESIS

QUE PARA OBTENER EL GRADO DE:  
DOCTORA EN INGENIERÍA ELÉCTRICA

PRESENTA:

M. en I. Alejandra Cruz Bernal

DIRECTORES:

Dra. Dora Luz Almanza Ojeda  
Dr. Mario Alberto Ibarra Manzano

SALAMANCA, GTO.

Octubre 2020

*A Dios y María, Nuestra Madre...  
Con profundo amor para mis hijos  
Miguel Abraham y Alejandro Esteban.  
A mis guerreras incansables mi Madre Carmelita Bernal,  
mis hermanas Gely, Carmen y Kary mi apoyo y fortaleza.*



# Agradecimientos

*De menos infinito a más Infinito  
Siempre Gracias Dios, por que Tú no  
tienes límites.*

Este trabajo de Tesis no hubiese sido posible sin el apoyo de mis hijos, mi Madre, mis hermanas y la familia Vilches Cruz. El apoyo de mi segunda familia Romo Bernal, mis tíos Francisco y Chatis, por supuesto Lilis. Su ejemplo, entereza y palabras siempre son una pauta en mi vida. Alex Herrera, gracias.

Con gratitud para mis compañeros y compañeras del Laboratorio DS-PLab gracias por su apoyo y comprensión. De igual forma, para mis compañeros y amigos del Depto. de Ingeniería Robótica de la UPG gracias por sus palabras, muestras de apoyo y aliento.

Mi agradecimiento y reconocimiento especial para la Dra. Dora Luz Almanza Ojeda y el Dr. Mario Alberto Ibarra Manzano por su apoyo y orientación durante todo este proceso; así como para cada uno de los investigadores e investigadoras que me brindaron sus conocimientos sin otro afán que el de formar personas de bien.

# Agradecimientos Institucionales



Agradezco a la **Coordinación General de Universidades Tecnológicas y Politécnicas** y a la **Universidad Politécnica de Guanajuato** por su invaluable apoyo, para el desarrollo de los estudios de doctorado, otorgando todas las facilidades requeridas para tal fin.

Así mismo, agradezco a **PRODEP** por el apoyo brindado a través del: *Programa para el Desarrollo Profesional Docente, para el tipo Superior*, a través de la convocatoria: *Becas para Estudios de Posgrado de Alta Calidad*, **Referencia: F-PROMEPE-01/Rev-04 SEP-23-002-B**



Agradezco infinitamente a la **Universidad de Guanajuato** por ser la Casa de Estudio que me abrió sus puertas tan generosamente. Agradezco particularmente, la oportunidad brindada para hacer uso de su espacio y equipo al *Laboratorio de Procesamiento Digital de Señales* del **Campus Irapuato-Salamanca, División de Ingenierías**.

# Resumen

*Defiende tu derecho a pensar, porque  
incluso pensar de manera errónea es  
mejor que no pensar.*

Hipatia

Nuestro mundo es visualmente rico en una enorme variedad de texturas, por ende, es una característica fundamental en un amplio rango de imágenes, que van desde una vista satelital hasta microscópicas. En este trabajo, presentamos una técnica que involucra el cálculo de medidas de textura como entropía, homogeneidad, energía y agrupamiento de prominencias, tomando como base los Histogramas de Sumas y Diferencias. Parte esencial de nuestra metodología es el análisis estadístico aplicado a estas medidas con el fin de detectar marcas de interés en áreas tan disimiles como lo son, medicina y robótica.

En medicina, el desarrollo de herramientas que coadyuven en la detección temprana de cáncer de mama es fundamental. Por tanto, proponemos llevar a cabo la detección de microcalcificaciones en mamografías analizando los histogramas correspondientes al descriptor de agrupamiento de prominencias. Por otro lado, el detectar objetos en imágenes de rango, adquiridas por sensores 3D de bajo costo como Kinect es un tópico de interés en robótica. Nuestra propuesta implica procesar la información contenida en estas imágenes, utilizando características de textura, como, homogeneidad, energía y entropía. El resultado de este proceso es una imagen de la cual se computa la superficie y de esta los correspondientes descriptores. El análisis estadístico de estos últimos permite detectar un objeto contenido en una región específica. Diversas pruebas fueron realizadas utilizando bases de datos propias y genéricas. Se evaluó la efectividad de nuestra técnica en ambas áreas, mostrando que los resultados son confiables y los tiempos de procesamiento eficientes con respecto de propuestas similares.

***Abstract***

Our world is visually rich in a huge variety of textures; therefore, it is a fundamental feature in a wide range of images, that ranging from a satellite view to microscopic. This proposal presents a method that allows us to compute the entropy, homogeneity, energy and cluster prominences measurements considering the Sums and Differences Histograms technique. An essential part of this proposal is the statistics analysis applied to these measurements in order to detect marks in areas so dissimilar as medicine and robotics.

For medicine, we focus to develop a tool in order to collaborate in an earlier detection of breast cancer. The detection of microcalcifications is carried out analyzing the histograms corresponding to cluster prominences. On another hand, the object detection in range images is a recently topic of interest in robotics. Our proposal implies carried out the processing of information contains in range image using texture features as entropy, homogeneity and energy. The result of this process is an image of wich is computed its surface. The statistics analysis applied to the descriptors computed from this surface, allows us to detect object in a region of interest. Several tests are realized using generic and own database. The effectiveness of this proposal is evaluated in both areas, showing reliable results, as well processing time efficient with respect to similar techniques.

# Índice

<b>Agradecimientos</b>	<b>II</b>
<b>Agradecimientos Institucionales</b>	<b>III</b>
<b>Resumen</b>	<b>IV</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos . . . . .	1
1.2. Motivación . . . . .	2
1.3. Justificación . . . . .	4
1.4. Hipótesis . . . . .	5
1.5. Antecedentes y Conceptos Generales . . . . .	5
1.5.1. Agarre . . . . .	5
1.5.2. Sistemas en Tiempo Real . . . . .	6
1.5.3. Procesamiento de Información Dispersa . . . . .	8
1.5.4. Imágenes Digitales . . . . .	10
1.5.5. Sensor Kinect . . . . .	19
1.5.6. Puntos de Interés y Descriptores . . . . .	24
1.5.7. Detección y Segmentación de Objetos 3D . . . . .	25
1.5.8. Textura . . . . .	28
1.5.9. La Técnica SDH . . . . .	33
1.6. Definición del Problema . . . . .	36
1.7. Descripción General de la Propuesta . . . . .	37
1.7.1. Aportaciones . . . . .	39
1.8. Organización del Contenido . . . . .	40
<b>2. Metodología Parte I: Características de Textura en Imá- genes DICOM y de Rango</b>	<b>42</b>
2.1. Procesamiento de Imágenes DICOM mediante la Técni- ca SDH . . . . .	42
2.1.1. Extracción de la Característica de Textura CP . . . . .	43

2.1.2. Análisis de Calcificaciones . . . . .	45
2.1.3. Clasificación de las Mamografías . . . . .	50
2.2. La técnica SDH Usando Imágenes de Rango . . . . .	51
<b>3. Metodología Parte II: Segmentación y Detección en Es-</b>	<b>56</b>
<b>cala de Grises</b>	
3.1. Consideraciones de Superficie para Imágenes de Rango	57
3.2. Proceso Denso . . . . .	61
3.3. Proceso Disperso . . . . .	63
3.3.1. Consideraciones de Matrices Dispersas . . . . .	63
3.3.2. Selección de la Información para el Proceso Dis-	
perso . . . . .	64
3.3.3. Pseudocódigo de la técnica SKD . . . . .	65
3.3.4. Algoritmo para la detección de un objeto en la	
ROI . . . . .	71
3.3.5. Proceso de Validación . . . . .	74
<b>4. Análisis de Entropía en las Imágenes de Rango</b>	<b>76</b>
4.1. Entropía para Imágenes de Rango . . . . .	76
4.2. Entropía como Característica de Textura . . . . .	78
4.2.1. Maximización y Minimización de la Entropía . . . . .	79
4.2.2. Minimización de la Entropía y su Relación con	
la Energía . . . . .	80
4.2.3. Maximización de la Entropía y su Relación con	
la Homogeneidad . . . . .	82
<b>5. Resultados</b>	<b>83</b>
5.1. Resultados de la Metodología I . . . . .	83
5.2. Resultados de la Metodología II . . . . .	88
5.2.1. Resultados Previos de Imágenes en Niveles de Gris	89
5.2.2. Resultados Previos de Imágenes de Rango . . . . .	91
5.2.3. Análisis Cuantitativo . . . . .	94
5.3. Resultados del Análisis de la Entropía . . . . .	102
5.3.1. Análisis Cuantitativo . . . . .	103
<b>6. Conclusiones y Trabajos Futuros</b>	<b>109</b>
6.1. Conclusiones . . . . .	109
6.2. Trabajos Futuros . . . . .	112
<b>Bibliografía</b>	<b>113</b>

# Índice de figuras

1.1.	a) Aplicación del robot Scorbot en el juego Tic-Tac-Toe. b) Manipulador industrial Yaskawa Motoman (c) Banda transportadora y robots manipuladores en laboratorio de Automatización (Universidad Politécnica de Guanajuato).	3
1.2.	Esquemas Ilustrativos. a) Diagrama general de un manipulador industrial de 3 GDL [115]. b) Analogía ilustrativa de la estructura general de un robot manipulador en comparación con la morfología de un humano [28].	6
1.3.	Esquemas representativos de agarres. a) Tipos de agarre de acuerdo a la forma de ser contenido o sujetado el objeto [21]. b) Ejemplo de agarre de una copa de cristal por un manipulador de servicio [129].	7
1.4.	Diagrama ilustrativo de una a) imagen bidimensional $f(s, t)$ la cual a través de un proceso de b) muestreo y cuantización, se obtiene como resultado una c) imagen digital $f(x, y)$ [46].	11
1.5.	Diagrama ilustrativo de una imagen digital [46] ubicando en la coordenada espacial $(x, y)$ un píxel; tal que, $f(x, y)$ es igual a un nivel de intensidad en la escala de grises, el cual corresponde a dicho píxel.	12
1.6.	Imagen en a) nivel de gris y b) a color [97]. Los valores que se muestran en las matrices se encuentran normalizados.	13
1.7.	Diagrama ilustrativo de una arquitectura básica PACS [26].	14
1.8.	Ejemplo ilustrativo de una mamografía en formato DICOM. Nótese la información médica contenida en las cuatro esquinas de la imagen ( <i>imagen cortesía del Hospital General de Irapuato</i> ).	16
1.9.	Ejemplo de imágenes de a) mallado [95]; b) disparidad [58]; c) nube de puntos [144] y d) profundidad [31].	18
1.10.	Formación de las imágenes desde un sistema estéreo[27]	19

1.11. Sensor Kinect XBox360 [103] . . . . .	20
1.12. Espacio de color RGB [105] . . . . .	21
1.13. Ejemplo de las imágenes de salida del sensor Kinect [135]	23
1.14. Ejemplo de imagen obtenida del sensor Kinect para la base de datos DSPLab. a) Imagen a color en formato VGA y b) representación en imagen monocromática de la información adquirida por el sensor de profundidad. . . . .	23
1.15. Ejemplo de correspondencias bajo diferentes tipos de varianzas (ruido). Técnicas como (a) BRAND [15], b)RCS [161], SIFT [88] y d) HGND [167] presentan robutez ante la rotación, escalamiento o iluminación. . . . .	25
1.16. Ejemplo de algunos descriptores mencionados en 1.5.7. a) Geometric Scale Space [170]; b) Local Reference Frame [102]; y c) Scale Invariant Features [83]. . . . .	27
1.17. a)La imagen a color (Baby Middlebury dataset) es sólo referencia. b) Imagen de rango (disparidad). c) Presenta la superficie formada por los diferentes niveles de gris correspondientes a la imagen de disparidad. . . . .	34
1.18. Diagrama general de la propuesta. . . . .	37
2.1. Esquema general de la técnica empleada para la detección de microcalcificaciones en mamografías. . . . .	43
2.2. Histogramas de características de textura usando SDH para una imagen en nivel de gris. En el primer renglón de izquierda a derecha las características de energía, entropía, homogeneidad y agrupamiento de prominencias. En el segundo renglón se muestran los histogramas cada característica. . . . .	44
2.3. Diagrama descriptivo correspondiente al <i>análisis de las calcificaciones</i> (analysis of calcifications) del esquema 2.1.	46
2.4. Ejemplo de una interpolación global para $h_{cp}$ . Los * muestran los valores discretos de $h_{cp}$ . Mientras que, la línea azul corresponde a la interpolación polinomial de $f(cp)$ , y los $\circ$ marcan los puntos de inflexión ( $Z_{cp}$ ). . .	48
2.5. Los * muestran los valores de $h_{cp}$ . La línea azul corresponde a la interpolación global $f(cp)$ y los $\circ$ son ( $Z_{cp}$ ). Las dos primeras imágenes primer renglón, corresponden al primer caso de la ecuación 2.4, mientras que la última al segundo caso de la misma ecuación. Eje horizontal valor de CP. Eje vertical frecuencia (píxel/imagen). . .	49

2.6.	Comparación de una interpolación global (imagen izquierda) $h_{cp}$ , con respecto de una local (imagen derecha) $h_{cp}^*$ . . . . .	50
2.7.	Ejemplos de la técnica SDH aplicada a) DSPLab Dataset (adquisición de las imágenes por medio del sensor Kinect; en un ambiente no controlado) y b) Middlebury dataset (adquisición de ambas imágenes por medio de un sistema par-estéreo bajo condiciones controladas). Diferentes distancias [1, 1] y [3, 3]. La imágenes a color son sólo referencia. . . . .	53
3.1.	Diagrama general para la detección de un objeto en una región de interés. . . . .	57
3.2.	(a) Superficie imagen de rango. En (b) la figura de (a) es representada como una semiesfera en el espacio $\mathbb{R}^3$ . (c) superficie generada de la imagen de homogeneidad y (d) curvatura correspondiente a dicha característica; e) y f) imágenes de la división en las direcciones $x$ y $y$ de la imagen del inciso (d). . . . .	58
3.3.	De a) la imagen de rango (Bowling1 [58]) se extraen los descriptores que forman las superficies correspondiente a las características de c) energía; d) entropía y e) homogeneidad. Mientras que en b) se muestra la superficie formada por (a). . . . .	59
3.4.	Ejemplo de la representación como imágenes de algunas características de textura tal como a) homogeneidad, b) entropía y c) energía obtenidas a partir de la imagen de rango (Bowling1[58]) representadas en imágenes formadas por medio de isolíneas. . . . .	60
3.5.	Ejemplo de la representación de la característica de textura homogeneidad como imagen ( $M_{homogeneity}$ ) correspondiente a la $M_{range\_cubito4}$ de nuestra base de datos DSPLab. La división de a) superficie de la imagen de homogeneidad, en superficies direccionales se muestran en b) $M_x$ , c) $M_y$ y d) $M_z$ . . . . .	61
3.6.	Diagrama esquemático de la primera fase (proceso <i>denso</i> ) 3.2 de la técnica propuesta. . . . .	62
3.7.	Diagrama representativo [139] de la formación de los vectores normales densos en la superficie de textura, representada por la función en forma implícita $F(x, y, z) = 0$ . . . . .	63

3.8.	Diagrama esquemático que ilustra el inicio del proceso disperso. a) Representación de los vectores contenidos en las superficies direccionales $M_x$ , $M_y$ y $M_z$ . b) Representación en $\mathbb{R}^2$ de la posición de los vectores en estas superficies. . . . .	65
3.9.	Diagrama representativo a) de los vectores más sobresalientes en cada dirección, b) para obtener puntos de interés. . . . .	66
3.10.	Ilustración de los pasos 1 a 6 del algoritmo 1 de la técnica SKD aplicada a la imagen bird (en niveles de gris representada en hsv) de base de datos conocida. A su izquierda su representación en su forma Gaussiana y a la derecha como superficie. Segundo renglón representación de la forma Gaussiana de diferentes descriptores de textura. Tercero, superficie densa de la que se seleccionara la información para el proceso disperso. . . . .	69
3.11.	Diagrama general de la técnica SKD aplicada a la detección de un objeto ubicado en una ROI. . . . .	71
3.12.	Ejemplo ilustrativo de la representación de la <i>conectividad</i> - <i>m</i> entre los puntos de interés ( $S_K$ ). . . . .	74
3.13.	Diagrama esquemático para validar puntos de interés (con base en la propuesta de Muja y Lowe en [106] y [91]). . . . .	75
4.1.	Cálculo de la entropía para una imagen de rango (salsita-DSPLab dataset). El valor de entropía calculado es de 0.7. Este se encuentra marcado en los respectivos histogramas por un recuadro negro. a) Representación de la entropía como imagen, así como, b) su respectivo histograma de frecuencias. c) Representación de la maximización de la entropía como imagen y d) su correspondiente histograma. e) Representación de la minimización de la entropía como imagen, así como, f) su respectivo histograma (más detalles en la sección 4.2.1). . . . .	79
5.1.	a) Mamografía en formato <i>DICOM</i> 3,0. b) Imagen resultante de la mamografía procesada para estimar CP. . . . .	84
5.2.	Ejemplo de los resultados de calcificaciones detectadas. a) Calcificaciones y tejido normal detectado por el experto. b) Calcificaciones detectadas con la interpolación local. . . . .	84

5.3.	Ejemplo de los resultados de calcificaciones detectadas. a) Calcificaciones y tejido normal detectado. b) Calcificaciones detectadas con la interpolación local. . . . .	85
5.4.	Resultados de la clasificación. (a) Mamografía en formato DICOM, los círculos resaltan las zonas con calcificaciones encontradas por un experto, (b) atributo de CP con las calcificaciones encontradas por el método propuesto, (c) acercamiento a las microcalcificaciones detectadas. . . . .	86
5.5.	Ejemplo ilustrativo de los puntos de interés obtenidos usando la técnica SKD calculando la homogeneidad como característica de textura, aplicados a la imagen bird de base genérica conocida. . . . .	88
5.6.	Ilustración comparativa de los puntos de interés generados con técnicas de trabajos previos como son NARF; Harris; FAST y SURF. Las primeras dos imágenes corresponden a los puntos de interés obtenidos en las direcciones $x$ y $y$ con la técnica propuesta. . . . .	89
5.7.	Ilustración de los puntos de interés generados con la técnica SKD en la dirección $y$ ; así como otras técnicas ya conocidas y revisadas en la subsección 1.5.7. . . . .	90
5.8.	Ilustración que muestra el seguimiento de los puntos de interés (rojos imagen disparidad izquierda y verdes para la derecha) computados de las imágenes de disparidad de baby1 [58]. Estos son evaluados de acuerdo con lo presentado en 3.3.5. . . . .	91
5.9.	Ilustración de los puntos de interés “calculados” con la técnica SKD, así como otras técnicas de trabajos previos. La imagen original bowling2 [58], es rotada $60^\circ$ para producir la imagen sintética. . . . .	93
5.10.	Ilustración comparativa de los puntos de interés que se obtienen aplicando las técnicas b) SKD (dirección en $x$ ), c) Harris; d) NARF; e) SURF; f) FAST; a) imagen de profundidad correspondiente a nuestra base de datos. . . . .	93
5.11.	La figura presenta los resultados correspondientes a la rotación de la información contenida en la imagen de rango ( $60^\circ$ ). Las pruebas realizadas con respecto de otras bases de datos son: b) nuestra técnica; c) SURF; d) HARRIS; e) FAST; f) NARF. a) Es la imagen de profundidad, solo como referencia. . . . .	94

5.12. Resultados obtenidos de aplicar la técnica SKD, a las imágenes de profundidad de la base de datos DSPLab. Primer renglón cubo de Rubik, segundo renglón salsa, tercer renglón taza. . . . .	95
5.13. Comparación cualitativa con respecto a la base de datos Middlebury ([112] y [58]), de los resultados obtenidos para la detección de puntos de interés. Nuestra técnica SKD (segunda columna), tercera columna <i>S&amp;T</i> , cuarta Harris, quinta FAST y sexta SURF. Primer renglón Art, segundo renglón Laundry. . . . .	96
5.14. Ejemplo de imágenes sintéticas generadas con ruido Gaussiano ( <i>D</i> ) y sal y pimienta ( <i>S&amp;P</i> ) y algunas combinaciones entre estas. . . . .	97
5.15. a) Ejemplo de la generación de imágenes sintéticas usando la base de datos Middlebury b) Comparativa del índice de repetibilidad usando la técnica SKD con respecto de las técnicas <i>S&amp;T</i> , Harris, SURF y FAST. . . . .	98
5.16. Resultados obtenidos de imágenes sintéticas con $\alpha = 1$ y a) $\theta = 30^\circ$ ; b) $\theta = 60^\circ$ ; c) $\theta = 150^\circ$ d) $\theta = 180^\circ$ . e) $\theta = 0^\circ$ y diversas escalas $\alpha$ . . . . .	99
5.17. Gráfica comparativa del tiempo de procesamiento. El eje vertical izquierdo corresponde al tiempo en s, entre bases de datos de dimensiones similares. El de la derecha, corresponde a una base de datos de una dimensión mayor. La dimensión de cada base se expresa en $K \times L$ píxeles. . . . .	100
5.18. Los descriptores $N_S$ corresponden a la superficie direccional $M_x$ . Estos han sido seleccionados del vector $N_D$ . Mientras que los $S_K$ se obtienen de $N_S$ . . . . .	101
5.19. Resultados comparativos de la maximización y minimización de la entropía. Primera columna imágenes a color sólo como referencia. De la imagen superior a la inferior: cubo de Rubik, cubo amarillo, bote blanco, tacita, de la base de datos DSPLab. Segunda y tercer columna minimización de la entropía y sus resultados. De la cuarta a la sexta columna los resultados de la maximización. . . . .	102
5.20. Ejemplo comparativo de la cantidad de descriptores contenidos en cada vector de atributos obtenidos en el proceso denso ( $N_D$ ) y disperso $N_S$ y $S_K$ . . . . .	104
5.21. Ejemplo ilustrativo del total de atributos contenidos en $M_y$ en $N_S$ la segmentación de estos, en $S_{K_{contour}}$ para el contorno de la escena y de la ROI ( $S_{K_{ROI}}$ ). . . . .	105

---

5.22. Ejemplo ilustrativo de la comparación de los $S_k$ obtenidos para el contorno de la escena con respecto a los de la ROI. . . . .	105
5.23. Ejemplo esquemático de la cantidad de descriptores requeridos para el cómputo de los puntos de interés que definan el contorno de la escena y la ROI. . . . .	106
5.24. Relación establecida entre los descriptores correspondientes a los diferentes contornos obtenidos partiendo de la información contenida en $M_y$ . . . . .	107

# Índice de Tablas

1.1. Tabla con los descriptores de texturas más comunes expresados desde la matriz de coocurrencia [52] e histogramas de sumas y diferencias [150]. . . . .	35
5.1. Matriz de confusión correspondiente a la función $g(cp)$ , para $K = 3$ con una distancia Euclidiana. . . . .	86
5.2. Resultados del rendimiento presentado por las métricas con respecto del método propuesto. . . . .	87
5.3. Tabla comparativa de nuestra método con respecto de otros trabajos previos. . . . .	87
5.4. Tabla comparativa de nuestra método con respecto de otros trabajos previos. . . . .	92
5.5. Tabla comparativa del número de correlaciones reales en promedio y la precisión obtenida aplicando el proceso de validación referido en 3.3.5 a la base de datos Middlebury. . . . .	92
5.6. Relación presentada de los objetos detectados haciendo uso de la técnica propuesta. . . . .	101
5.7. Valores promedio del número de descriptores procesados por la técnica propuesta. . . . .	107
5.8. Estimación del tiempo promedio (segundos) de ejecución de la técnica propuesta con respecto de otras técnicas revisadas en la introducción a trabajos previos. . . . .	107
5.9. Razón establecida entre el promedio del tiempo de ejecución y el número de descriptores calculados $\Delta(t)/\Delta(\#_{descriptores})$ . . . . .	108

# Capítulo 1

## Introducción

*El hombre inteligente no es el que tiene  
muchas ideas, sino el que sabe sacar  
provecho de las pocas que tiene.*

Anónimo

A través de este capítulo se presenta una introducción del trabajo de Tesis desarrollado para el procesamiento y análisis de imágenes en escala de grises.

Los resultados obtenidos de dicho procesamiento son presentados para formatos de imágenes como lo son imágenes de mamografías e imágenes obtenidas haciendo uso del sensor Kinect.

Para ello, en esta sección se determinan los objetivos, se desarrolla el planteamiento del problema y se establece la justificación. Finalmente, concluimos presentando la organización del trabajo de tesis.

### 1.1. Objetivos

#### *General:*

Desarrollar una técnica que permita procesar imágenes en escala de grises, como pueden ser las mamografías y las imágenes de profundidad o disparidad, haciendo uso de técnicas enfocadas en la extracción de características de textura; con el fin de generar herramientas que ayuden en la detección global de microcalcificaciones en una mamografía o bien, puntos de interés en imágenes de profundidad o disparidad.

#### *Específicos:*

1. Obtener diferentes características de textura tales como, homo-

- geneidad, agrupamiento de prominencias, entropía, energía, entre otros, a partir de la información contenida en imágenes de mamografías, así como, de profundidad extraídas por un sensor RGB-D.
2. Desarrollar e implementar un detector de puntos de interés con base en el cálculo y análisis de las normales obtenidas de la información contenida en las imágenes de rango.
  3. Diseñar, implementar y validar el uso de las características de textura en la detección temprana de zonas con micro-calcificaciones en imágenes DICOM como una herramienta confiable que apoye al experto, en el diagnóstico del cáncer de mama.
  4. Analizar la invarianza al ruido de los puntos detectados tal como, cambios de luminosidad, rotación o escalamiento de la estrategia propuesta.
  5. Determinar la región de interés en la cual se encuentre el elemento a identificar mediante el cálculo de normales densas y luego dispersas sobre la imagen de profundidad.

## 1.2. Motivación

La detección de objetos en escenarios reales es un tópico de interés en constate investigación en visión por computadora. Las técnicas para llevar a cabo dicha tarea involucran desarrollar algoritmos computacionales que permitan un procesamiento en tiempo real. En ese sentido, nuevas tecnologías de software y sensores 3D permiten que, los costos de los sistemas de adquisición sean bajos en relación con sistemas de visión convencionales. Por lo que, el sector industrial ha centrado su interés en el desarrollo de aplicaciones para mejorar o automatizar sus procesos productivos con este tipo de tecnologías. Por ende, el sector educativo se ha visto en la necesidad de involucrarse también con el fin de formar a profesionistas que tengan o puedan generar competencias en el manejo de estas tecnologías.

En el ámbito industrial, los robots manipuladores, son mecanismos re-programables capaces de desarrollar diversas actividades de manera repetitiva y empleando la misma cantidad de tiempo en cada repetición. Estos robots realizan tareas que involucran, principalmente, la manipulación de objetos como piezas, herramientas, materiales de diferentes tamaños y formas, entre otras. La manipulación de objetos por un robot industrial requiere de sensores adicionales que proporcionen la coordenadas exactas del objeto a sujetar y que le entregue al robot

información del escenario, para evitar colisiones mientras realiza la trayectoria y toma del objeto.

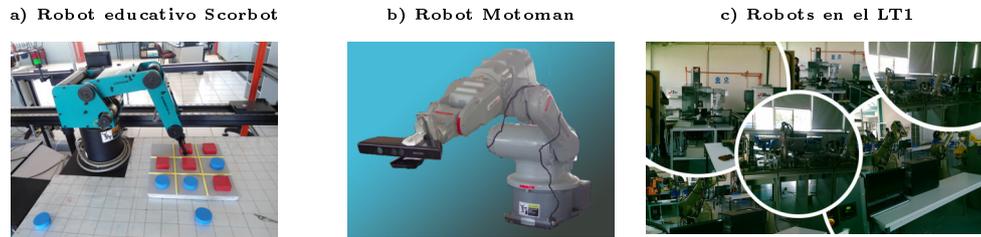


Figura 1.1: a) Aplicación del robot Scorbobot en el juego Tic-Tac-Toe. b) Manipulador industrial Yaskawa Motoman (c) Banda transportadora y robots manipuladores en laboratorio de Automatización (Universidad Politécnica de Guanajuato).

La automatización de tareas por un robot manipulador es un tema de investigación ampliamente estudiado desde finales del siglo pasado, y que con el crecimiento de las capacidades en la tecnología de cómputo, permiten continuar con la propuesta de nuevas y mejores técnicas de solución al problema. Dentro de estas nuevas propuestas se cuenta con cámaras o sensores de visión, que prometen desarrollar las mismas tareas que un sistema de visión industrial, pero con un costo más asequible. Esto implica el desarrollar nuevas técnicas para el procesamiento de imágenes adquiridas por este tipo de dispositivos; con las cuales nos permitan realizar la detección, modelado y reconstrucción de objetos con el fin de llevar a cabo tareas de automatización y manipulación, por ejemplo.

Así la principal motivación de este proyecto es diseñar y desarrollar diferentes algoritmos de procesamiento de imágenes que permitan detectar características para obtener una zona de manipulación del objeto con base a su tamaño y posición. Las características propuestas pueden ser evaluadas para detectar zonas pequeñas, como las microcalcificaciones en mamografías para garantizar su nivel de descripción de las zonas de interés en una imagen. Una vez obtenidas las características y la zona del objeto, es de nuestro mayor interés proporcionar este algoritmo como una metodología y herramienta de enseñanza de la robótica en manipuladores industriales, para estudiantes de licenciatura en carreras afines, en particular de ingeniería robótica que es mi adscripción de trabajo. La Figura 1.1 muestra los robots manipuladores con los que se cuenta en el laboratorio de Automatización de la Universidad Politécnica de Guanajuato y que es utilizado en las materias propias al área

de robótica y automatización.

### 1.3. Justificación

Un área de interés fundamental en visión por computadora es la detección y clasificación de objetos 3D para su manipulación durante la realización de tareas o en funciones más avanzadas, como lo es la interacción hombre-máquina. En ello se requiere una respuesta en tiempo real por parte del sistema lo cual involucra cálculos de software y respuesta de actuadores mecánicos o hardware. La introducción de sensores RGB-D permite el procesamiento de imágenes representadas con color (imagen RGB) y en escala de grises. El procesamiento de imágenes con color es más común debido a que simplifica la detección de un objeto en una escena por sus componentes. Por otro lado, la imagen de profundidad (D), representada en intensidades de gris, permite conocer en perspectiva, la posición de los objetos en coordenadas 3D mediante un método de proyección de puntos utilizando la óptica y parámetros de la cámara.

En este contexto, esta Tesis busca hacer un análisis de las técnicas clásicas para la extracción de características de textura, utilizando un sensor RGB-D comercial y económico que permita ofrecer una alternativa de solución a problemas de detección de objetos en un escenario de interior. Adicionalmente, se pretende compilar una estrategia, rápida, económica, básica pero fundamentada en investigación, que permita la demostración completa del proceso que implica la detección de objetos en el espacio 3D procesando solamente la imagen de profundidad. Esta estrategia será implementada como metodología de base para el desarrollo de cursos de robótica, visión por computadora y tópicos avanzados de robótica para alumnos de licenciatura en ingeniería, principalmente en Robótica, Manufactura e Ingeniería Automotriz de la Universidad Politécnica de Guanajuato, donde soy profesor titular de tiempo completo. La propuesta es tener un prototipo demostrativo y toda una metodología de cómo el procesamiento de imágenes interviene en la interacción humano-robot. Las tareas que se pretenden demostrar y desarrollar de forma didáctica son la captura de las imágenes de color y profundidad en escenarios que involucren la manipulación de objetos. Esto con el fin de realizar la validación de las características de textura en imágenes en nivel de gris, así como, el proceso de reproyección del objeto de un plano 2D al espacio 3D con coordenadas en el mundo. Se espera también que la formación brindada permita a los egresados tener una base de conocimientos en un área de interés particular que les ayude a proyectarse realizando un posgrado.

## 1.4. Hipótesis

El análisis de imágenes de profundidad para la detección de puntos de interés desde una perspectiva de normales y los principios de superficies mínimas permite detectar puntos que presentan una invarianza ante rotación, escalamiento y ruido.

## 1.5. Antecedentes y Conceptos Generales

Las áreas de la Robótica involucran tareas de percepción, interacción y manipulación de los objetos en un medio ambiente que puede ser estructurado o no. En las últimas décadas, la Robótica Industrial, ha integrado los sistemas de visión como un elemento fundamental de los procesos de manufactura avanzada. En este contexto, es elemental que el robot industrial o manipulador “aprenda” de forma precisa, el cómo tomar los objetos a manipular, teniendo siempre una posición visual del objeto, la cual le permita llevar a cabo la sujeción y contención de este.

El realizar y programar tareas en un robot depende en gran medida, del nivel de exactitud con el que se solicite llevar a cabo dichas tareas. En el sector industrial los robots son programados para realizar tareas muy específicas. Este tipo de programación generalmente sólo requiere del mantenimiento que el operador le pueda dar al código con el cual el robot interactúa con su medio. Además, por razones de seguridad, no se puede tener contacto directo entre el operador y el espacio de trabajo del manipulador, por lo que el manipulador trabaja sólo en zonas muy controladas. Por ejemplo, en una línea de ensamblaje, donde las rutinas deben ejecutarse en tiempo real, el robot debe de tener un conocimiento del entorno, así como de su propio espacio de trabajo, con el objetivo de saber exactamente en qué posición debe tomar los objetos o cómo deben estos llegar para realizar algún proceso en particular.

### 1.5.1. Agarre

Considerando la estructura básica de cualquier manipulador industrial 1.2 a, un brazo robótico o manipulador contiene elementos, segmentos o eslabones unidos por articulaciones; que permiten un movimiento relativo entre cada dos eslabones consecutivos. De acuerdo con el número de articulaciones (parámetros independientes) se define los *grados de libertad* (GDL, por sus siglas) de este. El extremo final del robot se le conoce como “muñeca”, como se puede apreciar en la Figura 1.2. Por lo tanto, su posición está definida por los GDL del brazo. Por

el contrario, los dispositivos que son colocados en la muñeca presentan su propia morfología, por lo que, el número de GDL dependerá exclusivamente de esta. Los elementos que pueden ser conectados a la muñeca reciben diversos nombres tal como, *elemento terminal* [125], *efector final* [11], *órgano terminal* [154] o *actuador final* [104]; los cuales pueden ser elementos de aprehensión o herramientas, su elección dependerá de la tarea específica a realizar.

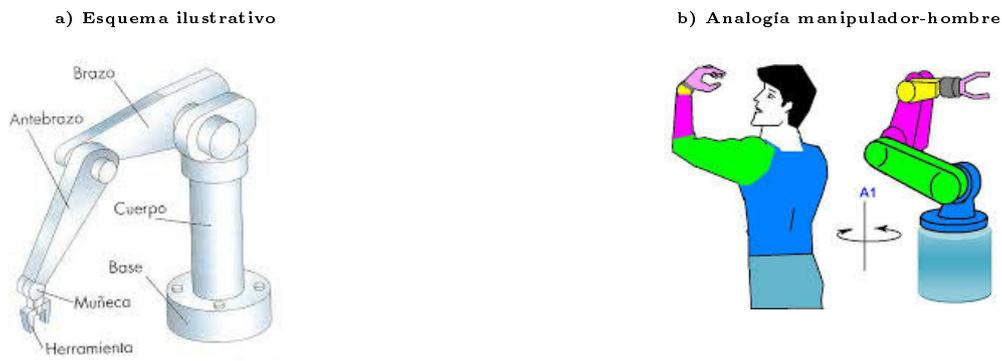


Figura 1.2: Esquemas Ilustrativos. a) Diagrama general de un manipulador industrial de 3 GDL [115]. b) Analogía ilustrativa de la estructura general de un robot manipulador en comparación con la morfología de un humano [28].

La manipulación del objeto involucra en primera instancia la sujeción, contención o “agarre” (*grasping*, en inglés) de este. Por tanto, El desarrollo de técnicas para llevar a cabo la tarea de *agarre*, se ha vuelto uno de los tópicos con mayor auge en el área de Robótica, debido principalmente a la complejidad que involucra la generación de una trayectoria. Esto es, llevar al robot de una posición a otra, mediante una serie de puntos intermedios que deben estar referidos preferentemente a un sistema de coordenadas referenciado, como lo es el cartesiano o polar, por ejemplo.

La forma en la que es “tomado” el objeto implica una emulación del cómo el ser humano lleva a cabo este tipo de actividad (ver Figura 1.3).

### 1.5.2. Sistemas en Tiempo Real

De acuerdo con lo expuesto por Kopetz en su libro [73], un sistema de tiempo real es aquel en el cual los aspectos de comportamiento temporal forman parte de su especificación. Esto es, el funcionamiento de estos sistemas, no sólo depende de la exactitud de los resultados aritméticos y lógicos, sino también del instante en que estos se producen.

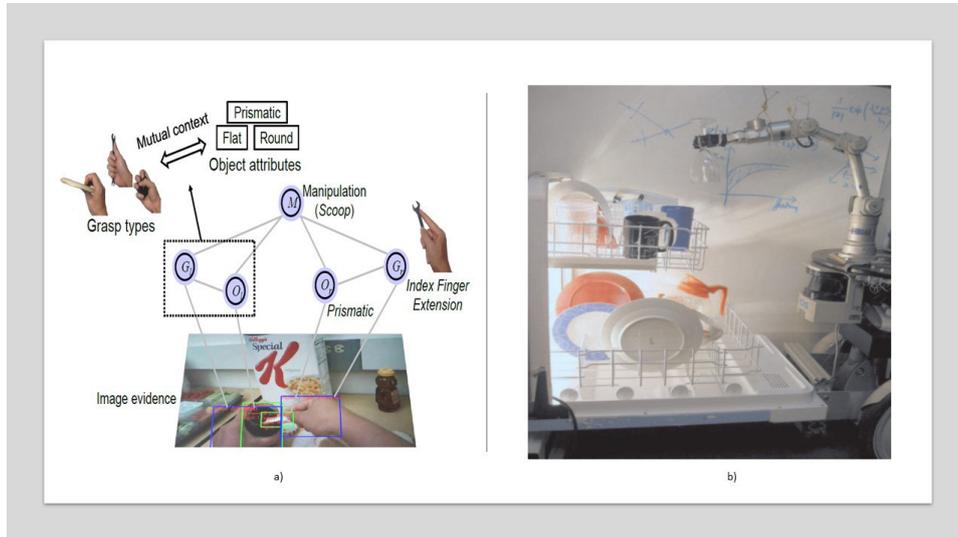


Figura 1.3: Esquemas representativos de agarres. a) Tipos de agarre de acuerdo a la forma de ser contenido o sujetado el objeto [21]. b) Ejemplo de agarre de una copa de cristal por un manipulador de servicio [129].

En áreas como la robótica donde el robot interactúa de forma directa con su entorno, en intervalos de tiempo definidos, es imperativo la supervisión de los datos de entrada para reaccionar en consecuencia. A esta “reacción” se le conoce como *vencimiento*.

Si un resultado ha sido útil aún cuando ha pasado el vencimiento, entonces, el sistema se clasifica como *blando* o *suave*, de lo contrario se clasificará como *firme*. Más aún si las condiciones temporales son restrictivas, es decir, el tiempo límite establecido de ejecución permite la obtención de resultados útiles entonces, los sistemas son conocidos como sistema en tiempo real *duro* o de *misión crítica*. Por ejemplo, un sistema de tiempo real suave es el de la transmisión multimedia; en la cual la pérdida ocasional de algún cuadro de video no trae grandes consecuencias. Por otro lado, un sistema de control que presenta una respuesta tardía, con la cual se provocará una falla o ineficiencia en el proceso puede ser clasificado como un sistema firme o incluso duro.

En el desarrollo de esta propuesta, se considerará un sistema de tiempo real suave. Para lograr el objetivo de trabajar en tiempo real se realiza el procesamiento de la información a través de algoritmos tipo *sparse* o *disperso*, lo cual permite que el rendimiento temporal del sistema sea suficiente, para que la respuesta sea considerada como útil. Para reducir el problema de la incertidumbre en relación con el procesamiento en tiempo real subdividiremos el problema de la percepción

(objeto-robot) en dos subprocesos básicos, el primero implica el procesamiento denso de la información y el segundo un procesamiento tipo disperso.

### 1.5.3. Procesamiento de Información Dispersa

Los algoritmos que tratan información *dispersa* se fundamentan en el procesamiento de las matrices denominadas *matrices dispersas* [120]. Una característica fundamental en cualquier algoritmo que procesa este tipo de información es el no almacenamiento de la información nula. Por lo que, se debe de definir estructuras de datos que sean adecuados para la implementación eficiente de métodos de solución estándar, ya sean directos o iterativos.

#### ***Matrices Dispersas***

Una matriz se considera *dispersa* cuando el número de elementos no nulos es del orden de  $\sqrt{n}$ ; siendo  $n$  el número de elementos de la matriz [120]. Existen básicamente dos tipos de matrices dispersas las estructuradas y no estructuradas. Las primeras tienen la característica de seguir un “patrón” en relación con la información contenida; esto es, la información *no cero* o *no nula* se presenta en un número determinado de “submatrices densas” acomodadas en un patrón regular a lo largo de un reducido número de diagonales principales. Por el contrario, en las matrices no estructuradas, la información se encuentra dispersa a lo largo de la matriz.

El trabajar con matrices dispersas nos permite tener algunas ventajas, como son, ahorrar cantidad de memoria que se ocupa almacenando información nula. Por lo que, se puede definir o aplicar métodos para realizar cualquier tipo de operación matricial, pero con la condición de tener un almacenamiento disperso. Los errores de resolución para matrices dispersas, no son equivalentes a resolver una matriz densa con  $n$  elementos sino una con  $\sqrt{n}$ .

Algunos trabajos en el área de visión por computadora haciendo uso de información dispersa, los podemos ver en Frome *et al.* en [40]. En el cual, se muestra una técnica para el reconocimiento de objetos 3D en escenas con mucho ruido o con un alto grado de oclusión. Los autores describen como llevando a cabo un “muestreo” de la información, es posible generar una serie de descriptores regionales, a los que ellos nombran como *contextos de formas 3D* (3D shape context, en inglés) y *contextos de formas armónicas* (harmonic shape context, en inglés). El concepto general es que dichos descriptores caracterizan la forma regional en la superficie de un objeto. Mientras que, Johnson *et al.*

en [68] hacen uso de la nube de puntos obtenidos a partir de un sensor 3D. Partiendo de esta información, la cual se considera como dispersa, los autores generan un mallado dando como resultado una superficie poligonal, con esta, y la adyacencia entre puntos establecen una correspondencia para la generación primero de imágenes 2D, posteriormente, de la superficie del objeto.

Sin embargo, aún cuando son procesos más rápidos o que consumen menos recursos de cómputo, se debe hacer notar que una de las desventajas que presentan es una cierta falta de repetibilidad con respecto de los puntos de interés detectados. De igual forma, es posible tener discriminación de información relevante, que debería ser considerada para la formación de los descriptores [101]. Sin embargo, este tipo de técnicas son empleadas con mayor frecuencia cuando se requiere de procesamiento de video en tiempo real. Uno de los trabajos más destacados lo podemos encontrar en la técnica propuesta por Dóllar *et al.* [35]. Esta técnica se fundamenta en el trabajo presentado por Laptev *et al.* [78]. En la misma se describe cómo, a través de la obtención de características espacio-temporales, es posible llegar a la representación compacta 3D de la escena, así como, del objeto de interés al cual se quiere dar seguimiento; tal como lo demuestra Dóllar en [35].

### ***Técnicas dispersas aplicadas en la percepción visual***

Cuando se trabaja con percepción visual para la detección, identificación y reconstrucción de objetos, se genera información espacial-visual de la cual se puede abstraer al objeto de la escena. Esto se puede lograr, realizando procesamiento en tiempo real a través de *algoritmos dispersos* que permitan redefinir la información espacial. Esto es, en espacios de interés (**SOIs**, por sus siglas en inglés) para llevar a cabo una segmentación adecuada del objeto respecto a su entorno, como lo presenta Zhou *et al.* en [166]. En este trabajo los autores presentan una serie de propiedades geométricas, con las cuales sugieren se puede garantizar la invarianza y repetibilidad en los puntos de interés detectados.

En particular, haciendo uso del concepto de *causalidad*, en el cual se sugiere que, todos los valores mínimos y máximos de la función de curvatura Gaussiana pertenecen a la forma original, dentro del espacio de la escala definida por dicha curvatura. Esta misma propiedad, la retoma Bariya *et al.* en [10]. En dicho trabajo, los autores realizan un embebido denso 2D del campo normal de la superficie, a través de una serie de suavizados Gaussianos. En contraste con lo propuesto por Zuo

*et al.* [166] quien obtiene el suavizado de la superficie a través del Flujo de Ricci (Ricci Flow, en inglés). Algunos trabajos similares los podemos encontrar en Zhuo *et al.* y su trabajo realizado en la correspondencia de características en escenas 3D aglomeradas [167], aplicando en principio la técnica presentada por Lowe en [91] en el cual los descriptores locales son referenciados a una escala variante, mientras que, Zhuo propone hacerlo a través de una escala fija. Lim y Lee en su trabajo [83] hacen uso de lo propuesto por Lowe, haciendo una extensión a superficies 3D. Una propuesta similar se encuentra en [84], desarrollada por Lin *et al.* quienes presentan una técnica para la detección de bordes usando información dispersa. De forma semejante, You y Huang en [163] desarrollan una técnica para el reconocimiento de objetos, basado en características que tengan similitudes con respecto a una disimilitud de los objetos contenidos en la escena.

En nuestra propuesta nosotros establecemos que, las imágenes de rango presentan una superficie geométrica similar a la de una semi-esfera o curvatura Gaussiana. Con lo cual nos permite, a través de un suavizado por medio de características de textura generar una superficie plana. A partir de esta información, de manera similar a los trabajos propuestos por Frome *et al.* en [40] y Zhou *et al.* en [166], se lleva a cabo una selección o muestreo de la información que se obtiene de este proceso denso. Lo que nos permite formar una matriz con información dispersa la cual contiene la información correspondiente a los descriptores de región y superficie.

#### 1.5.4. Imágenes Digitales

Una imagen es definida como una función bidimensional  $f(s, t)$ , donde  $s$  y  $t$  son componentes en un espacio continuo. El transformar esta función en  $f(x, y)$  a través de un proceso de *muestreo* (digitalización de las coordenadas) y *cuantización* (digitalización de la amplitud) permite obtener una **imagen digital**. Por lo tanto, los valores de  $f(x, y)$  son discretos y finitos, tal como se muestra en el diagrama ejemplificativo de la Figura 1.4.

La imagen digital es representada como una matriz de dimensiones  $M \times N$ , donde  $(x, y)$  son sus correspondientes coordenadas discretas, tal que,  $x = 0, 1, 2, \dots, M - 1$  corresponde a los renglones y  $y = 0, 1, 2, \dots, N - 1$  corresponde a las columnas. Una forma de representar dicha matriz es la siguiente,

$$A = \begin{pmatrix} a_{[0,0]} & a_{[0,1]} & \cdots & a_{[0,N]} \\ \vdots & \vdots & \cdots & \vdots \\ a_{[M,0]} & a_{[M,1]} & \cdots & a_{[M,N]} \end{pmatrix}$$

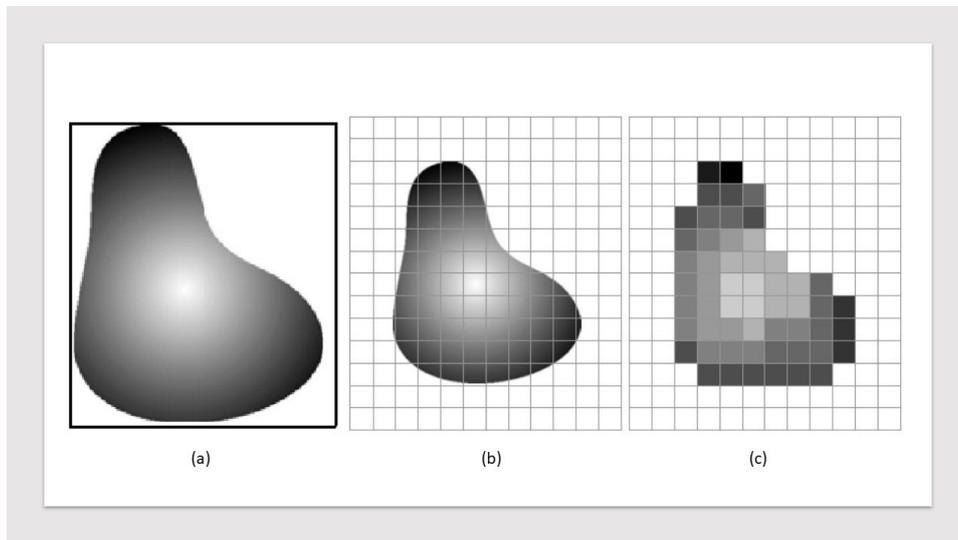


Figura 1.4: Diagrama ilustrativo de una a) imagen bidimensional  $f(s, t)$  la cual a través de un proceso de b) muestreo y cuantización, se obtiene como resultado una c) imagen digital  $f(x, y)$  [46].

Por otro lado, de acuerdo con González y Woods en [46] la sección del plano real que se encuentra definido por las coordenadas de una imagen es llamado “dominio espacial”, donde  $x$  y  $y$  son nombradas “variables espaciales” o “coordenadas espaciales”. Los valores contenidos en cada elemento de la matriz  $A$  son conocidos como *elemento de la imagen*, *elemento de la pintura*, *píxel* o *pel* (*picture element*, en inglés).

Si la imagen que se adquiere es bidimensional la representación de la imagen se conoce como imagen en *niveles* o *intensidades de gris*. Por lo tanto, en el proceso de adquisición de imagen, los píxeles pueden tomar valores entre  $0, \dots, 2^n - 1$ . Si se considera una codificación del píxel donde  $n = 8 \text{ bits}$ , entonces el nivel máximo de intensidad en la escala de grises será representado por el valor de 255 (ver imagen, 1.5). Al ser el *píxel* la unidad básica en una imagen implica que, no hay un elemento de menor dimensión. Por ejemplo, una imagen con una resolución de  $500 \times 300$  significa que dicha imagen está conformada por una matriz de 150,000 píxeles.

En una imagen de color los píxeles son representados normalmente en espacios de color. El más utilizado es el denominado **RGB** (Red, Green and Blue, en inglés), compuesto por un valor para el *componente* o *canal* rojo, uno para el verde y otro para el azul. Esto implica que la matriz que los representa es una matriz tridimensional (ver imagen ilustrativa en 1.6). El procesamiento de la misma es un problema más

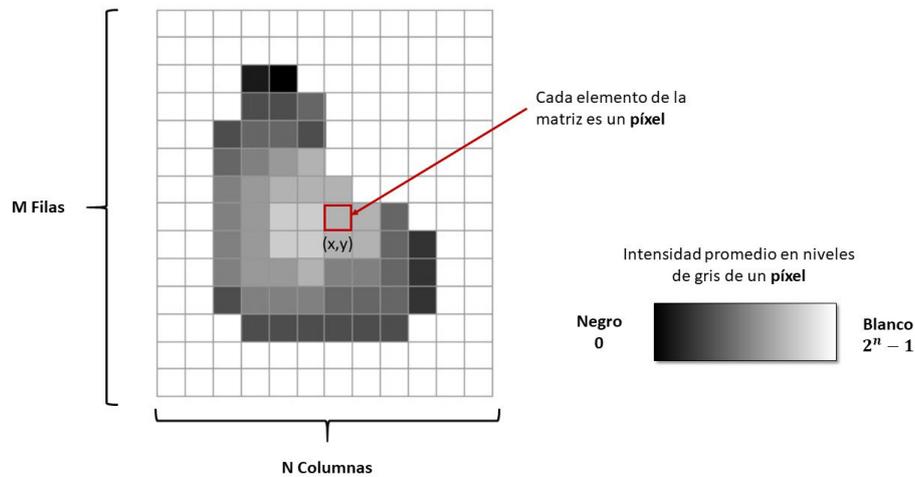


Figura 1.5: Diagrama ilustrativo de una imagen digital [46] ubicando en la coordenada espacial  $(x, y)$  un píxel; tal que,  $f(x, y)$  es igual a un nivel de intensidad en la escala de grises, el cual corresponde a dicho píxel.

a resolver, ya que las señales de color no son homogéneas; esto es, en la posición  $(x, y)$  correspondiente a cada canal el valor del píxel puede cambiar. Esto contrasta con las imágenes en escala de grises, donde los bits que representan esta escala son todos del mismo tipo y toman el valor de un número que representa la intensidad de píxeles; por lo tanto, se pueden procesar como una sola entidad o matriz. En cualquier representación, cuando los valores son normalizados, entonces la escala será entre  $0, \dots, 1$ ; donde 0 es el valor más cercano al negro (0) y 1 corresponderá a una intensidad máxima (255).

El procesamiento y análisis de imágenes en escala de grises comprende áreas tan diversas como lo es medicina, ingeniería, industria, física, biología, arte, deportes, entretenimiento o comerciales, por mencionar algunas de estas. De nuestro interés particular, para efectos de esta Tesis son las imágenes médicas en especial, las mamografías en formato DICOM; así como, las imágenes de rango (profundidad o disparidad), las cuales son procesadas para aplicaciones en robótica, que pueden ser desde la industrial hasta la de servicio [30, 22, 135].

### ***Imágenes DICOM***

En el campo de la medicina, la *imagen médica* procede de un conjunto de técnicas y procesos usados para crear imágenes de un cuerpo o de partes de este, tal como lo define González en [46]. Por lo tanto, una

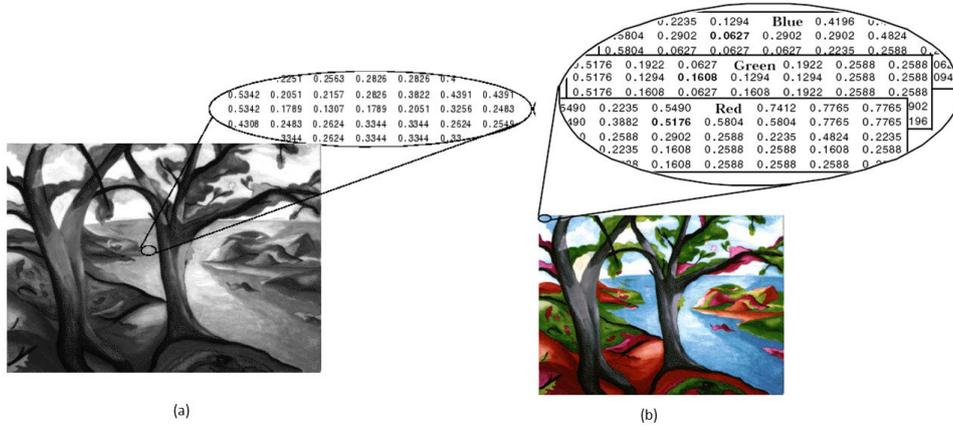


Figura 1.6: Imagen en a) nivel de gris y b) a color [97]. Los valores que se muestran en las matrices se encuentran normalizados.

imagen médica carece de sentido sino incluye toda la información necesaria del paciente, para que el especialista la pueda interpretar. Aun cuando pueden ser diversos los equipos médicos, que deberán interactuar entre sí, con el fin de llevar a cabo la adquisición, almacenamiento, distribución, procesamiento, etc., lo que implica que el trabajar con este tipo de imágenes se torne también un reto de comunicación.

La solución a este problema se le planteó al ACR (American College of Radiology, por sus siglas en inglés) y NEMA (National Electrical Manufacturers Association, por sus siglas en inglés), que en 1983 formaron un comité cuya misión era desarrollar un estándar para satisfacer todas estas necesidades de comunicación. En 1993 veía la luz la primera versión 3.0 del estándar DICOM (Digital Imaging and Communications in Medicine, por sus siglas en inglés) que ha sido el primer estándar instaurado en informática médica y que, actualmente tiene el soporte de la mayoría de los fabricantes de equipos médicos [26].

DICOM por tanto, es un estándar aplicable a una gran variedad de sistemas médicos que permite producir, almacenar, visualizar, procesar, enviar, recuperar, buscar e imprimir imágenes médicas y documentos estructurados derivados, así como el manejo de volumen de trabajo. DICOM específicamente según Serna [130] describe: i) el contenido de la información, incluyendo la estructura y codificación; ii) servicios DICOM para la administración de la información; y iii) protocolo de mensajería. Por ello, el incremento en el uso de sistemas PACS (Picture

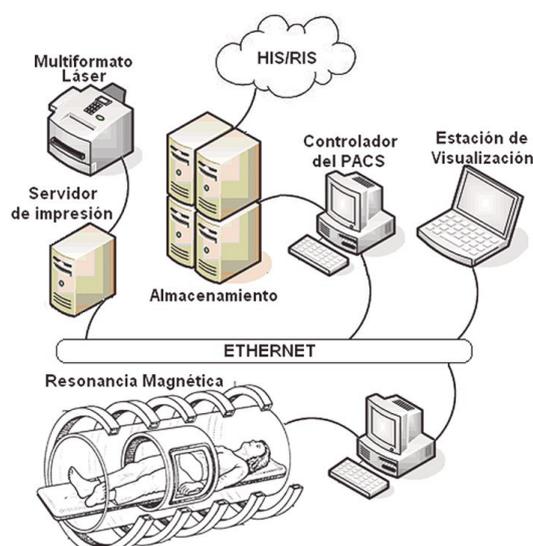


Figura 1.7: Diagrama ilustrativo de una arquitectura básica PACS [26].

Archiving and Communication System, por sus siglas en inglés) permite proveer una forma de servicio informático en el cual, los equipos de diagnóstico por imágenes, equipos de visualización, base de datos y sistemas de almacenamiento establezcan comunicación a través de la red Ethernet a la que están conectados, tal como se muestra en la Figura 1.7.

En la actualidad, estos sistemas son ampliamente usado en diversas especialidades médicas como lo es radiología, cardiología, oftalmología, endoscopia, patología, dermatología y oncología. En esta última, es de nuestro particular interés el procesamiento de mamografías. La *mamografía* es la radiografía proyectiva (exposición de un objeto a radiaciones electromagnéticas de alta energía) correspondiente a las mamas y otros tejidos blandos. Su uso fundamental es en el diagnóstico del cáncer de mama, aunque puede tener otros usos, tales como la localización de tejidos sospechosos antes de realizar una biopsia. Un ejemplo, de una mamografía en formato DICOM se muestra en la Figura 1.8.

### ***Detección de Microcalcificaciones***

La detección de calcificaciones en mamografías es un proceso lento y con un alto grado de error, esto debido a que aún se lleva a cabo mediante inspección visual realizada por el médico especialista. La precisión está directamente relacionada con la experiencia, la capacidad visual y la fatiga. Dos de estos factores son temporales, generando problemas de estabilidad en la detección e incluso, pueden ser facto-

res decisivos que afecten la sensibilidad del proceso. Como resultado, existen fallas en la detección del cáncer de mama, debido a interpretaciones incorrectas, lo que trae como consecuencia un mayor número de casos situados como falsos positivos. Otro aspecto a considerar, es la exactitud con la que el radiólogo considera la importancia médica del tamaño de las regiones con calcificación. Ya que, en la mayoría de los casos desprecian o malinterpretan la misma [90]. Esto es debido a que en la generalidad, el tamaño de las microcalcificaciones oscilan en el rango de  $[0.1 - 1]mm$  [64] por ende, no son fáciles de observar. En un estudio realizado por Cheung *et al.* en [24], validan la existencia de una correlación entre las dimensiones de microcalcificaciones sospechosas y los resultados de patología.

Por otro lado, diversos sistemas de diagnóstico asistido por computadora (CAD system, abreviatura en inglés) ayudan al procesamiento de mamografías; para que provean resultados más exactos en un menor tiempo. Dichos sistemas tienen un amplio significado en los avances tecnológicos que apunta a detectar, segmentar y clasificar agrupamiento de microcalcificaciones en mamografías digitales. Por lo tanto, los sistemas CAD tienden a ser usados clínicamente desde hace más de dos décadas tal como lo presentan estudios recientes realizados por Cheung *et al.* ([24] así como, Van *et al.* [151]). Quienes consideran este tipo de sistemas como un segundo lector en el diagnóstico llevado a cabo por los radiólogos. Este creciente uso, es debido a la sensibilidad en la detección ya que, sin un sistema CAD la detección certera corresponde a un 80% mientras que, con CAD se incrementa en un 90% de acuerdo con lo reportado por Cheng *et al.* en [23].

La importancia de este tipo de sistemas CAD radica en la ayuda que proporciona para localizar alteraciones morfológicas e inferir la presencia o ausencia de anormalidades, principalmente, pequeñas calcificaciones. Estas, aparecen como pequeñas manchas blancas en las mamografías. Por lo que, se identifican dos tipos de calcificaciones, las primeras se conocen como microcalcificaciones y las segundas como macrocalcificaciones. En este sentido, las macrocalcificaciones son amplias, granuladas y generalmente, son asociadas con la edad. Mientras que, las microcalcificaciones son vinculadas regularmente, como signos tempranos de cáncer de mama, sea que estos presenten o no, una masa visible. Por lo tanto, la relación de las microcalcificaciones con diversos tipos de cáncer se establece de acuerdo con, su morfología, distribución o ambas, en el tejido mamario como es presentado por Burnside *et al.* en [19]. En dicha referencia, se encuentra que, estas microcalcificaciones pueden ser asociadas con tumores benignos, lesiones atípicas (cystis, en inglés) o lesiones cancerosas.

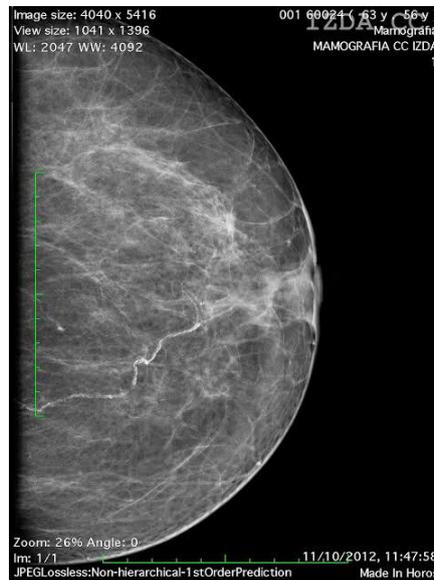


Figura 1.8: Ejemplo ilustrativo de una mamografía en formato DICOM. Nótese la información médica contenida en las cuatro esquinas de la imagen (*imagen cortesía del Hospital General de Irapuato*).

Es importante hacer notar que, las imágenes obtenidas con formato DICOM son de bajo contraste, de acuerdo con su procedimiento de adquisición. Esto implica que el procesamiento de estas se convierta en una tarea desafiante. Por otro lado, es cierto que existen imágenes de mamografías adquiridas con mejor calidad, como lo son por ejemplo, las que se obtienen por medio de resonancia magnética. Sin embargo, se ha comprobado que, para el diagnóstico de lesiones malignas que aparecen como microcalcificaciones, el uso de estas no es recomendado actualmente. Este hecho se fundamenta en un estudio realizado por Bennani en [14], quien presenta una revisión sistemática y un meta análisis del diagnóstico de este tipo lesiones que se manifiestan como microcalcificaciones en las mamografías.

Algunos de los métodos propuestos en trabajos previos para segmentar diversos tipos de microcalcificaciones de acuerdo con su naturaleza; tales como, técnicas clásicas, difusas, substracción de la imagen bilateral y múltiple escala son revisados por Cheng *et al.* [23] así como, el último estudio realizado por Alam *et al.* en [4]. Propuestas recientes para llevar a cabo el mismo proceso de detección y segmentación, pero que hacen uso de técnicas como la de aprendizaje profundo (Deep learning, en inglés) presentada por Samala *et al.* [122] o la de contornos activos como lo muestra el trabajo de Wang *et al.* en [156], entre otros.

Por otra parte, Yadollahpour en [5] presenta una revisión de varios métodos considerando análisis de textura para la detección de masas y microcalcificaciones usados en la detección temprana del cáncer de mama. Samala en [123] presenta un estudio correspondiente a los avances y cambios realizados para la detección de microcalcificaciones en mamografías digitales y tomosíntesis digital del seno partiendo desde la perspectiva de un sistema CAD.

La transformada de Hough y un método basado en umbralización fueron considerados por Fanizzi *et al.* en [38] técnicas con las que, es posible evidenciar la potencialidad de agrupar simples microcalcificaciones, no como una “sucesión de eventos”, sino como un agrupamiento usando un conjunto de reglas expertas de codificación; con las cuales es posible obtener los resultados correspondientes a dichas regiones, en donde se encuentran las lesiones de interés. Filtros de alta frecuencia fueron implementados por Lauria *et al.* [80] como un preprocesamiento en orden para llevar a cabo la segmentación de posibles áreas sospechosas en la mamografía. El análisis de la microcalcificación apunta actualmente, a encontrar y clasificar regiones de interés (ROI, por sus siglas en inglés) por medio de dos redes neuronales. La primera es una red neuronal hacia delante (feedforward network, en inglés) mientras que la segunda red neuronal utiliza el método de componente principal (PCA, por sus siglas en inglés) finalizando así el proceso de clasificación.

Por otro lado, un estudio cuantitativo de acuerdo con diversas propuestas usando análisis de textura en sistemas CAD es presentado Ali y Hamed en [5] y Wan *et al.* en [157]. Kim en [71] compara la región circundante a través de otro análisis convencional de textura con respecto de la detección de agrupamiento de microcalcificaciones en mamografías digitalizadas. Los resultados de estimación de dichas clasificaciones son usualmente evaluados utilizando la curva ROC (Receiver Operating Curve, en inglés), La cual describe la capacidad para discriminar dichas aproximaciones tal como lo presenta Dheeba *et al.* en [33] haciendo uso de algoritmos heurísticos. Jalalian en [63] obtiene características estadísticas de textura basada en la matriz de coocurrencia a partir de la segmentación de un volumen de interés. Los estados que se presentan de los resultados obtenidos, son clasificados por medio de una red neuronal multicapa activa conocida como perceptrón, la cual presenta una alta exactitud en las clasificaciones realizadas.

### ***Imágenes de Rango***

Las imágenes de rango son representaciones 2D de escenas 3D; por lo que, su principal ventaja es la información geométrica contenida

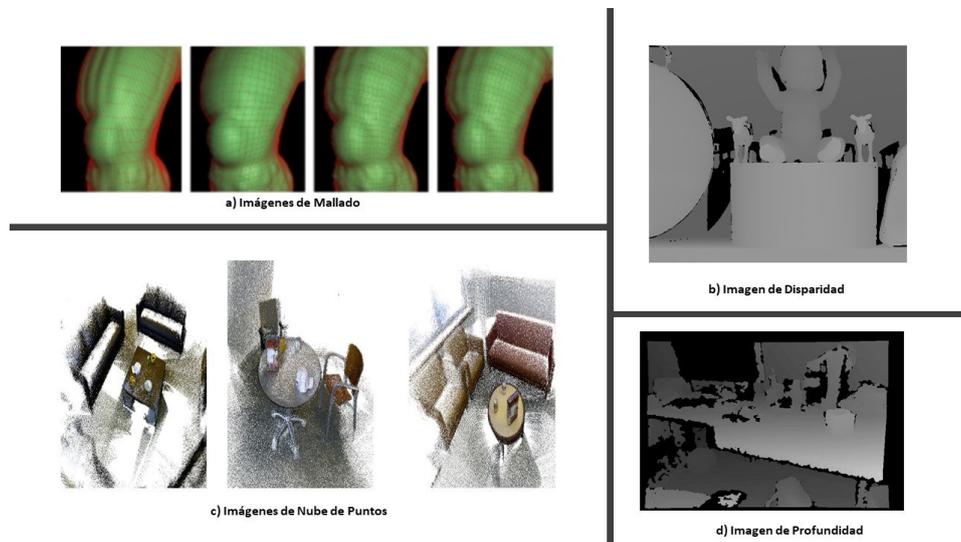


Figura 1.9: Ejemplo de imágenes de a) mallado [95]; b) disparidad [58]; c) nube de puntos [144] y d) profundidad [31].

en la imagen. Esto implica que, cada valor entero de un píxel en una imagen de rango es la representación de la distancia más cercana entre los objetos contenidos en una escena “vistos” desde un punto en el espacio (ver ejemplo, Figura 1.10). Por lo que, la posición de cada píxel determina la dirección del haz de incidencia sobre cada elemento de la escena. Por ejemplo, en un escaneo 3D la información adquirida es capturada desde un punto fijo en el espacio. Esto es, el sensor no se mueve mientras que los puntos 3D son generados.

En el procesamiento de este tipo de imágenes es bastante razonable, transferir conceptos de rango basados en percepción 3D. Por lo tanto, una imagen de rango puede ser calculada eficientemente a partir de una nube de puntos usando por ejemplo, *z-buffer* [137, 136], en una amplitud tan larga como la posición original conocida del sensor. Algunas otras técnicas para la formación de imágenes de rango partiendo de mapas de disparidad generados por sistemas estéreo visión, los podemos encontrar en [58] y [112]; o bien trabajos recientes respecto a la generación de mapas de profundidad haciendo uso del sensor Kinect son investigados en [135] y [51].

De acuerdo con lo expuesto por Guo *et al.* [47], una imagen de rango puede ser representada de tres formas distintas. Esto es, como una imagen de profundidad (o disparidad), una nube de puntos o un mallado poligonal, como se muestran en la Figura 1.9. Por lo anterior, en este trabajo de Tesis se considera como imagen de rango, a toda imagen de

profundidad o disparidad.

### 1.5.5. Sensor Kinect

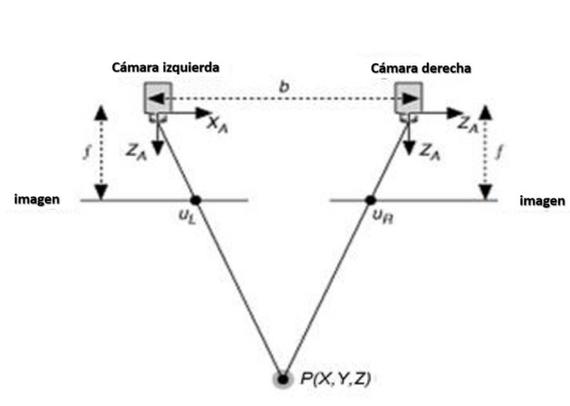


Figura 1.10: Formación de las imágenes desde un sistema estéreo[27]

El ser humano está habilitado para llevar a cabo la adquisición de imágenes por medio de la visión binocular (estereoscópica) que tiene el sentido de la vista, con el que, además podemos tener una percepción de la *profundidad*.

Una emulación de la visión humana lo podemos encontrar en un sistema estéreo, como el presentado en la Figura 1.10. Este se conforma generalmente por un par de cámaras monoculares; ya que, a través de una segunda perspectiva se permite la localización 3D de los puntos en el mundo real. Existen diversas técnicas [32] para la estimación de la posición de los puntos dentro de las dos imágenes planas, correspondientes a cada cámara. Esto es, porque la información obtenida desde las dos imágenes implica una fuerte aplicación de la trigonometría. Cada cámara captura la posición lateral de los objetos bajo dos perspectivas diferentes lo que, por ende, ocasiona un paralaje debido a la distancia existente entre dichas cámaras. Sin embargo, el computar la fusión de ambas imágenes es una tarea bastante ardua, la cual conlleva la correspondencia de los puntos localizados en estas. En algunos casos la alternativa será localizar puntos seleccionados en las respectivas imágenes, en los que se encuentran bordes o límites relevantes que permitan una correspondencia más efectiva [119].

Las cámaras llamadas **RGB-D** tales como el sensor Microsoft Kinect [103] (ver Figura 1.11) y la versión libre del Kinect conocida como

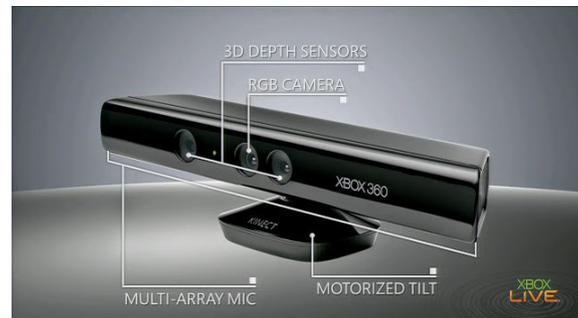


Figura 1.11: Sensor Kinect Xbox360 [103]

Asus Xtion PRO LIVE [117] proveen información al mismo tiempo, en el espacio de color RGB y de profundidad (D). Esta última se adquiere ya que, el sensor Kinect posee un proyector laser (IR projector, en inglés) y una cámara infrarroja (IR camera, en inglés); con una resolución de  $1290 \times 960$  píxeles, que funciona a 30 cuadros por segundo (fps, por sus siglas en inglés). Tiene además, una cámara **RGB** con una resolución de  $640 \times 480$  píxeles (VGA) que funciona a  $30\text{fps}$ . El dispositivo en general, posee un ángulo de visión de  $58^\circ$  horizontalmente y  $45^\circ$  verticalmente y  $70^\circ$  sobre la diagonal; pero, se puede incrementar el vertical a  $\pm 27^\circ$ , por medio de su base móvil (ver Figura 1.11). Según las especificaciones del dispositivo, este es capaz de detectar profundidades en un rango de distancia al sensor desde los  $0.7\text{m}$  hasta los  $6\text{m}$ . La mínima distancia que se puede utilizar del sensor es aproximadamente  $0.7\text{m}$  dando lugar a un campo visual de  $0.87\text{m}$  en el sentido horizontal y  $0.63\text{m}$  en el sentido vertical. Teniendo en cuenta las características de la cámara se obtiene una resolución de  $1.3\text{mm}$  por píxel.

### ***Cámara RGB***

El funcionamiento de la cámara RGB contenida en el sensor Kinect es similar al de cualquier cámara digital, por lo que, su sensor captura colores utilizando un filtro de Bayer. El problema del sensor es que no distingue los colores sino sólo variaciones de intensidad; así, para obtener una imagen en color es necesario descomponer la imagen en los primarios (rojo, verde y azul). Por tanto, la abreviación **RGB** (Red, Green y Blue, por sus siglas en inglés) corresponde a los tres canales de color usados para estimar la información correspondiente para cada píxel que forma la imagen. Por cada uno de estos canales usualmente se almacena un valor de 8 bits. La combinación de todas estas representaciones permiten estimar valores en el espacio de color RGB.

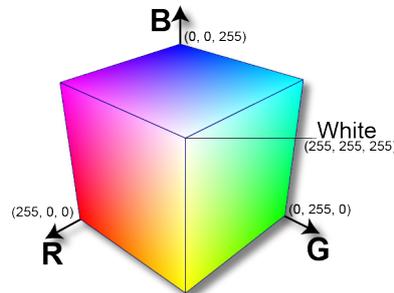


Figura 1.12: Espacio de color RGB [105]

Una representación de dicho espacio la podemos ver en la Figura 1.12. En este se considera que, si todos los canales de color para un píxel en particular se encuentran en los valores más bajos, entonces el color resultante será negro. En el caso contrario si todos los valores de color son los más altos, entonces el color resultante será blanco.

### ***Sensor de Profundidad***

El sensor de profundidad está formado por dos componentes: un proyector de luz infrarroja (IR) y un sensor CMOS monocromático estándar, como el de cualquier cámara común. El sensor de la cámara IR opera a  $30\text{ Hz}$  capturando imágenes de  $1200 \times 960$  píxeles. Estas imágenes se “reducen” a  $640 \times 480$  píxeles representados con 11 bits. Esta representación proporciona 2048 niveles de sensibilidad. Dicha sensibilidad se pierde, debido a que, el sensor Kinect transmite su información a través de un puerto USB por lo que, las imágenes terminan transmitiéndose solamente con 8 bits.

Los componentes del sensor de profundidad se encuentran alineados sobre el eje horizontal del dispositivo, a una distancia de  $75\text{ mm}$  (denominada línea base), con ejes ópticos paralelos. Dicha configuración, permite llevar a cabo los cálculos de profundidad, los cuales se basan en un principio similar al de la triangulación activada entre emisor y cámara. La técnica implementada se conoce como Codificación de Luz (Light Coding, en inglés) [42], la cual consiste en dos fases, un proceso de calibración y otro de funcionamiento. El error generado usando esta técnica es menor a los  $10\text{ cm}$  en distancias superiores a los  $4\text{ m}$  y menor a los  $2\text{ cm}$  para distancias inferiores a los  $2.5\text{ m}$ .

Así, la información de profundidad hace referencia a la distancia entre el sensor y regiones particulares de la escena. Para ello, el proceso de la detección de profundidad utiliza un método de luz estructurada para “medir” la profundidad. Esto es, se proyecta un patrón conocido

como “de puntos o motas” generado a partir de un conjunto de rejillas de difracción colocadas enfrente del sensor IR. Estos puntos son capturados por la cámara IR y luego comparados con el patrón conocido (almacenado en el chip interno del sensor Kinect).

Por otro lado, se sabe que cualquier perturbación capturada es una variación en la superficie la que se puede identificar como más cercana o lejana al sensor de la cámara IR. En este caso, por cada píxel en la imagen IR se utiliza una pequeña ventana de correlación ( $9 \times 9$  o  $9 \times 7$ ); esto con el fin de comparar el píxel capturado con el patrón contenido en el sensor en esa posición así como, los 64 píxeles que son vecinos, considerando la ventana horizontal. La mejor coincidencia ofrece un desplazamiento de la profundidad conocida, en términos de píxeles, como *disparidad*.

El dispositivo Kinect realiza una interpolación adicional, con respecto de la mejor coincidencia para obtener una precisión de subpíxel, esto es, de  $1/8$  de píxel. Dada la profundidad conocida del plano-patrón y la disparidad, la profundidad puede ser estimada para cada píxel calculado, por triangulación. Cabe hacer notar que, la posición de un píxel en la imagen RGB siempre tendrá una referencia diferente, a la que se tenga en la imagen de profundidad. Esto es, porque existe una diferencia en la posición de estas dos cámaras; la cual es, una separación horizontal de  $2.5\text{ cm}$ . Para corregir esta diferencia, es necesario llevar a cabo un proceso de calibración entre las cámaras por medio de herramientas como openNI o Matlab, entre otras para realizar la correspondiente corrección. La Figura 1.13 presenta un ejemplo del registro de imágenes obtenidas por medio del sensor Kinect.

### ***Imágenes RGB-D***

Las imágenes que se obtienen de sensores 3D de bajo costo como lo es Kinect [103] nos permiten obtener al mismo tiempo, información de naturaleza diferente pero complementaria. Esto es, se combina información geométrica (*profundidad-Depth*, en inglés) con la de color RGB (ver ejemplo ilustrativo en 1.14). Los canales de color como se mencionó son representados por medio de una matriz de enteros de  $8\text{ bits}$ . Mientras que, los datos de profundidad pueden ser representados por medio de una matriz de enteros de  $16\text{ bits}$ ; aún cuando, el sensor transfiere los datos como enteros de  $11\text{ bits}$ , para ahorrar espacio. Kinect cuantifica la medición de profundidad en un rango de 1 a 10,000 valores, que representa la distancia de profundidad en milímetros. En este caso, la información del canal de profundidad es representada como una imagen monocromática tal como es presentado por Cruz *et al.* en [30].

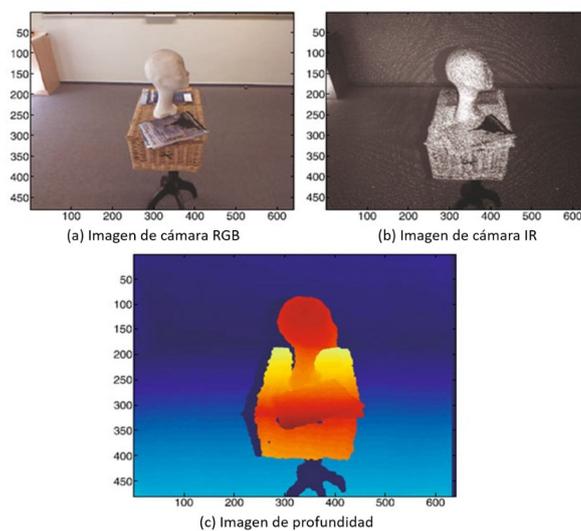


Figura 1.13: Ejemplo de las imágenes de salida del sensor Kinect [135]

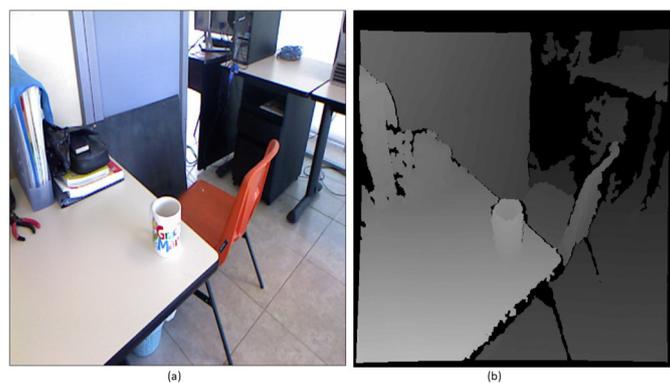


Figura 1.14: Ejemplo de imagen obtenida del sensor Kinect para la base de datos DSPLab. a) Imagen a color en formato VGA y b) representación en imagen monocromática de la información adquirida por el sensor de profundidad.

### 1.5.6. Puntos de Interés y Descriptores

Davies en su libro *Machine Vision* [32] presenta, cómo la detección de *puntos salientes* obtenidos principalmente de los detectores de esquinas, o puntos con patrones de intensidad similares pueden ser considerados como *puntos característicos* o *puntos de interés*. Aún cuando, no siempre como lo comenta Davis, un punto saliente o característico es un punto de interés. El mismo autor, retoma la definición de punto de interés propuesto por Haralick y Shapiro en [54], quienes consideran que estos puntos deberán de presentar ciertas características de forma simultánea, tal como, distintividad e invariancia. Lo que implica que, estos deberán ser únicos e invariantes a las distorsiones geométricas. Como tal, se puede mostrar la existencia de cierta *invarianza* a los cambios moderados en la escala o rotación. Adicionalmente, Haralick y Shapiro presentan una serie de propiedades deseables que estos puntos deben de tener tal como estabilidad, singularidad e interpretabilidad [54]; con las cuales se establece una correlación confiable.

Por otro lado, un objeto es posible representarlo por medio de un conjunto de puntos característicos y sus descriptores. La *representación* es una descripción matemática del objeto. Esta, puede ser a través de sus características externas, como lo es el *contorno* o de sus características internas, lo que se conoce por *región* [46]. Generalmente, se elige una representación externa cuando el objetivo principal se centra en las características de forma. Mientras que, una representación interna se lleva a cabo cuando el principal interés se centra en las propiedades de reflectividad, tales como color y textura. En cualquier caso, las características seleccionadas como descriptores deberían ser tan insensibles como fuera posible a cambios de tamaño, traslación y rotación.

Así, una *característica* puede estar asociada a propiedades globales o locales del objeto. Las *características globales* describen a un objeto como una sola variable o vector; por ejemplo, un histograma de color. La principal ventaja de usarlas es la eficiencia computacional, ya que no hay necesidad de analizar el objeto en secciones. Por otro lado, las *características locales* describen a un objeto a través de múltiples descriptores asociados a puntos característicos. Por tanto, un *descriptor local* codifica información “local” contenida en la vecindad o región del punto característico; tal que, un objeto sea representado a través de un conjunto de descriptores.

En el presente trabajo, nuestra propuesta considera la extracción de características de región tal como lo son las geométricas y de textura. Esta última (textura), es calculada a través de un proceso denso.

### 1.5.7. Detección y Segmentación de Objetos 3D

Un proceso fundamental en la detección de objetos involucra, el cómputo de características 3D; el cual, a su vez, debe de estar ligado a la detección de puntos de interés. Por lo tanto, un extractor local de características de acuerdo con lo citado por Tombari en [147] es una fase esencial en este propósito. En [48] Guo *et al.* proponen que, todo *descriptor* debe ser compacto, robusto y descriptivo ante un conjunto de información que se considera como ruido.

En la Figura 1.15 se muestran algunos ejemplos de los diferentes tipos de detectores-descriptores de puntos de interés. Por ejemplo, BRAND (Binary Robust Appearance and Normal Descriptor, en inglés) 1.15 a) aplicado a nube de puntos, muestra su robustez ante la rotación. Las técnicas RCS (Rotational Contour Signatures, en inglés) 1.15 b) y HGND (Histograms of Gaussian Normal Distribution, en inglés) 1.15 d) aplicados a imágenes de profundidad, muestran su invarianza ante el escalamiento y rotación. Mientras que, por otro lado, técnicas como el clásico SIFT (Scale Invariant Feature Transformation, en inglés) 1.15 c) aplicado a imágenes en niveles de gris, muestra invarianza a rotación, escala e iluminación.

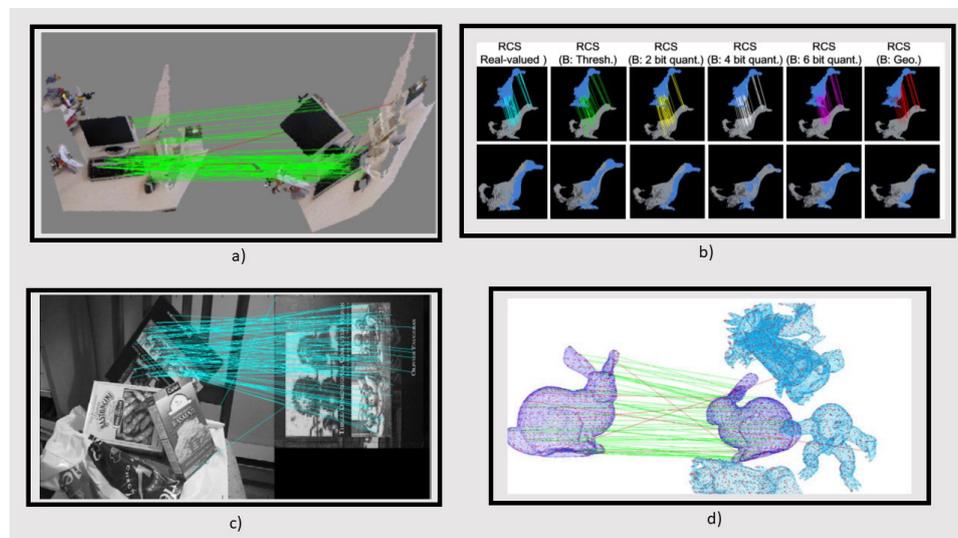


Figura 1.15: Ejemplo de correspondencias bajo diferentes tipos de varianzas (ruido). Técnicas como (a) BRAND [15], b)RCS [161], SIFT [88] y d) HGND [167] presentan robutez ante la rotación, escalamiento o iluminación.

De manera general, es posible identificar a lo largo de las últimas décadas, una referencia cronológica de algunos de los detectores y des-

criptores locales más destacados.

### ***Descriptores Locales 3D***

Es importante hacer notar que, cuando los puntos de interés han sido detectados se tenga también, los descriptores correspondientes a la superficie circundante a cada punto. De acuerdo con el trabajo presentado por Guo *et al.* [47], se tienen tres tipos de clasificaciones: firma (signature based, en inglés), histograma (histogram based, en inglés) y transformaciones basadas en métodos (transform based, en inglés).

Steder *et al.* desarrollan el descriptor Alineación de Normales de Características Radiales NARF (Normal Aligned Radial Features, en inglés) [136, 137] aplicado a nube de puntos. Los autores presentan cómo a través de alinear vecindades (llamadas “parches”) a lo largo de la normal correspondiente al punto de interés, es posible posteriormente, realinear con respecto de un patrón inicial. El fijar este a una única orientación, permite que exista una invarianza con respecto a la rotación. Un trabajo similar, pero desarrollado para imágenes RGB-D es presentado por do Nascimento *et al.* en [107]. En particular ellos proponen fusionar la información geométrica y de intensidad en un parche local para codificar las variaciones locales y desplazamientos de la superficie normal. Esto es llevado a cabo, dentro de un descriptor binario robusto a, apariencia y vectores normales; el cual es nombrado como BRAND. Este descriptor presenta ventajas sobre otras técnicas clásicas, como lo son SIFT, SURF, spin image, and CSHOT en términos de, precisión en la coincidencia y robustez.

Por otro lado, Tombari *et al.* presentan en [145] una versión 3D (USC unique shape context, en inglés), con respecto de la técnica contexto de forma (shape context, en inglés) presentada por Belongie *et al.* en [12]. En [145] Tombari propone que, además de considerar una vecindad esférica, se asocie cada punto de interés a una determinada repetibilidad. Lo que permita un inequívoco LRF y más aún, se añada invarianza ante transformaciones rígidas. La técnica USC presenta mejoras en el decremento de memoria, así como, el incremento en la exactitud de los puntos puestos en coincidencia; en comparación con el 3DSC (3D shape context, en inglés) propuesto en [40].

Los métodos **basados en transformaciones** son aquellos en los que, primero se convierte la información contenida en una imagen de rango, teniendo como base un dominio espacial, en otro dominio, por ejemplo, frecuencia. Esto, con el fin de representar la vecindad de un punto determinado, de la superficie definida en el dominio de origen. Esto es, a través de un proceso de codificación, la información obtenida de la imagen es transformada a un nuevo dominio. Por ejemplo,

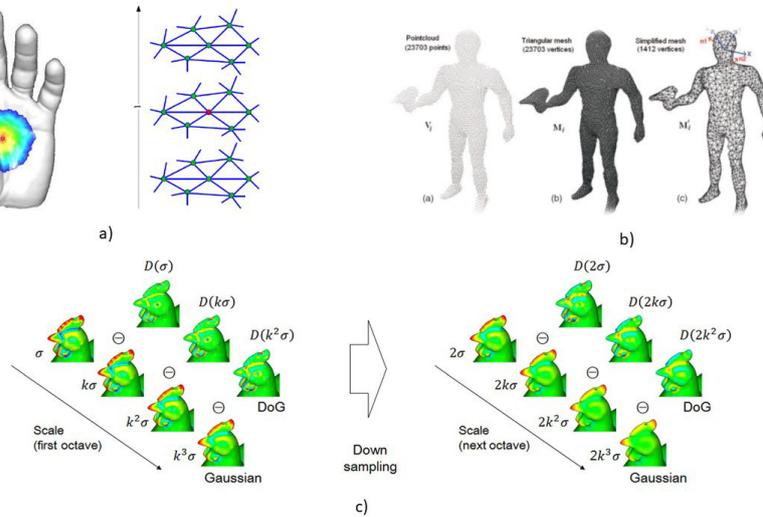


Figura 1.16: Ejemplo de algunos descriptores mencionados en 1.5.7. a) Geometric Scale Space [170]; b) Local Reference Frame [102]; y c) Scale Invariant Features [83].

en [59] se lleva a cabo una transformación de una superficie local, por medio de la transformada de Laplace-Beltrami obteniéndose valores espectrales, con los cuales es generado el histograma. Otra propuesta la encontramos en [140]; la cual considera una transformación de un Manifold de Riemannian restringido por *heat kernel* llamado HKS (Heat Kernel Signature, en inglés). Algunas extensiones de este descriptor las encontramos en [18, 15]. Finalmente, Knopp *et al.* desarrollan en [72] una extensión de SURF llamada 3D SURF. Estos métodos presentan un buen desempeño ante transformaciones rígidas, deformaciones isométricas y escalado.

### ***Detectores y Descriptores Locales***

Las aproximaciones globales de detectores y descriptores pueden extraer información característica teniendo como referencia la superficie geométrica total del objeto, o el modelo de este, por lo que estos métodos son más usados en la recuperación de formas así como, en clasificación. Algunos trabajos que presentan un análisis de dichas propuestas las encontramos en Bustos *et al.* Brostein y Marton *et al.* [20, 17, 96], respectivamente.

Por otro lado, las aproximaciones locales como se revisó previamente extraen y describen regiones específicas alrededor de puntos determinados. Trabajos recientes que recopilan la descripción y evaluación de

estas técnicas en algunas de las áreas de aplicación más sobresalientes, las podemos encontrar en Bronstein *et al.* [16], Guo *et al.* [49], Gao *et al.* [41], Bennamoun *et al.* [13]. Cada uno de estos trabajos destaca el desarrollar primero un detector para que, con los resultados obtenidos del mismo sea el insumo para el descriptor. Wu *et al.* [158] lleva a cabo un exhaustivo análisis y evaluación de detectores y descriptores locales.

Por lo anterior expuesto en las subsecciones, se dice que, los métodos locales son más eficientes y robustos que los globales para la detección, segmentación o estimación de la pose, como lo presentan Guo *et al.* en [48] y Wu *et al.* [158].

Por lo tanto, la propuesta del detector y descriptor que se lleva a cabo en el presente trabajo de Tesis es desarrollada con respecto a métodos locales. Tal que, primero se lleva a cabo un proceso de detección de un conjunto de puntos distintivos los cuales se busca sean invariantes a transformaciones de escala o rotación. Entonces, el descriptor codifica la información más relevante de la vecindad de dichos puntos, lo que formará el vector característico, el cual permitirá que se lleve a cabo una detección de puntos clave que permitan definir el contorno de los objetos contenidos en la escena, una región de interés en particular y finalmente el objeto contenido en esta.

### 1.5.8. Textura

Nuestro mundo es visualmente rico en una enorme variedad de texturas, por ende, es una característica fundamental en un amplio rango de imágenes, que van desde una vista satelital hasta microscópicas. Esto es debido a que, que la textura visual es la *apariencia* que tiene una superficie con una textura táctil propiamente dicho. Es por ello, que el estudio de la textura como elemento visual posee al mismo tiempo, características ópticas y táctiles, lo que la convierte en un tópico de interés en diversas áreas, destacándose los trabajos desarrollados en visión por computadora.

Tuceryan [1] presenta algunos ejemplos de una compilación de diferentes definiciones, interpretaciones y conceptos en relación con la textura, que son considerados básicos en la literatura de visión por computadora. Para efectos de esta Tesis se retoma una de las definiciones descritas por Tuceryan, la cual nos indica que, *“una imagen tiene una textura constante si un conjunto de estadísticas u otras propiedades locales de la función de imagen son constantes, varían lentamente o son aproximadamente periódicas”*.

Es importante destacar con respecto del cómputo realizado para analizar la textura de una imagen, en la generalidad, este es orientado

para establecer una metodología. A través de esta, se define el rango en el que se establece la diferencia entre un máximo o un mínimo sobre la vecindad de la región que se está examinando. Por ejemplo, tal como lo presenta Russ en [119] para una superficie plana o con una región uniforme, el rango es muy pequeño ya que la superficie “rugosa” es mínima. Por el contrario, valores amplios corresponderán a superficies con áreas de rugosidad muy amplias.

### ***Análisis de la Textura***

El problema del análisis de textura ha sido estudiado desde los años 60's, tal como se presenta en [69], implicando una evolución en los últimos 50 años. Este hecho permitió realizar una serie de categorizaciones a partir de los 90's tal como es presentado por Tuceryan en [1]. En dicho trabajo, se muestran varios métodos para extraer características de textura con las cuales se puedan resolver problemas relacionados al procesamiento de textura, tales como clasificación, segmentación, síntesis y forma de estas. A partir de esta década los progresos se han dado de manera significativa en el análisis de textura, de tal forma que, incluyen aplicaciones médicas [7, 141, 100] inspección de calidad [159, 2], recuperación de imagen basada en contenido [55], análisis de rostros [155, 134, 165], biométricos [45, 93], reconocimiento de objetos [133, 111, 164], síntesis de textura para compresión de imágenes [43], así como, visión robótica para navegación aérea y vehículos autónomos [6, 60, 37, 108, 92, 118].

El análisis de textura se implementó primero para imágenes bidimensionales, como lo expone Haralick en [53]. En su trabajo muestra como la extracción de características es llevada a cabo a través de una matriz de dependencia espacial, SDM (Spatial Dependence Matrix, en inglés). Básicamente, este método calcula las propiedades estadísticas partiendo de las interrelaciones espaciales de los píxeles a lo largo de la imagen. Aún cuando la técnica SDM fue introducida en 1973 existen trabajos recientes presentando una extensión de SDM, aplicado a imágenes 3D como lo muestran los trabajos de Mahmoud *et al.* [94] Kurani *et al.* [76] o Othmen *et al.* en [110]; cuyas propuestas incluyen la extracción de características de imágenes texturizadas 3D. Un ejemplo interesante también es el presentado por Sampson en [124]; quien calcula características texturales desde imágenes de rebanadas de manzanas secas. Yahia *et al.* en [160] usa SDM en combinación con 3D-LBP para llevar a cabo el estudio de reconocimiento de expresiones faciales; y finalmente Lloyd *et al.* en [87] utilizan esta técnica para la detección de actividades violentas o anormales en multitudes a través de videos. En particular, en este último trabajo se hace uso de la técnica SDM

aplicada a información contenida en las imágenes de profundidad.

Por otro lado, el descriptor Local Binary Patterns (LBP, por sus siglas en inglés) propuesto por Ojala en [109], el cual consiste en calcular un código binario para cada píxel de la imagen. Por lo tanto, con la técnica LBP es posible llegar a una distribución de la textura semejante a los histogramas en escala de grises, tal que existe una distribución de  $2^8 = 256$  posibles valores. Quizás la desventaja que presentan estas técnicas sea, que pierden el sentido de una región texturizada, para convertirse más en un descriptor global que local. Otro tipo de técnicas llamadas de proyección, se encuentran representadas en trabajos como wavelet descriptor presentado por Chuang *et al.* [25] o zernike moments trabajo realizado por Teague [143]. En estas técnicas se proyecta la imagen en otra básica, para entonces, así poder llevar a cabo el cálculo del descriptor. De igual forma estos métodos se encuentran reconocidos como de extracción y detección de características aplicados a imágenes 2D. Extensiones a 3D de estos métodos los podemos encontrar en una retrospectiva de dichas técnicas y sus derivaciones realizadas por Liu *et al.*, [85]; Ahmad *et al.*, en [3] y Raad *et al.* en [113]. Las diferencias en los estudios antes mencionados radica en los diferentes enfoques que se tienen para presentar la información respecto a la representación o síntesis de la textura, o bien, las diversas aplicaciones que se pueden tener en estas áreas de interés.

El objetivo principal de la clasificación de textura se encuentra encaminado al diseño de algoritmos, que permitan determinar si una región o imagen, pertenecen a una determinada clase de textura, tal como lo muestran los trabajos de Lazebnik *et al.* mostrando en [82], la representación de texturas dispersas; en ese sentido Liu y Fieguth muestran una técnica para la clasificación de texturas haciendo uso de características aleatorias[86]; mientras que, Varma y Zisserman en [153] muestran cómo clasificar textura partiendo del procesamiento de la información contenida en una sola imagen.

La síntesis de textura hace referencia a la generación de nuevas texturas teniendo como referencia las ya conocidas; tal como lo presentan los primeros trabajos desarrollados por Zhu *et al.* con su FRAME (Filters, random fields and maximum entropy, en inglés) [168]. En trabajos recientes Gatys *et al.* desarrollan una técnica usando redes neuronales convolutivas [43] con la finalidad de formar una representación de la imagen que sea comprensible al ser humano.

Se conoce como forma a partir de la textura (shape from texture, en inglés) en primera instancia a una clase general de problemas en visión conocidos como “shape from X” (forma a partir de X). Esto fue primero destacado a través del trabajo de Gibson en [44], en el cual se

establece la relación existente entre las variaciones en las propiedades de la textura y la forma de la superficie con la cual se puede inferir una forma tridimensional de los objetos. Ahondando un poco más, Stevens observó que ciertas propiedades de textura son perceptualmente significativas cuando se lleva a cabo, la extracción de la geometría de la superficie [138]. Aquí existen tres efectos que las superficies geométricas presentan con respecto de la apariencia de la textura en imágenes: escorzo (recurso artístico que se utiliza para dar la sensación de profundidad); escalamiento de los elementos de textura así como, un cambio en su densidad. El escorzo es debido a la orientación de la superficie en la cual el elemento de la textura es relacionado. Los cambios referidos al escalamiento y la densidad son ocasionados de acuerdo con la distancia existente entre los elementos de textura y el observador. Stevens concluye que la propiedad de percepción más estable que permite extraer información geométrica de la superficie es la dirección en la cual la imagen no es escorzada, a lo que él llamó la “característica de la dimensión”. Stevens sugiere que, se puede calcular una relativa información de profundidad usando el recíproco del escalamiento en la característica de la dimensión.

Por otro lado, la generación de formas 3D a partir del análisis de texturas 2D, encuentran su representación en los trabajos realizados por Mel en [99], y de forma más reciente Tomabari *et al.* en [146]. Finalmente, se considera que, la segmentación de textura consiste en encontrar las diferentes texturas existentes en una imagen como lo muestran los trabajos iniciales en esta área presentados por Reed y Wechsler [114]; dichos conceptos son retomados por Shotton *et al.* [133] para el reconocimiento y segmentación de múltiples objetos. Esto generalmente resulta complicado en el sentido de que, no es claro cuántas texturas existen y a qué clase de textura pertenecen. Por lo que, sólo se encuentran los límites de las texturas para realizar una sencilla clasificación. Dicho proceso de segmentación se realiza por medio de dos aproximaciones, la primera basada en regiones y la segunda en fronteras.

### ***Regiones Texturizadas***

El proceso de subdividir una imagen en regiones es llamado *segmentación*. Una región es descrita a través de la cuantificación de su contenido en textura. En procesamiento de imágenes esto es llevado a cabo, mediante tres enfoques diferentes, estadístico, estructural y espectral esto es, de acuerdo con lo expuesto por Russ en [119]. El primero hace referencia a la caracterización de texturas tal como, suavidad, granulado, rugosidad y sus combinaciones. Con respecto del análisis estructural, este se enfoca en la descripción de los arreglos que se presentan en

las primitivas de las imágenes, por ejemplo, el espacio existente entre líneas paralelas. Por otro lado, las técnicas espectrales están fundamentadas en las propiedades del espectro de Fourier y sus aplicaciones. Esto, con el fin de detectar una periodicidad global en una imagen por medio del reconocimiento de una alta energía presentes en los pequeños picos del espectro.

En imágenes naturales los bordes que delimitan texturas disímiles son los bordes que pueden establecer también, diferentes regiones u objetos. Se considera que, la primera operación del reconocimiento de objetos en una escena natural es, por tanto, la división de la imagen por medio de la segmentación de las diferentes texturas. Al segmentar la imagen en regiones, nos da la posibilidad de iniciar tareas como la de separar los objetos presentes en estas, para posteriormente, reconocer los mismos. Así, la tarea para la segmentación de objetos implica, por lo tanto, la identificación de áreas con texturas similares o uniformes. Esto implica que, los píxeles son clasificados por lo tanto, con base en una característica local propia como lo es, la textura. Esta forma de segmentar tiene como ventaja el que se tengan bien definidas las regiones a través de fronteras cerradas entre cada textura; aún cuando, su principal desventaja, será el que necesariamente se tenga definido el número de texturas existentes en la imagen.

Por otro lado, cuando se lleva a cabo una segmentación basada en *fronteras*, esta se fundamenta en detectar las diferencias en las texturas adyacentes. En ese sentido, por lo tanto, no es necesario que se tenga determinado el número de texturas, aún cuando pueda presentar ciertas indefiniciones en regiones que no queden bien separadas.

Desde los primeros trabajos relacionados a este tópico, como el realizado por Reed y Wechsler en [114] la segmentación de textura tiene como meta el dividir una imagen en regiones disjuntas dentro de una textura homogénea; mismo concepto que es nuevamente analizado bajo un nuevo enfoque por Shotton *et al.* en su estudio presentado en [133].

Por lo anterior parecería que, el problema de tratar imágenes de rango o en escala de grises como imágenes texturizadas es un problema complejo. Esto sin duda, es debido a la heterogeneidad presente en este tipo de imágenes. Así en este trabajo de Tesis se propone que, el proceso denso que se lleva a cabo con imágenes de rango sea tratada como una imagen texturizada, como lo puede ser cualquier imagen en niveles de gris. Esto con la finalidad de que la información obtenida a partir de este proceso, sirva como información de entrada a un siguiente proceso.

Por ejemplo, aún cuando en apariencia la imagen de rango parece presentar una textura homogénea (ver Figura 1.17-b), se establece que la información de profundidad o disparidad puede conformar superficies

en diversas regiones con diferentes grados de rugosidad (ver Figura 1.17-c). Por lo tanto, regiones con valores similares conforman un objeto o un área particular de la escena. Esto es posible, debido a la forma en cómo se adquiere la información; por ejemplo, el sensor Kinect Xbox360 utiliza la técnica de luz estructurada usando una proyección de un patrón de motas conocido (ver subsección 1.5.5). Dicha adquisición, establece una relación entre la disparidad y la profundidad con un error máximo que oscila en el orden de los 2 a 4 píxeles. Esto implica que, objetos que se encuentran a una distancia menor a los 80 *cm* o posteriores a los 400 *cm* no son percibidos de forma correcta.

### 1.5.9. La Técnica SDH

Desde un punto de vista estadístico y de acuerdo con lo propuesto por Lateghn *et al.* en [79] se puede considerar que aún en imágenes con patrones de textura aleatorios, estos pueden ser completamente descritos por la función de densidad de probabilidad (PDF, por sus siglas en inglés).

El primer estimador no paramétrico de la PDF se desarrolló con base en histogramas de frecuencia conjuntos [79] (joint histograms; jHST, por sus siglas en inglés). A pesar de la sencillez del mismo, este se convierte en algo impráctico de aplicar, incluso en regiones texturizadas pequeñas debido a que, el histograma puede terminar generándose con información poco relevante o igual a cero. Esto es debido a que, el número de entradas en el jHST excede por mucho, el número de píxeles contenidos en una región de textura de tamaño típico (fenómeno conocido como la “*maldición de la dimensionalidad*” o “*Curse of Dimensionality*”, en inglés) [79]. Por lo tanto, una alternativa para dar solución a este problema es considerar solamente, pares de píxeles (píxel de referencia y vecino) que sirvan como entradas al jHST. Esta manera de interpretar y construir el jHST para análisis de textura, la podemos encontrar en la técnica 2D-jHST conocida como matrices de *coocurrencia* de niveles de gris (Co-occurrence grey level matrix; GLCM, por sus siglas en inglés) [52].

De lo anterior, Unser en [149, 150], retoma la técnica 2D-jHST, la cual combina con filtros elementales, para estimar histogramas 1-D correspondientes a las sumas y diferencias de las intensidades de los píxeles vecinos. Dicha técnica se conoce como *Sum and Difference Histograms technique* (técnica de Histogramas de Suma y Diferencia o SDH, por sus siglas en inglés).



Figura 1.17: a) La imagen a color (Baby Middlebury dataset) es sólo referencia. b) Imagen de rango (disparidad). c) Presenta la superficie formada por los diferentes niveles de gris correspondientes a la imagen de disparidad.

En [150] el autor establece que, haciendo uso de dichos estimadores probabilísticos, estos pueden calcular la información de manera similar a como se lleva a cabo con las matrices de coocurrencia, pero presentando una reducción de los requerimientos de memoria en su procesamiento. Esto es porque en las matrices de coocurrencia generadas, su tamaño se incrementa en  $2^k$  veces el número de canales necesarios para su representación. Por ejemplo, los datos de 8 bits con los que se representa una banda original de una imagen satelital, se consideran 256 ( $2^8 = 256$ ) posibles valores, por ende, la matriz de coocurrencia sería igual a  $256 \times 256 = 65536$  celdas. Además, este hecho, incrementa el uso de memoria y tiempo computacional de forma exponencial, esto es, en  $2^k$ .

Es importante hacer notar que, haciendo uso de la técnica GLCM se generan diferentes matrices de coocurrencia para cada relación espacial entre el píxel de referencia y su vecino. Estas relaciones se pueden dar en 8 direcciones, aún cuando generalmente se calculan cuatro, según se considere la dirección del vecino: arriba, al costado o en diagonal (las otras cuatro direcciones se pueden calcular a partir de matrices simétricas, de la dirección calculada). Cada uno de estos cálculos puede ser interpretado a través de una función de probabilidad. Tal que, lo contenido en la matriz de coocurrencia se pueda expresar como una probabilidad  $P(i, j | d, \theta)$  donde, dos píxeles tienen relativas coordenadas polares  $(d, \theta)$  e intensidades respectivas de  $i, j$ . Por lo que, es necesario, realizar una doble sumatoria, para obtener un valor que describa la textura correspondiente a la imagen analizada. En este sentido, Unser en [150] demuestra que, una sumatoria a lo largo de los ejes paralelos a la diagonal principal de la matriz, resulte un conjunto de estadísticos de primer orden.

Tabla 1.1: Tabla con los descriptores de texturas más comunes expresados desde la matriz de coocurrencia [52] e histogramas de sumas y diferencias [150].

Descriptor de Textura	GLCM [52]	SDH [150]
Media	$\sum_i \sum_j i \cdot \hat{P}(i, j)$	$\frac{1}{2} \sum_i i \cdot \hat{P}_s(i)$
Varianza	$\sum_i \sum_j (i - \mu)^2 \cdot \hat{P}(i, j)$	$\frac{1}{2} [\sum_i (i - 2\mu)^2 \cdot \hat{P}_s(i) + \sum_j j^2 \cdot \hat{P}_d(j)]$
Energía	$\sum_i \sum_j \hat{P}(i, j)^2$	$\sum_i \hat{P}_s(i)^2 + \sum_j \hat{P}_d(j)^2$
Correlación	$\sum_i \sum_j (i - \mu) \cdot (j - \mu) \cdot \hat{P}(i, j)$	$\frac{1}{2} [\sum_i (i - 2\mu)^2 \cdot \hat{P}_s(i) - \sum_j j^2 \cdot \hat{P}_d(j)]$
Entropía	$\sum_i \sum_j -\hat{P}(i, j) \cdot \log(\hat{P}(i, j))$	$\sum_i -\hat{P}_s(i) \cdot \log(\hat{P}_s(i)) - \sum_j -\hat{P}_d(j) \cdot \log(\hat{P}_d(j))$
Contraste	$\sum_i \sum_j (i - j)^2 \cdot \hat{P}(i, j)$	$\sum_j (j)^2 \cdot \hat{P}_d(j)$
Homogeneidad	$\sum_i \sum_j \frac{1}{1 + (i - j)^2} \cdot \hat{P}(i, j)$	$\sum_j \frac{1}{(1 + j)^2} \cdot \hat{P}_d(j)$
Agrupamiento de prominencias	$\sum_i \sum_j (i + j - 2\mu)^4 \cdot \hat{P}(i, j)$	$\sum_i (i - 2\mu)^4 \cdot \hat{P}_s(i)$

Por otro lado, retomando el ejemplo de la imagen satelital (1 banda,  $2^k = 2^8$  niveles de color), al remplazarse una doble sumatoria realizada con la técnica GLCM por una sumatoria con la técnica SDH, este permitirá un cálculo ocho veces más rápido y necesitará una memoria ocho veces más pequeña que la evaluación habitual basada en las matrices de coocurrencia. Finalmente, el mismo autor presenta una serie de aproximaciones correspondientes a los diferentes descriptores de textura [53], pero, llevados a cabo a través de los estimadores de funciones de probabilidad, como lo podemos ver en la Tabla 1.1.

Algunos trabajos en los que se encuentra la aplicación de la técnica SDH, o propuestas con técnicas afines son consideradas en [127, 126, 79, 61, 39]. Sandid *et al.* en [126] sugieren codificar patrones locales utilizando un SDH tridimensional aplicado al reconocimiento de materiales. Dicha propuesta consiste, en analizar los valores máximos obtenidos de los histogramas de sumas y diferencias con respecto del píxel central, y sus vecindades codificando en el espacio de color correspondiente a la imagen. Adicionalmente, para robustecer la técnica, en lugar de hacer uso del valor exacto del píxel, hacen uso de los valores promedios considerando ventanas de  $3 \times 3$ ; alrededor del píxel evaluado sobre la componente de color correspondiente. En [127] el autor retoma dicha propuesta e incorpora los valores máximos obtenidos del histograma de diferencias, así como la información de rotación de la máscara en un rango de  $[d, \theta | d \in \{1, 2\}, \theta = \{0, 45, \dots, 315\}]$ . A pesar de que al incorporar la información de color y rotación el descriptor se vuelve más robusto, el tiempo de procesamiento es afectado significativamente. Ibarra en [61] aplica directamente la técnica de SDH para obtener las características de textura correspondientes a una imagen RGB transformada en el espacio de color *CIE-L\*ab*. Los componentes cromáticos *a* y *b* se utilizan como características de color durante la clasificación y el componente *L\** se utiliza para calcular características de textura.

Lateghan propone una aproximación para calcular un LBP haciendo uso de Modelos de mezcla Gaussiana; los cuales se basan en la estimación de PDF

aplicada a la salida de un banco de filtros (WFT wavelet frame transform, en inglés) aplicado a una región de interés. De forma similar, Feng *et al.* en [39] presenta un histograma de diferencia estructural para representar las texturas de la imagen y realizar la clasificación de esta. Una particularidad de dicha técnica es la sustitución del cálculo de las diferencias en la vecindad, por umbrales, entonces se establece un primer histograma de dichas características. Esto es, siempre y cuando existan patrones uniformes, de lo contrario, se considera la información obtenida a partir de las diferencias de las vecindades con respecto del píxel central. De igual forma, se aplica un LBP que “refina” dicha información de textura; lo cual, se puede considerar como una generalización del LBP propuesto por Guo *et al.* en [50]. Estos tres elementos descriptores de la imagen son los que conforman el SDHR (structural difference histogram representation, en inglés).

## 1.6. Definición del Problema

En la actualidad el apostar por sistemas de adquisición de imágenes de bajo costo sobre todo en el sector educativo es fundamental. La información que se obtiene a través de estos sistemas quizás no sea la óptima, pero, nos permite mostrar a través de diversos procesos que es posible obtener información relevante o de utilidad. Así, el procesamiento de la imagen de profundidad (o disparidad) deberá ser similar al que se lleva a cabo con cualquier tipo de imagen que se encuentra en escala de grises como lo es, por ejemplo, las imágenes médicas. De estas se destacan las mamografías por su proceso de adquisición ya que, en este la imagen digital obtenida es de bajo contraste debido al nivel mínimo de radiación con el que se debe llevar a cabo el proceso de formar la imagen. De forma similar, las imágenes de rango presentan visualmente poca variación de intensidades, especialmente si son escenas con aglomeración de objetos, o si estos se encuentran en un mismo plano.

Por lo tanto, en una imagen de rango las intensidades similares en una escena deberían definir cada región u objeto contenido en esta al mismo nivel de profundidad; sin embargo, es difícil precisar los límites entre cada uno de los elementos contenidos. Por ello, es esencial identificar los cambios de intensidad que permitan distinguir las diferentes regiones en la imagen.

En el caso de las microcalcificaciones se denota que, el agrupamiento de píxeles representativos de este tipo de masas anómalas presentan un nivel alto de intensidad, en comparación con el resto de la información. En este caso, aunado al problema del bajo contraste en estas imágenes la principal dificultad para llevar a cabo la segmentación e identificación de estas masas, radica en su tamaño, ya que, este oscila entre  $0.1 - 1 \text{ mm}$ . Lo que implica que, en el contexto de la imagen esto sea como un “punto luminoso” que no alcanza a irrumpir la homogeneidad de la información contenida.



Figura 1.18: Diagrama general de la propuesta.

El proceso de segmentación con el cual se particiona una imagen en regiones o extrae un objeto de interés se vuelve una tarea ardua de realizar cuando la información a procesar presenta una alta similitud. En este trabajo se propone mostrar cómo el análisis de textura es una alternativa confiable para detectar y segmentar objetos o regiones de interés en imágenes en escala de gris. El principio para considerar en ambos casos es el mismo, buscar una alta disimilitud a través del uso de los Histogramas de Sumas y Diferencias.

Establecer la metodología y técnicas necesarias para lograr una segmentación de una imagen en regiones y luego una detección más fina del contorno del objeto es uno de los principales intereses de este trabajo. Diferentes técnicas existentes en trabajos previos serán estudiadas, para lograr un conocimiento sólido de los diferentes algoritmos y que esto permita proponer diferentes soluciones a procesos de detección, localización y modelado de objetos. Toda la teoría descrita en este documento busca ser un referente para la enseñanza de visión por computadora en alumnos de nivel superior, permitiendo acceso a información detallada y descrita con ejemplos.

## 1.7. Descripción General de la Propuesta

En esta Tesis se presenta un análisis de textura aplicado a imágenes en escala de grises, las cuales son codificadas bajo diferentes formatos de acuerdo con el sistema de adquisición del cual fueron obtenidas. Por ejemplo, las imágenes en formato DICOM se obtienen a través de un sistema PAC; mientras que, imágenes de profundidad se obtienen haciendo uso del sensor Kinect 360.

Para efectos de esta propuesta, se llevará a acabo un análisis de la textura,

en el cual se estudiará la correspondencia que existe, entre los niveles de gris y la formación de diferentes regiones texturizadas. Esto con el fin de llevar a cabo la detección y segmentación de objetos o áreas de interés. En este orden de ideas, una aproximación estadística de la información contenida en la imagen se desarrolla, con el fin de realizar el cálculo de las diferentes texturas contenidas en dichas imágenes.

En un primer enfoque de dicho análisis, se calcula el descriptor de textura conocido como *agrupamiento de prominencias* (“cluster prominence” CP, por sus siglas en inglés). Este descriptor permite realizar el modelado de imágenes médicas para efectuar la detección y segmentación de microcalcificaciones en mamografías. Un segundo análisis se efectúa en zonas específicas en las que el histograma de CP presenta una alta frecuencia. Para evaluar la búsqueda, así como validar la presencia de calcificaciones se hace uso de la técnica de Vandermonde. Finalmente, a través de una red neuronal se realiza el proceso de clasificación de dichas áreas o regiones de interés.

Por otro lado, la técnica llamada Sparse Keypoint Detector (SKD por sus siglas en inglés) nos permite llevar a cabo dos procesos diferentes, la detección de los contornos en una escena, así como, la detección y segmentación de un objeto en particular, utilizando imágenes de rango. La técnica es subdividida en dos etapas, la primera hace referencia a un proceso denso mientras que la segunda a uno disperso. Ambos procedimientos involucran la obtención de puntos de interés representativos, los cuales son obtenidos por medio de descriptores locales. Estos son calculados en la superficie correspondiente a la imagen de rango previamente adquirida desde el sensor Kinect.

Un análisis de función de masa de probabilidad (probability mass function; PMF, por sus siglas en inglés), es aplicado a los descriptores correspondientes de los puntos de interés, que definen el contorno de los objetos en la escena. Esto es, con el fin de determinar un espacio de trabajo, llamado ROI (Region of Interest, en inglés). El objetivo de establecer esta área es para identificar la región donde se encuentra ubicado el objeto a detectar. Adicionalmente, los objetos contenidos en la ROI también son definidos por su contorno, a través de los puntos de interés. El objeto 3D es detectado, con base en el análisis estadístico de los descriptores contenidos en la ROI. Finalmente, con el contorno del objeto, se determinan los mejores posibles puntos de sujeción para este.

Adicionalmente, la propuesta es evaluada cualitativa y cuantitativamente obteniendo resultados positivos, con relación a su desempeño respecto a otras propuestas presentadas en el estado del arte. Dicha evaluación es llevada a cabo a través de las bases de datos Middlebury [58, 112], así como, nuestra propia base DSPlab [57].

El diagrama esquemático del procesamiento general que se lleva a cabo es presentado en la Figura 1.18.

### 1.7.1. Aportaciones

Esta Tesis describe diversas contribuciones en el área de Procesamiento Digital de Imágenes a través del análisis de textura en imágenes representadas en niveles de gris. Dentro de las aportaciones de este trabajo se consideran las siguientes:

- Una estrategia para la detección de microcalcificaciones en mamografías basado en el cómputo y análisis de los Histogramas de Suma y Diferencia.
- Un procesamiento automático de la información contenida en las mamografías, con la cual se tenga una herramienta que apoye en la detección temprana de microcalcificaciones, con un desempeño similar o superior al de otras técnicas presentadas en trabajos previos.
- Una técnica para la extracción de puntos de interés partiendo de la información contenida en una sola imagen de rango, sin preprocesamiento y de fácil implementación en comparación con técnicas similares consideradas como referencia en el estado del arte.
- Además, la técnica propuesta permite un procesamiento tipo disperso, con el cual, se decrementa el tiempo de ejecución en comparación con el de otras técnicas similares del área.
- En un segundo nivel de procesamiento, el método propuesto permite la detección de la región de interés y del contorno de los objetos en la escena.

A continuación se presentan las contribuciones dos artículos en revistas indizadas JCR (Journal Citation Reports, en inglés) y 1 conferencia.

#### Artículos en revistas indizadas

- Cruz-Bernal A., Flores-Barranco M.M., Almanza-Ojeda D.L., Ledesma S., and Ibarra-Manzano M.A., “Analysis of the Cluster Prominence Feature for Detecting Calcifications in Mammograms”, *Journal of Healthcare Engineering*, vol. 2018, Article ID 2849567, 11 pages, 2018. ISSN: 2040-2295 (Print), ISSN: 2040-2309 (Online) DOI:10.1155/7158
- Cruz-Bernal A., Almanza-Ojeda D.L., Ibarra-Manzano M.A., “Object Detection from a Range Image Using Sparse Keypoint Detector Technique”, *IEEE Latin America Transactions*, no. 5, May 2018. DOI: 10.1109/TLA.2018.8408451

### Conferencias

- Cruz-Bernal A., Almanza-Ojeda D.L., Ibarra-Manzano M.A., “Contour Detection at Range Images Using Sparse Normal Detector”, LNCS 9703 Springer, (2016), ISBN 978-3-319-39392-6, DOI: 978-3-319-39393-3\_12

## 1.8. Organización del Contenido

El contenido de este documento de Tesis se describe a continuación. En el Capítulo 2 se establece la metodología empleada para llevar a cabo la extracción de características de textura en imágenes en formato DICOM y de Rango utilizando la técnica SDH. En el Capítulo 3 se presenta la propuesta para la detección de puntos de interés así como la detección de un objeto en una región de interés en particular. En el Capítulo 4, se introducen los conceptos generales de entropía y se establece la definición de entropía como medida, pero calculada como atributo de textura. El Capítulo 5 se mostrarán los resultados obtenidos más relevantes. El último capítulo está dedicado a las conclusiones y trabajos futuros del presente proyecto de tesis.

+

## Capítulo 2

# Metodología Parte I: Características de Textura en Imágenes DICOM y de Rango

*“For since the fabric of the universe is most perfect and the work of a most wise Creator, nothing at all takes place in the universe in which some rule of maximum or minimum does not appear.”*

Leonhard Euler

La metodología que se presenta a continuación corresponde al procesamiento denso que se lleva a cabo para dos tipos diferentes de imágenes representadas en niveles de gris. Cada caso es tratado ampliamente de acuerdo con los enfoques que se tienen para cada tipo de imagen. En el primer caso se tratan imágenes en formato DICOM con el objetivo de encontrar una región de interés, la cual corresponderá a una microcalcificación. Por otro lado, el segundo corresponde al procesamiento de imágenes de rango. En este caso, la información resultante de este proceso contendrá el vector de características por medio del cual en un proceso que llamamos en este trabajo disperso se lleva a cabo desde la detección de puntos de interés hasta la de un objeto en una región en particular.

### **2.1. Procesamiento de Imágenes DICOM mediante la Técnica SDH**

Un primer procesamiento de imágenes en niveles de gris haciendo uso de la técnica SDH es la que se propone para el análisis de mamografías, con el fin de llevar a cabo la detección o no de microcalcificaciones. Particularmente,

haremos uso de imágenes en formato DICOM. Las ventajas de este tipo de imágenes fueron descritas en la subsección 1.5.4 de la Introducción. Es importante hacer notar que, el procesamiento de estas imágenes intenta ser una herramienta que ayude a los médicos a buscar signos tempranos de cáncer de mama, a través de la detección de la presencia de microcalcificaciones en la imagen en general. Cabe aclarar que, no se pretende especificar ubicación o cantidad de estas.

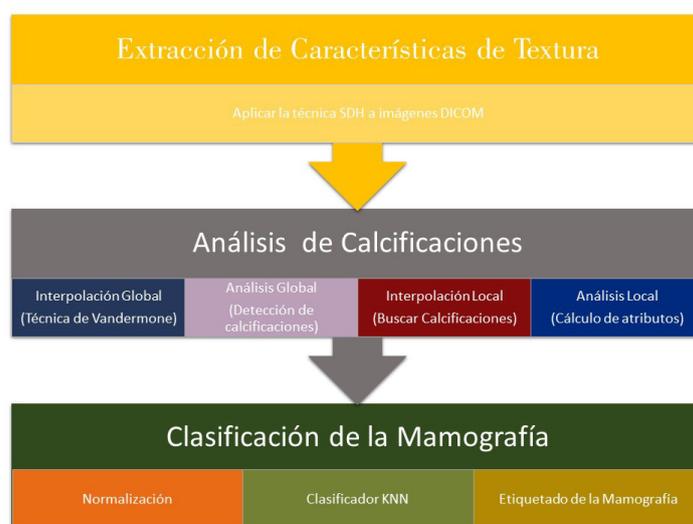


Figura 2.1: Esquema general de la técnica empleada para la detección de microcalcificaciones en mamografías.

La metodología que se propone para llevar a cabo el procesamiento de imágenes médicas, propiamente mamografías corresponde al diagrama esquemático 2.1. En el mismo se describen tres procesos que pueden englobar la metodología propuesta. El primer bloque corresponde al cómputo de características de textura, de manera particular nos centramos en la extracción de agrupamiento de prominencias. El segundo proceso nos permite llevar a cabo un análisis detallado de dichas característica haciendo uso de la técnica de Vandermonde; para que, en el último bloque se establezca una clasificación de las calcificaciones detectadas por medio del clasificador KNN.

### 2.1.1. Extracción de la Característica de Textura CP

La propuesta que se presenta en este trabajo considera aplicar la técnica SDH para extraer la característica de textura se denominada *agrupamiento de prominencias* “CP” (cluster prominence, en inglés), con el fin de identificar cambios en las intensidades de los niveles de gris en la mamografía. En esta aproximación, tales cambios obedecen a conceptos básicos de mate-

mática, en los cuales se define un punto de inflexión como un punto sobre la curva que separa la parte convexa de la concava, tal que,  $f''(x) = 0$ . De acuerdo con el criterio de la segunda derivada, si  $f'''(x_1) > 0$  se tendrá un mínimo (*concauidad*), en caso contrario será un máximo (*convexividad*).

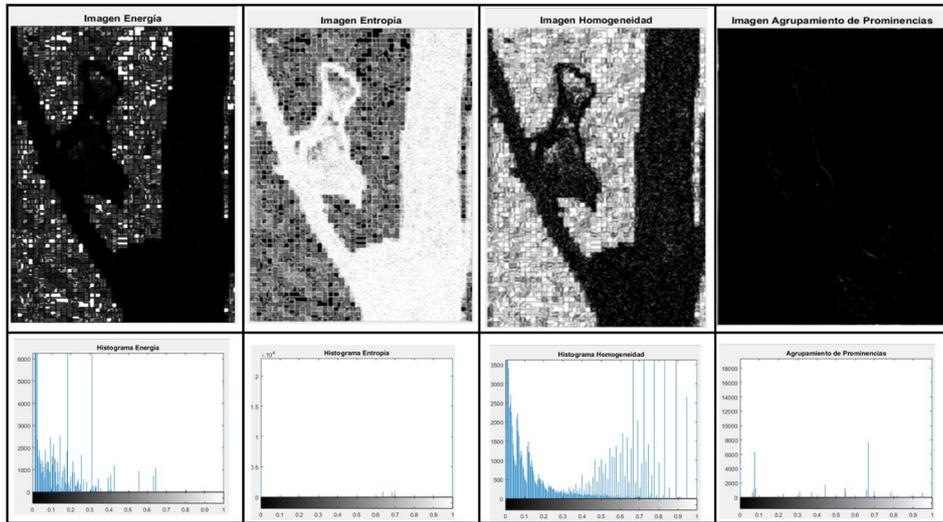


Figura 2.2: Histogramas de características de textura usando SDH para una imagen en nivel de gris. En el primer renglón de izquierda a derecha las características de energía, entropía, homogeneidad y agrupamiento de prominencias. En el segundo renglón se muestran los histogramas cada característica.

Es importante notar que, se elige CP sobre otras características de textura debido a que representa una medida de asimetría; para la cual un valor alto en dicha característica indica amplios cambios en la escala de grises tal como, lo establece Yang *et al.*, en [162]. En otras palabras, CP es una medida de cuán uniforme es la distribución de la información en esta escala.

En este sentido, en la figura 2.2 la imagen obtenida de la característica de textura CP, es la que se encuentra prácticamente “ennegrecida”. Sin embargo, con respecto de esta misma característica en su histograma correspondiente se expone la información que tiene una mayor frecuencia en relación con la escala de grises, tal que, existen dos extremos en este. El primero se ubica en el rango del eje horizontal entre  $[0 - 0.2]$  y el segundo entre  $[0.6 - 0.8]$ . Dichos rangos se encuentran normalizados correspondiendo a los niveles de gris de la imagen. Ahora bien, en el caso de las mamografías, las calcificaciones son pequeñas aglomeraciones de puntos con valores cercanos a 1 de intensidad. Estas se encuentran posicionadas como un valor prominente del histograma de CP, es decir, por ejemplo, en cualquiera de los dos rangos que se presentan en 2.2. Por lo que, en nuestra propuesta se utilizará un segundo análisis

sobre CP mediante un modelo polinomial, esto con el fin de establecer la pertenencia o no de las puntos detectados como un agrupamiento que defina una microcalcificación.

Por lo tanto, haciendo uso de esta propiedad se propone que, valores cercanos a 1 de CP representa una posible microcalcificación. Para validar esta aseveración se realiza un análisis polinomial utilizando la técnica de Vandermonde.

El proceso da inicio con el cómputo de la característica de textura CP a través de

$$\sum_i (i - 2\mu)^4 \cdot \hat{P}_s(j)$$

presentada en la tabla 1.1. Por otro lado, el tamaño de la ventana que recorre la mamografía es de  $3 \times 3$  píxeles. Mientras que, el desplazamiento solamente se lleva a cabo en sentido horizontal. Esto es, con una diferencia de 1 píxel debido al tamaño de las microcalcificaciones; ya que estas son menores a  $\frac{1}{4}$  de pulgada. La relación del tamaño de la ventana y el desplazamiento que se elige, se encuentra en correlación con respecto del efecto que se desea producir en el procesamiento de las imágenes.

### 2.1.2. Análisis de Calcificaciones

El siguiente paso (diagrama descriptivo en la figura 2.3) es analizar las posibles calcificaciones utilizando la técnica de Vandermonde. Para lo cual, primero se lleva a cabo una interpolación global, haciendo uso de una función  $f(cp)$  con la que se describe los mejores valores de CP obtenidos. Esta función es empleada para analizar el rango del histograma con altas posibilidades de encontrar calcificaciones, por medio de  $n$  puntos de inflexión, referidos aquí como  $Z_{cp}$ .

Cada punto de inflexión en este caso, representa zonas destacables de aglomeraciones en las mamografías. Por lo tanto, realizando un análisis deberá realizarse para evaluar la búsqueda, así como validar la presencia de calcificaciones en una zona específica, la cual presenta una alta frecuencia en el histograma. Posteriormente, la función  $g(cp)$  se obtiene aplicando una interpolación local, a la información contenida en el histograma ( $h_{cp}$ ). Esto es, la interpolación local  $g(cp)$  se encuentra definida en el rango marcado por los puntos de inflexión de  $f(cp)$ ; el cual permite realizar un análisis más detallado de una región en específico. Adicionalmente, estableciendo un rango definido de la función  $g(cp)$ , se genera un vector de características llamado  $\mathbf{X}$ . Este vector será el que contenga los descriptores para dar inicio al proceso de clasificación de la mamografía. Dichos atributos corresponden a, el número de píxeles ( $NP$ ), número de ceros ( $NZ$ ), media ( $\mu$ ), desviación estándar ( $\sigma$ ) y varianza ( $\sigma^2$ ) correspondientes a la interpolación local llevada a cabo con  $g(cp)$  y el histograma resultante de la interpolación de rango  $h_{cp}^*$ .

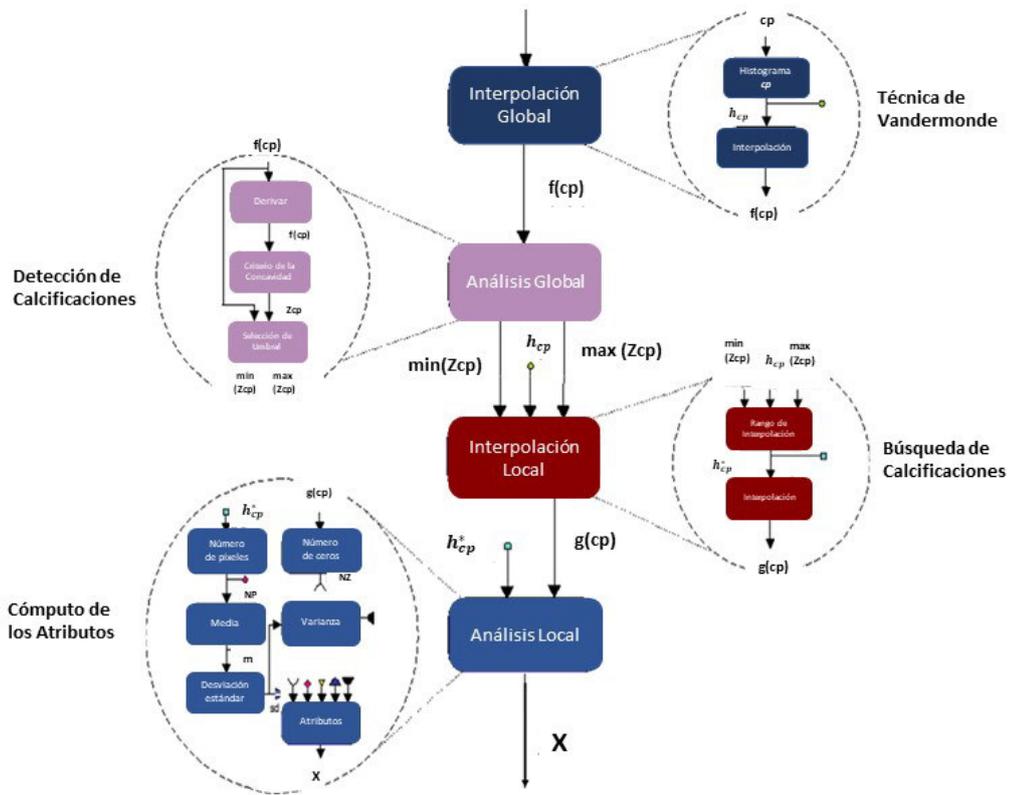


Figura 2.3: Diagrama descriptivo correspondiente al *análisis de las calcificaciones* (analysis of calcifications) del esquema 2.1.



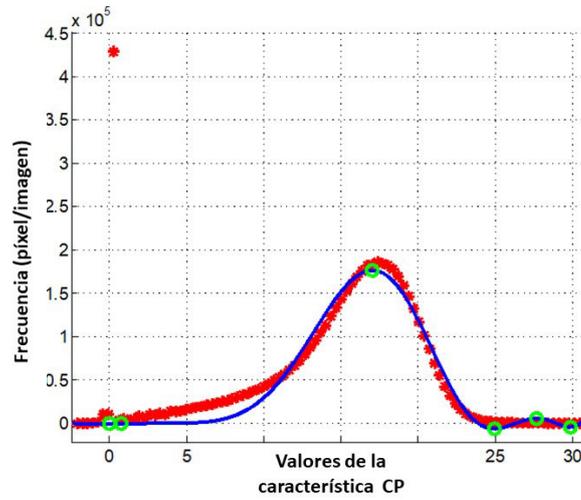


Figura 2.4: Ejemplo de una interpolación global para  $h_{cp}$ . Los \* muestran los valores discretos de  $h_{cp}$ . Mientras que, la línea azul corresponde a la interpolación polinomial de  $f(cp)$ , y los o marcan los puntos de inflexión ( $Z_{cp}$ ).

interpolación de dichos datos se muestra a través de la línea azul, calculada por medio de la ecuación 2.2. Es importante destacar que,  $f(cp)$  “sigue” la forma de la  $h_{cp}$ , excepto en los valores más altos de las características CP, así como donde se presentan valores locales frecuencia, que son considerados como mínimos y máximos de la función. Esto es debido a que, la matriz de Vandermode es un sensible a los cambios en los valores interpolados, por eso si existe un cambio, aunque este sea pequeño será detectado por la interpolación, por lo que, se ha utilizado esta característica para detectar las fluctuaciones en las frecuencia altas.

### *Análisis global: detección de microcalcificaciones*

Los resultados obtenidos de la subsección anterior mostrados en la figura 2.4, presentan los puntos de inflexión  $Z_{cp}$ . Por lo que, para establecer un rango de interés para la detección de calcificaciones, es necesario analizar el histograma en los últimos puntos de inflexión, por ello se establece un criterio basado en el análisis de la concavidad, tal que, “*la búsqueda de zonas con posibles calcificaciones se extiende a partir del penúltimo punto de inflexión hasta el último punto de inflexión, siempre que este sea un máximo; de lo contrario, se extiende desde el antepenúltimo punto de inflexión si el último punto de inflexión es un mínimo*”. La descripción matemática del anterior

supuesto es

$$th_{min} = \begin{cases} cp_{n-1}, & \text{if } f(cp_n) > f(cp_{n-1}) \\ cp_{n-2} & \text{de otra forma} \end{cases} \quad (2.4)$$

Mientras que,  $th_{min}$  describe el inicio del rango del umbral en dónde es posible encontrar calcificaciones,  $th_{max}$  determina el final de este. Por lo que, en correlación con la ecuación 2.4, definimos a  $th_{max}$  como

$$th_{max} = cp_n \quad (2.5)$$

por lo que, 2.5 se define con respecto del último punto de inflexión. Un esquema ilustrativo correspondiente a las ecuaciones 2.4 y 2.5 es mostrado en la figura 2.5. El establecer estos rangos, nos permite generar una nueva interpolación de  $h_{cp}$ . Esta interpolación es referida como  $g(cp)$  la cual interpretamos como una “interpolación local”.

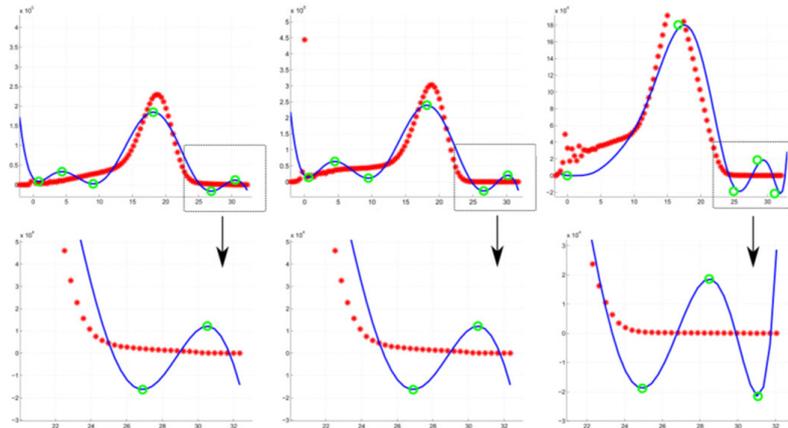


Figura 2.5: Los \* muestran los valores de  $h_{cp}$ . La línea azul corresponde a la interpolación global  $f(cp)$  y los o son ( $Z_{cp}$ ). Las dos primeras imágenes primer renglón, corresponden al primer caso de la ecuación 2.4, mientras que la última al segundo caso de la misma ecuación. Eje horizontal valor de CP. Eje vertical frecuencia (píxel/imagen).

***Interpolación local: búsqueda de microcalcificaciones***

La parte del histograma  $h_{cp}$  localizado en el intervalo  $[th_{min}, th_{max}]$  es referido como  $h_{cp}^*$ . En esta interpolación local, la función  $g(cp)$  ajusta los valores de la curva  $h_{cp}^*$  con mayor precisión y se utiliza para calcular los nuevos puntos de inflexión de  $Z_{cp}$ .

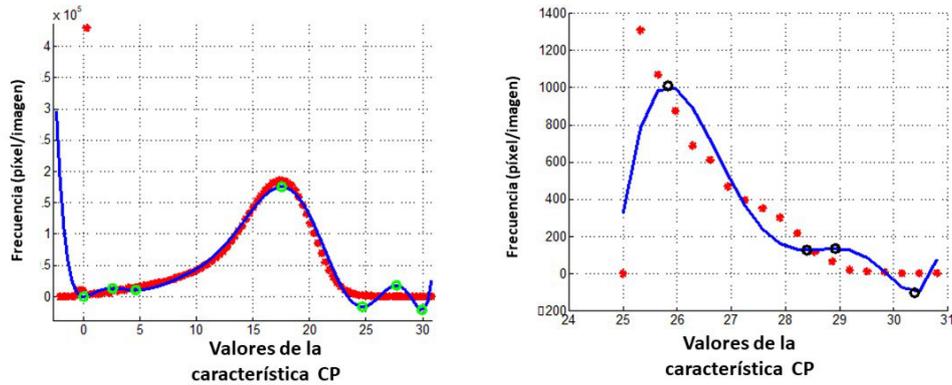


Figura 2.6: Comparación de una interpolación global (imagen izquierda)  $h_{cp}$ , con respecto de una local (imagen derecha)  $h_{cp}^*$ .

En la figura 2.6 se muestra un ejemplo comparativo de las dos interpolaciones realizadas. La imagen izquierda corresponde al proceso global ( $f(cp)$ ), mientras que el de la derecha al local ( $g(cp)$ ). En ambas imágenes los asteriscos corresponden a  $h_{cp}$  y a  $h_{cp}^*$ , respectivamente. Observemos como, el rango que se establece para este ejemplo se encuentra entre [25, 30]. Por lo que, en el mismo es donde se lleva a cabo la interpolación local. Cualitativamente, es visible que, el ajuste polinómico es mucho mejor, lo que se evidencia a través de los nuevos  $Z_{cp}$ , localizados como círculos negros en la imagen izquierda.

### *Análisis local para el cálculo de los atributos*

La función  $g(cp)$  se utiliza para calcular el vector  $\mathbf{X}$ . El cual es obtenido a partir de la mamografía para detectar si existen o no calcificaciones. Los atributos calculados son: 1) el número de ceros  $NZ$  y 2) el número de píxeles contenidos en  $h_{cp}^*$ . Partiendo del segundo atributo, es posible calcular tres valores estadísticos, media ( $\mu$ ), desviación estándar ( $\sigma$ ) y varianza ( $\sigma^2$ ); lo que nos proporciona un total de 5 atributos.

### 2.1.3. Clasificación de las Mamografías

Los cinco atributos se normalizan mediante la técnica de datos centrados-reducidos; los cuales se utilizan como entrada a un clasificador conocido como el de los k-vecinos más cercanos (k-nearest neighbors, en inglés o por sus siglas  $KNN$ ). El rendimiento de un clasificador  $KNN$  está determinado por la elección de  $K$  (clases) así como, de la distancia métrica aplicada; esta última es utilizada en la medición de un nuevo vector de atributos respecto de cada  $K$  clase. Por lo que, esta medida nos otorga el punto de referencia

con el cual se podrá discernir entre la pertenencia de cada caso en función de los atributos de dichas características a una clase. Por lo anterior, la métrica a utilizar será la distancia Euclidiana. Para, ayudar a clasificar datos en la frontera de las clases de datos individuales,  $K$  debe elegirse como un número impar, en nuestro caso  $K = 3$ .

El clasificador fue entrenado y probado utilizando la validación cruzada conocida como LOOCV (“Leave One Out cross validation”, en inglés) [77]. Esta técnica de validación garantiza un error muy bajo, además, se usa normalmente para bases de datos pequeñas que proporcionan un ajuste máximo del conjunto de entrenamiento y el conjunto de prueba independiente tal como lo muestra Jain *et al.*, en [62].

## 2.2. La técnica SDH Usando Imágenes de Rango

Las imágenes de profundidad o disparidad de acuerdo con la información revisada en la sección 1.5.4 contienen información geométrica de la escena. El análisis de la información geométrica obtenida de este tipo de imágenes puede ser llevado a cabo, a través de aproximaciones probabilísticas.

La información contenida en la imagen de rango ( $I_{Range}$ ) puede ser definida a través de una serie de particiones similares, a un grid rectangular de tamaño  $k \times l$ , donde ( $k = 1, \dots, M; l = 1, \dots, N \mid M \ll K$  y  $N \ll L$ ), donde  $[K, L]$  corresponde al tamaño de la  $I_{Range}$ . Por ejemplo, para saber la distribución de probabilidad de las diferentes intensidades de niveles de gris, se puede fijar en una partición cualquiera, la existencia de dos puntos definidos como  $S_1 = S_{k,l}$  y  $S_2 = S_{k+d_1, l+d_2}$ . Donde,  $S_2$  se encuentra a una distancia relativa de  $S_1$ , la cual se delimita a través de un desplazamiento  $d = (d_1, d_2)$ . Este desplazamiento nos permite, establecer una distancia polar  $(d_r, \theta)$ , donde  $0^\circ \leq \theta \leq 180^\circ$  con respecto al eje horizontal.

Adicionalmente, se debe de asumir que la matriz de coocurrencias empleada en el análisis de texturas 1.5.8, puede ser aproximada a través del histograma de sumas y diferencias, como es presentado por Unser en [150]. Esta estimación es llevada a cabo a lo largo del dominio de la  $I_{Range}$ , para calcular los estadísticos correspondientes a la imagen. En este caso, los ejes principales son definidos a través de los eigenvectores y eigenvalores. Por lo tanto, estos parámetros corresponden a la solución de la siguiente ecuación:

$$\mathbf{I}_{cov} \cdot v = \lambda v \quad (2.6)$$

$\mathbf{I}_{cov}$  es la matriz de covarianza asociada a un par de variables aleatorias, tal como,  $S_1$  y  $S_2$ . Así,  $\lambda$  es el eigenvalor de la ecuación 2.6 y el vector  $v$  es el eigenvector. Al obtener la solución de la ecuación 2.6, esta nos permite calcular un sencillo tensor de inercia en el espacio de  $\mathbb{R}^2$ . Esto es,

$$\mathbf{I}_{\text{cov}} = \sigma_s^2 \begin{pmatrix} I_{xx} & -I_{xy} \\ -I_{yx} & I_{yy} \end{pmatrix} = \sigma_s^2 \begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix} \quad (2.7)$$

donde, el momento de inercia  $I_{xx} = I_{yy}$  y estos a su vez, son iguales a 1. Es importante recordar que, el momento de inercia puede ser definido con respecto de un eje arbitrario, el cual no necesariamente pasa por el centro de masa.

Por otro lado, los productos de inercia son  $-I_{xy}$  y  $-I_{yx}$  donde,  $-I_{xy} = -I_{yx}$ , estos son representados por la variable  $\phi$ . Teniendo en cuenta que es un proceso estacionario, esto es *discreto*; la variable aleatoria solamente tomará uno de los valores de las intensidades de gris.

Solucionando la ecuación 2.6 por medio de la ecuación 2.7 tendremos que,

$$\left. \begin{array}{l} \lambda_1 = \sigma_s^2(1 + \phi) \\ \lambda_2 = \sigma_s^2(1 - \phi) \end{array} \right\} \text{ Eigenvalores}$$

y,

$$\left. \begin{array}{l} v_1 = \left[ \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right]^T \\ v_2 = \left[ \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2} \right]^T \end{array} \right\} \text{ Eigenvectores} \quad (2.8)$$

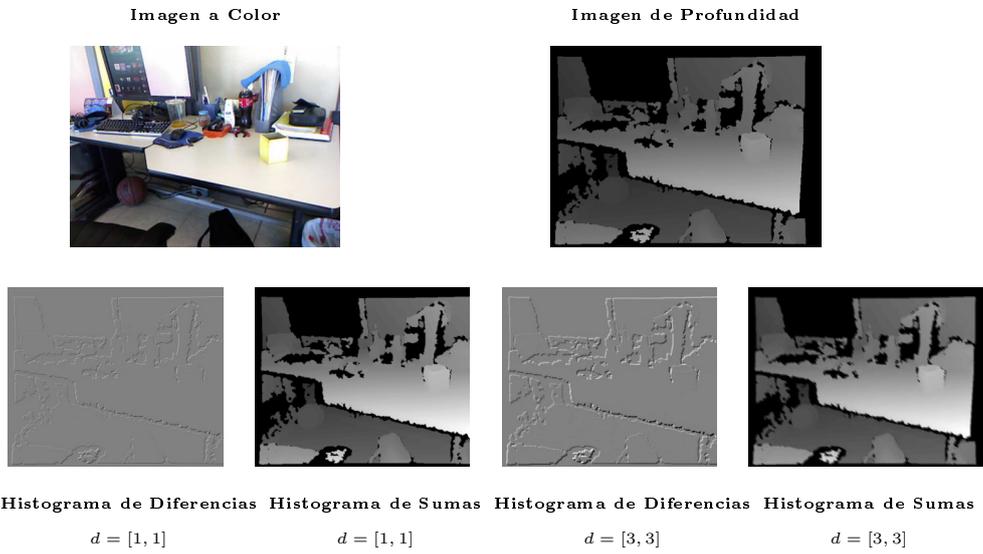
por tanto, el producto de inercia por la varianza se puede definir como,  $\sigma_s^2 \phi = E\{(S_1 - \mu)(S_2 - \mu)\}$ , donde,  $\mu = E\{S_1\} = E\{S_2\}$ , tal que

$$\sigma_s^2 = E\{(S_1 - \mu)^2\} = E\{(S_2 - \mu)^2\} \quad (2.9)$$

Los eigenvectores de la ecuación 2.8  $[v_1, v_2]^T$  determinan los ejes de inercia de cualquier matriz de covarianza de  $2 \times 2$ ; cuya varianza es descrita por medio de la ecuación 2.9. Tal que, todos los puntos obtenidos a partir de la ecuación 2.8 son variables aleatorias decorrelacionadas, cuyas varianzas son por lo tanto,  $\lambda_1$  y  $\lambda_2$  (eigenvalores).

Por consiguiente, la *función de probabilidad conjunta* (joint probability function; JPF, por sus siglas en inglés) de  $P(S_1, S_2)$  de todos los niveles de gris observables, se describe, a través de  $P(p_0, p_d | d_{rad}, \theta)$  donde,  $\{p_0, p_d\}$  corresponde a la probabilidad para las intensidades en los puntos  $(S_{k,l})$  y  $(S_{k+d_1, l+d_2})$  respectivamente, en el espacio  $\mathbb{R}^2$  y  $(d_{rad}, \theta)$  es la distancia polar, en la cual se fija la observación.

## a) Cubito



## b) Reinder

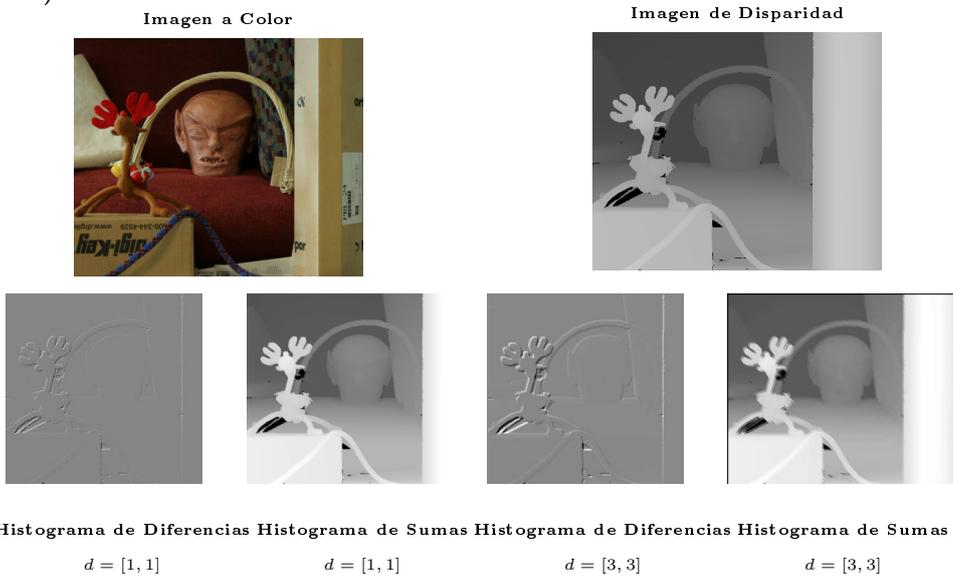


Figura 2.7: Ejemplos de la técnica SDH aplicada a) DSPLab Dataset (adquisición de las imágenes por medio del sensor Kinect; en un ambiente no controlado) y b) Middlebury dataset (adquisición de ambas imágenes por medio de un sistema par-estéreo bajo condiciones controladas). Diferentes distancias  $[1, 1]$  y  $[3, 3]$ . La imágenes a color son sólo referencia.

Es importante destacar que, el cálculo de la JPF a lo largo del eje princi-

pal es *solamente una aproximación*; la cual nos permite representar la JPF del total de la  $I_{Range}$ . Así, una transformación lineal correspondiente a la solución de la ecuación 2.6 considerando la ecuación 2.8 es expresada como

$$\varsigma_1 = \frac{\sqrt{2}}{2}(S_1 + S_2) \quad \text{y} \quad \varsigma_2 = \frac{\sqrt{2}}{2}(S_1 - S_2)$$

así,

$$P(S_1, S_2) = P(\varsigma_1, \varsigma_2) = P_s(\varsigma_1) \cdot P_d(\varsigma_2) \quad (2.10)$$

De lo anterior, los términos  $(S_1 + S_2)$  y  $(S_1 - S_2)$  representan respectivamente, la *suma* y *diferencia* de la distribución de probabilidad correspondientes a las intensidades de los niveles de gris, de los puntos  $S_1$  y  $S_2$ . Así, se define que  $S_{sum} = (S_1 + S_2)$  y  $S_{dif} = (S_1 - S_2)$ , con respecto del punto  $(k, l)$ . Entonces, del lado derecho de la función de probabilidad con respecto de la ecuación 2.10 se reescribe en términos de JPF tal que,

$$P_s(\varsigma_1) \cdot P_d(\varsigma_2) = \text{JPF}\{S_{sum}(\varsigma_1) = p_0\} \cdot \text{JPF}\{S_{dif}(\varsigma_2) = p_d\} \quad (2.11)$$

donde,

$$P_{sum}(p_0) = \text{JPF}\{S_{sum}(\varsigma_1)\} \quad (2.12)$$

y

$$P_{dif}(p_d) = \text{JPF}\{S_{dif}(\varsigma_2)\} \quad (2.13)$$

La ecuación 2.11 presenta la independencia entre los eventos calculados, para la probabilidad conjunta de las sumas  $P_{sum}$ , así como la referente a las diferencias  $P_{dif}$ . Por lo que, para cada uno de ellos se lleva a cabo la generación de los respectivos histogramas de frecuencia conjuntos (jHST), mencionados en la sección 1.5.9; tal que,  $\sum P_{sum}$  formará el histograma de sumas (ecuación 2.12), mientras que  $\sum P_{dif}$  el de las diferencias (ecuación 2.13).

En la Figura 2.7 podemos analizar, un ejemplo de la aplicación de la técnica SDH usando imágenes de rango. Se considera para efectos de ilustración, la relación de diferencias entre grupos de dos píxeles de la imagen original, a una distancia dada  $[d_1, d_2] = [1, 1]$  y  $[d_1, d_2] = [3, 3]$ , sin rotaciones y con una ventana de dimensiones  $3 \times 3$ . Los primeros dos renglones corresponden a la imagen cubito de DSPLab (2.7-a); mientras que, la Figura 2.7-b hace referencia a la imagen Reindere de la base de datos Middlebury. Tanto en la segunda como en la tercera columna de dichas figuras encontramos las imágenes correspondientes a los histogramas de sumas; en tanto, la segunda y cuarta columna son concernientes a los histogramas de diferencias. Conforme se hace más grande la distancia entre los puntos  $S_1$  y  $S_2$ , esto es, se varía la distancia  $[d_1, d_2]$  se genera *blurring* en la imagen. Por ende, se pierde nitidez, lo cual se muestra en las imágenes correspondientes a los histogramas de sumas en comparación con sus originales ya sea de profundidad o disparidad,

el resultado es el mismo. Este mismo efecto se produce de manera similar, si se incrementa el tamaño de la ventana que recorre la imagen original.

Por otro lado, en los histogramas de diferencias las regiones que se encuentran en primer plano presentan una mejor definición. Pero este resultado no garantiza que si se usa una ventana más amplia o una distancia mayor entre los pares de píxeles sea una estrategia conveniente y se obtengan mejores resultados.

El aumentar estos parámetros ( $d_r$  o tamaño de la ventana) incrementa el tiempo de cómputo, así como el blurring. Más aún, se genera una región de no información que “enmarca” la imagen (ver última columna, Figura 2.7-b), en la parte superior y lado izquierdo, de esta. Esto es debido a que la ventana que recorre la imagen de rango, el píxel de referencia  $S_1$  se encuentra en el centro. Mientras que,  $S_2$  se fija en cada uno de los píxeles a su alrededor separados por una distancia de 1 píxel, si se trabaja con una ventana de  $3 \times 3$ , por ejemplo. Por otro lado, si se analizan imágenes muy saturadas o desordenadas, entonces, se corre el riesgo de traslapar límites o incluso áreas correspondientes a diferentes partes de la escena, por el blurring generado con ventanas o distancias muy grandes. Es importante hacer notar que, las pruebas de ejemplo generadas se realizaron con una ventana de  $[3, 3]$ , con parámetros de  $[d_r = 1, \theta = 0^\circ]$ .

## Capítulo 3

# Metodología Parte II: Segmentación y Detección en Escala de Grises

*“Divide each difficulty into as many parts  
as is feasible and necessary to resolve it.”*

René Descartes

Es importante hacer notar que, la extracción de características locales con lleva el determinar un tamaño de región adecuado de acuerdo con el procesamiento a realizar. Tal que, si se selecciona un área muy pequeña esta podría contener solamente ruido; mientras que, una región demasiado amplia podría perder el sentido de “local”.

De acuerdo con lo presentado por Creusot *et al.*, [29] la extracción de características que se obtienen de la información contenida en las imágenes de rango deberían no ser afectadas por la escala, rotación o iluminación. Esta última, no implica un problema por el sensor usado para la adquisición de los datos. Sin embargo, en relación con la invarianza de la pose, esta es dada, siempre y cuando se establezca un marco de referencia. Creusot propone, utilizar un modelo de escala rígida adaptable al área en donde son localizables los que podrían ser los puntos clave.

Debido a que, la vecindad correspondiente a cada punto característico depende de una distancia, en nuestra propuesta se establece como tamaño de la región, para la obtención de los descriptores de textura una ventana de tamaño de  $3 \times 3$  píxeles. Mientras que, para los descriptores geométricos, la región de interés se fija en una ventana de radio no mayor a la distancia máxima calculada entre los descriptores. Esta es computada haciendo uso de la distancia de ajedrez (distancia Chebyshev, o  $L_\infty$ ); tal que, los puntos de interés seleccionados sean considerados solamente, los que pueden man-

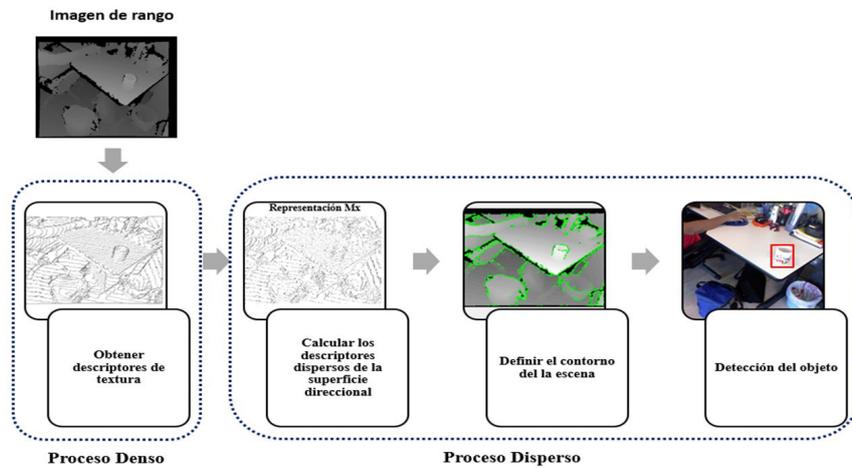


Figura 3.1: Diagrama general para la detección de un objeto en una región de interés.

tenerse a lo largo de las diferentes transformaciones.

### 3.1. Consideraciones de Superficie para Imágenes de Rango

La información de profundidad o disparidad es expresada en niveles de gris, con áreas que presentan diversos suavizados en estas. Dicho suavizado en una imagen a color corresponderían a los cambios de iluminación. Mientras que, en una imagen de profundidad o disparidad hace referencia a la distancia existente entre la escena y el sensor. Se dice entonces, que las imágenes de rango son invariantes a los cambios de luz, ya que esta no afecta su proceso de adquisición.

Por otro lado, si consideramos que la formación de diversas “capas” de superficies son descritas a través de superficies no Euclidianas, sino más bien Lambertianas (refleja la radiación incidente de manera uniforme, en todas las direcciones). Entonces, de acuerdo con Jeon *et al.* en [67], la *tonalidad* de una superficie puede ser determinada como el producto punto de la superficie normal y la dirección de la luz. Por otro lado, Jayasumana *et al.* [66] describen una aproximación a espacios tangenciales a través de un producto interno, en el que, se define una matriz simétrica positiva (SPD, por sus siglas en inglés)  $Sym_d^+$ . Los autores basados en el concepto de SPD generan una serie de kernels aplicados a varios problemas de visión por computadora entre ellos, la clasificación de texturas. Tabia *et al.* en [142] retoman este concepto y lo trasladan a las matrices de covarianza. Estas no son descritas a través

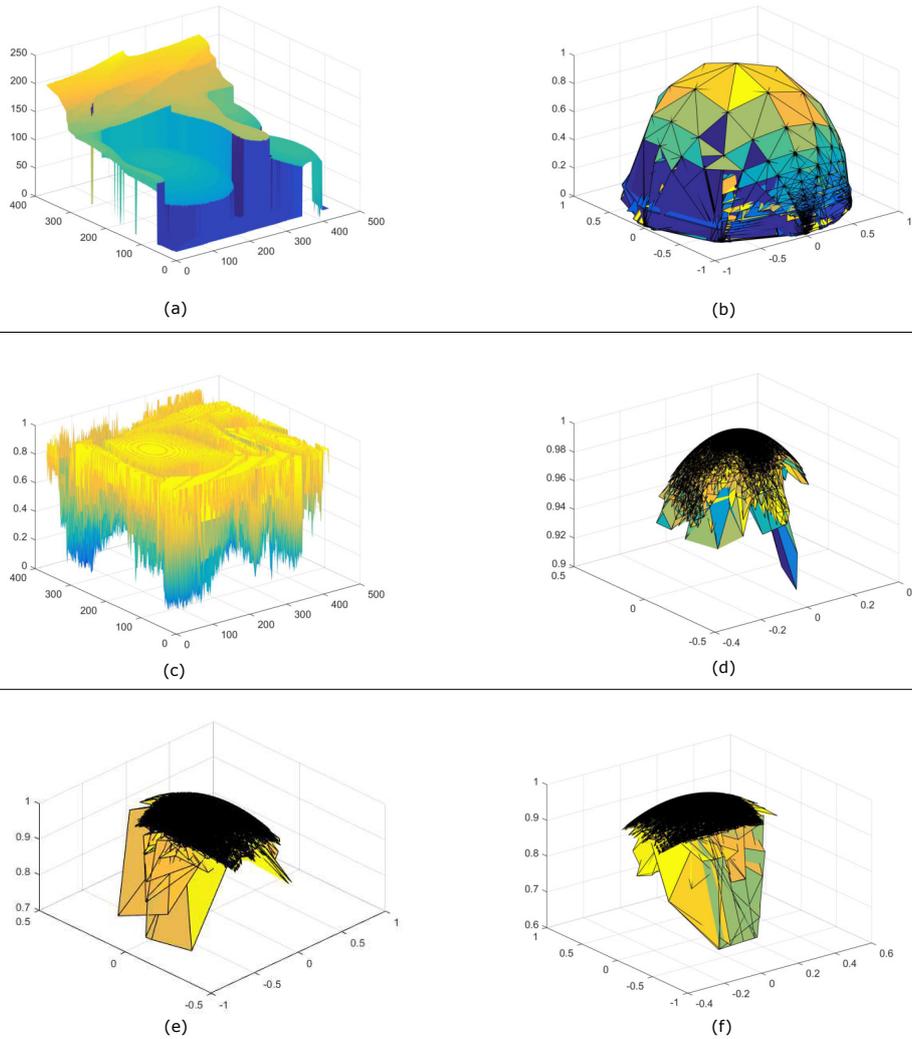


Figura 3.2: (a) Superficie imagen de rango. En (b) la figura de (a) es representada como una semiesfera en el espacio  $\mathbb{R}^3$ . (c) superficie generada de la imagen de homogeneidad y (d) curvatura correspondiente a dicha característica; e) y f) imágenes de la división en las direcciones  $x$  y  $y$  de la imagen del inciso (d).

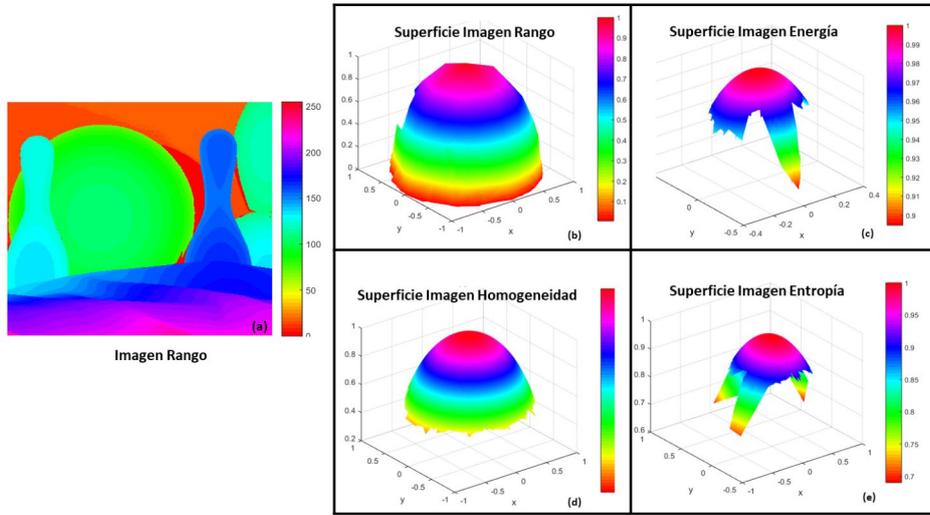


Figura 3.3: De a) la imagen de rango (Bowling1 [58]) se extraen los descriptores que forman las superficies correspondiente a las características de c) energía; d) entropía y e) homogeneidad. Mientras que en b) se muestra la superficie formada por (a).

de un espacio Euclidiano sino que, son conformadas como una variedad de tensores formados por SPDs.

En relación con este tópico, en este trabajo de Tesis se propone primero que, las imágenes de rango pueden describir un comportamiento similar al que se presenta en las superficies lambertianas en cada “superficie-capa” con las que se forman estas imágenes. La idea de describir espacios tangenciales con base en este tipo de superficies nos permite como lo plantea Jeon *et al.* en [67], seleccionar puntos en una región con una varianza más suave.

Segundo, consideramos que la información obtenida de calcular los descriptores de textura apunta a generar superficies que se pueden considerar como mínimas. En relación con la definición de superficies mínimas, se toma como referencia lo expuesto por Meeks y Pérez en [98] quienes establecen que, “Si  $\mathcal{M} \subset \mathbb{R}^3$  es una superficie mínima completa, inmersa y estable, entonces  $\mathcal{M}$  es un plano”. Esto es debido a que, la técnica propuesta para el cálculo de dichos descriptores asegura solamente la información más relevante de la imagen de rango, de acuerdo con el descriptor calculado; con lo cual se preserva la curvatura Gaussiana de la superficie original (ver figura,3.2).

Por lo tanto, se propone en este trabajo que, si la curvatura de la superficie se suaviza por medio de las características de textura es posible formar un plano proyectivo (segundo renglón figura3.2 c-d). Más aún, este no será solamente un plano Euclidiano con sus respectivas propiedades isotrópicas (propiedad a la *invarianza geométrica*), sino además, adopta las propiedades correspondientes a un manifold [8].

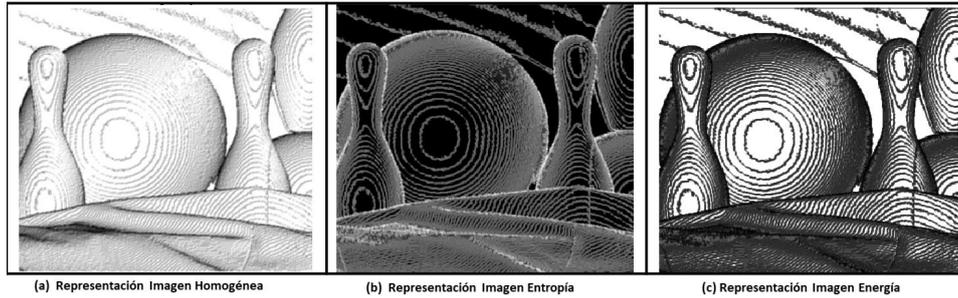


Figura 3.4: Ejemplo de la representación como imágenes de algunas características de textura tal como a) homogeneidad, b) entropía y c) energía obtenidas a partir de la imagen de rango (Bowling1[58]) representadas en imágenes formadas por medio de isolíneas.

La razón de lo anterior es, la minimización de la energía, que se presenta entre los descriptores con respecto de la superficie formada por la imagen de rango. Por ejemplo, en 3.3b la forma de la superficie es mucho más cercana a 3.3e; ya que, la superficie es formada por el descriptor de entropía. En este caso, la minimización de la energía es muy alta. Un ejemplo, de las superficies resultantes para cada descriptor es mostrado en 3.3b-d. La barra de color en cada imagen de la figura 3.3 nos indica, las diferentes intensidades en niveles de gris; las cuales están presentes en las figuras originales (ver, 3.2a y 3.3a).

Finalmente, se sugiere que, la extracción de las características de textura permite la formación de curvas geodésicas (ver, figura 3.4), con las cuales se destaca el contorno entre las diferentes regiones [128] correspondientes a los diferentes niveles en la imagen de rango. En este caso, la representación de dichas curvas es similar a “dibujar” líneas de isoelevación (líneas que, unen los puntos de igual altura), usualmente empleadas en mapas topográficos.

Por lo tanto, parte fundamental de esta propuesta, estriba en dividir direccionalmente la superficie formada por la imagen de rango, las propiedades geométricas de la curvatura Gaussiana [56, 116] contenidas en  $\mathbb{R}^3$  son transformadas a un plano en  $\mathbb{R}^2$ . Esta transformación es ejemplificada en la figura 3.2, en las direcciones e)  $x$  y f)  $y$  así como, a través de la ecuación 3.1. En dichas figuras, la parte sombreada en oscuro muestra la posible transformación o suavizado de la curvatura en un plano. Por otro lado, en la ecuación 3.1 se presenta, el mapeo que se lleva a cabo, entre los planos formados en cada dirección hacia la imagen de rango.

$$[M_x, M_y, M_z] : M_{texture} \mapsto M_{range} \quad (3.1)$$

donde,  $M_x, M_y, M_z, M_{texture} \subset \mathbb{R}^2$  y  $M_{range} \subset \mathbb{R}^3$ . Esta última proposición apunta a la reconstrucción que se puede llevar a cabo, a partir de la información contenida en la  $M_{range}$ . Todas estas superficies apuntan a imágenes

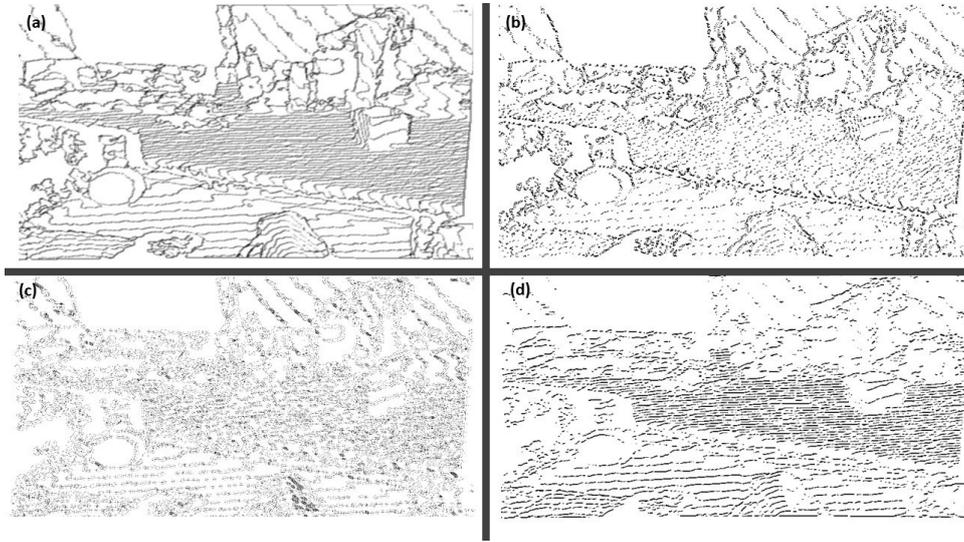


Figura 3.5: Ejemplo de la representación de la característica de textura homogeneidad como imagen ( $M_{homogeneity}$ ) correspondiente a la  $M_{range\ cubito4}$  de nuestra base de datos DSPLab. La división de a) superficie de la imagen de homogeneidad, en superficies direccionales se muestran en b)  $M_x$ , c)  $M_y$  y d)  $M_z$ .

de un mismo tamaño  $K \times L$ .

En este trabajo se designa a  $M_x$ ,  $M_y$  y  $M_z$  como las superficies en las direcciones  $x$ ,  $y$  y  $z$ , respectivamente. Mientras que,  $M_{texture}$  hace referencia a la superficie formada por medio de los descriptores de textura. Por ejemplo, si tenemos  $M_{homogeneity}$  hará referencia a la superficie obtenida a través del descriptor de homogeneidad y así, sucesivamente para cada característica.

La figura 3.5 muestra un ejemplo de las imágenes generadas como resultado de llevar a cabo el proceso de dividir la  $M_{texture}$ , en sus correspondientes superficies direccionales. Es de notar, cómo se preserva la información de la  $M_{texture}$  en cada una de las direcciones  $x$ ,  $y$  y  $z$ ; cuyas imágenes son una representación de sus respectivas superficies  $M_x$ ,  $M_y$  y  $M_z$  en  $\mathbb{R}^2$  (ver, figura 3.5b-d).

### 3.2. Proceso Denso

Si un descriptor representa el patrón de la imagen, en nuestra propuesta, dicha representación es realizada haciendo uso de una aproximación similar a un análisis de textura, como se propone en 2.2. Los resultados obtenidos de la adaptación de esta técnica, nos permiten dar inicio al cálculo y análisis de los puntos de interés y sus vecindades.

La primera fase, se ilustra en el esquema de la figura 3.6. En el mismo, se

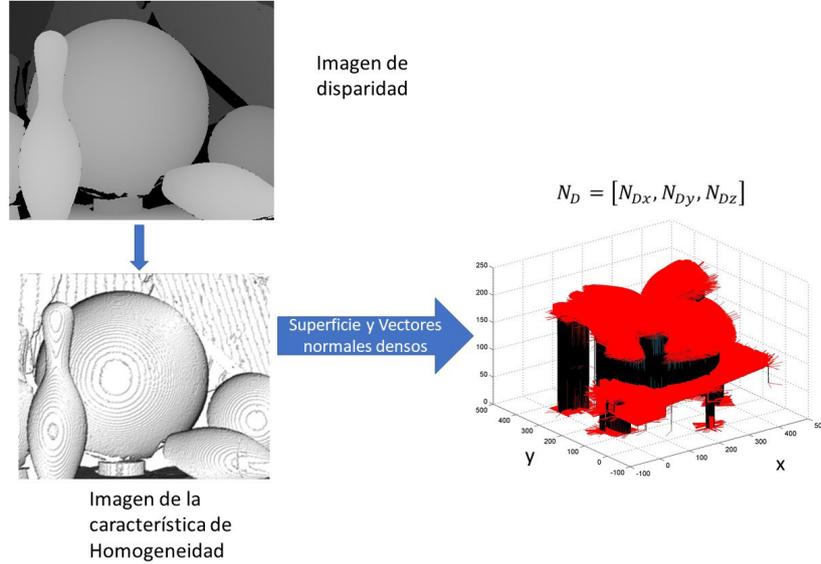


Figura 3.6: Diagrama esquemático de la primera fase (proceso *denso*) 3.2 de la técnica propuesta.

observa que a partir de la imagen de disparidad se calcula la característica de textura tal como, en este ejemplo, lo es la homogeneidad. Los descriptores calculados son representados por medio de una imagen y de esta se obtiene la superficie correspondiente, a la cual se le denominará *superficie densa*.

Adicionalmente, de acuerdo con lo revisado por Loncomilla en [89] así como, Kuo *et al.* en [75] en el presente trabajo se establece que, una característica local es el patrón de una imagen, la cual depende de su vecindad inmediata [148]. La posición referente a cada punto característico es proyectada con respecto de la ubicación que le corresponde al vector tangente, calculado a partir de la superficie formada por la imagen de textura.

Los vectores normales son computados después de realizar un ajuste bicúbico de los datos en las direcciones  $x$ ,  $y$  y  $z$ . Así, para cada punto, se realiza un producto cruz simple haciendo uso de los vectores tangentes calculados en cada dirección; con el fin de generar el vector normal correspondiente al vértice [31]. Por tanto, de acuerdo con la ecuación 3.2 el vector  $N_D$  como fue nombrado, representará los vectores contenidos en la superficie densa (diagrama ilustrativo, figura 3.7).

$$N_D = [N_{D_x}, N_{D_y}, N_{D_z}] \quad (3.2)$$

donde,  $N_{D_x}, N_{D_y}, N_{D_z} \in \mathbb{R}^2$  y de tamaño  $K \times L$ , el cual corresponde a la

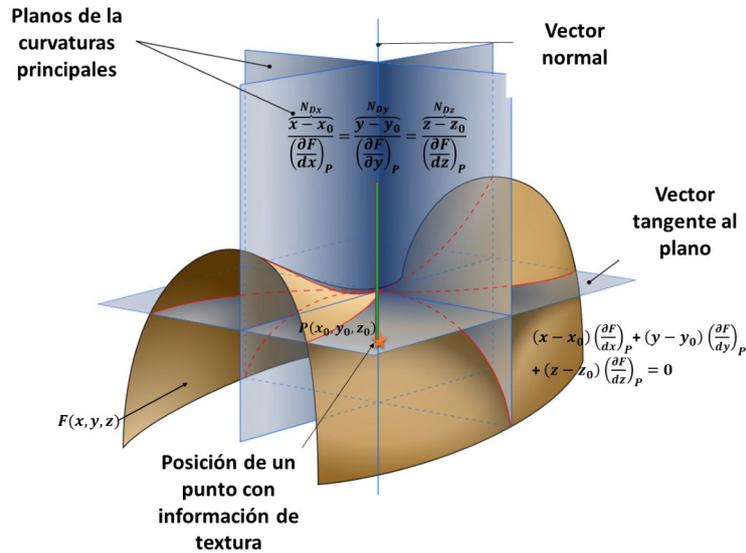


Figura 3.7: Diagrama representativo [139] de la formación de los vectores normales densos en la superficie de textura, representada por la función en forma implícita  $F(x, y, z) = 0$ .

misma dimensión que tiene la imagen de rango. El computo del vector  $N_D$  marca el final del proceso denso para dar paso al proceso disperso que será descrito en la siguiente sección.

### 3.3. Proceso Disperso

#### 3.3.1. Consideraciones de Matrices Dispersas

Para fines de esta Tesis, se propone primero, trabajar con matrices dispersas no estructuradas tal como se revisó en la subsección 1.5.3; segundo, llevar a cabo un almacenamiento de la información dispersa, bajo el esquema conocido como *formato coordinado* (COO, por sus siglas en inglés), el cual es el más sencillo de implementar según lo expuesto por Jaramillo *et al.* en [65]. Un ejemplo de este tipo de almacenamiento es el siguiente,

$$A = \begin{pmatrix} 0 & 0 & a & 0 & b & 0 \\ 0 & c & 0 & d & 0 & 0 \\ 0 & e & 0 & 0 & 0 & 0 \\ 0 & f & 0 & 0 & 0 & 0 \\ 0 & 0 & g & 0 & h & 0 \\ i & 0 & 0 & 0 & 0 & j \end{pmatrix} \quad \begin{aligned} AA &= [e & a & f & b & i & g & h & c & j & d] \\ AI &= [3 & 1 & 4 & 1 & 6 & 5 & 5 & 2 & 6 & 2] \\ AJ &= [2 & 3 & 2 & 5 & 1 & 3 & 5 & 2 & 6 & 4] \end{aligned}$$

de acuerdo con el ejemplo, el vector  $AA$  contiene los elementos *no-nulos* en forma dispersa.  $AI$  corresponderá al vector que contiene la posición respecto de la fila de dicho elemento, mientras que,  $AJ$  será para los índices correspondientes a la columna. En nuestra propuesta, el recorrido de la matriz será el convencional, de izquierda a derecha y de arriba hacia abajo. Por lo que, el orden de los vectores será,

$$\begin{aligned} AA &= [a \ b \ c \ d \ e \ f \ g \ h \ i \ j] \\ AI &= [1 \ 1 \ 2 \ 2 \ 3 \ 4 \ 5 \ 5 \ 1 \ 6] \\ AJ &= [3 \ 5 \ 2 \ 4 \ 2 \ 2 \ 3 \ 5 \ 6 \ 6] \end{aligned}$$

Al finalizar el proceso denso se lleva a cabo la selección de la información que se procesará en el disperso. Dicha selección genera matrices con un formato similar al de la matriz  $A$ . Tal que, el tratamiento entonces de dicha información es a través de un esquema COO.

### 3.3.2. Selección de la Información para el Proceso Disperso

El siguiente paso es llevar a cabo la división de la superficie densa, en las direcciones  $x$ ,  $y$ , y  $z$ , respectivamente. Esto es, nuevas superficies direccionales son generadas a partir de la información contenida en  $N_{D_x}$ ,  $N_{D_y}$  y  $N_{D_z}$  de acuerdo con lo establecido en la sección previa.

La figura 3.8 muestra un ejemplo de las imágenes generadas como resultado de llevar a cabo el proceso de dividir la  $M_{textura}$ , en sus correspondientes superficies direccionales. Es de notar, cómo se preserva la información de la  $M_{textura}$  en cada una de las direcciones  $x$ ,  $y$  y  $z$ ; esto es porque al llevarse a cabo una proyección del vector, la información correspondiente al descriptor de textura, también se “proyectará” a la correspondiente dirección.

Ahora bien, para llevar a cabo este proceso se establece en la ecuación 3.3 que sólo los vectores que apuntan hacia fuera de la superficie densa serán seleccionados. Por lo tanto, se establece que

$$\begin{aligned} N_{D_x} &\mapsto M_x, N_{D_x} > 0 \\ N_{D_y} &\mapsto M_y, N_{D_y} > 0 \\ N_{D_z} &\mapsto M_z, N_{D_z} > 0 \end{aligned} \tag{3.3}$$

un ejemplo ilustrativo de los resultados obtenidos descritos por la ecuación 3.3 se muestran en la figura 3.8. En la misma figura se presenta en el primer renglón las superficies direccionales en las que los vectores normales son representados de color rojo. En particular, se puede observar con respecto de las direcciones  $x$  y  $y$  los diferentes niveles que se forman en la superficie. Esto es debido a que, la información de la característica de textura en estas dos direcciones presenta valores disímiles, lo que apunta a la formación de estos niveles. Situación que no se presenta en la dirección en  $z$ . Esto es,

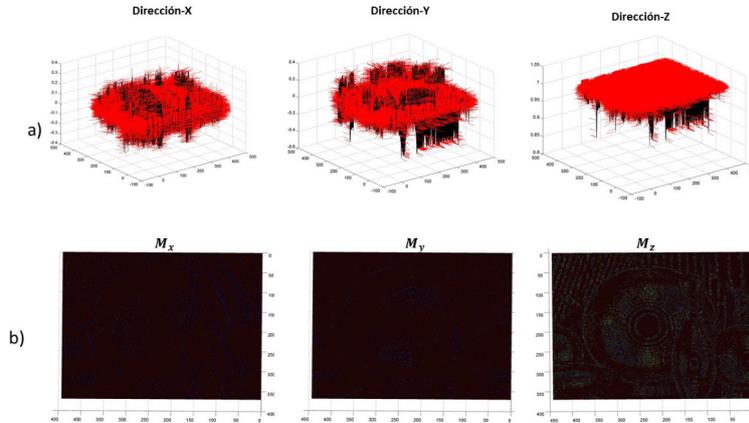


Figura 3.8: Diagrama esquemático que ilustra el inicio del proceso disperso. a) Representación de los vectores contenidos en las superficies direccionales  $M_x$ ,  $M_y$  y  $M_z$ . b) Representación en  $\mathbb{R}^2$  de la posición de los vectores en estas superficies.

debido a que, en dicha dirección es dónde se encuentra concentrado los valores homogéneos de la textura.

Los vectores más sobresalientes enmarcados en la circunferencia amarilla de la figura 3.9, con respecto de las direcciones  $x$  y  $y$ , son los vectores que se consideran como puntos de interés. Más aún, la ubicación o posición de dichos vectores, en estas superficies ( $M_x$  y  $M_y$ , respectivamente) implica también, el considerar en su vecindad las características de textura correspondiente. Además, es importante hacer notar, que en la dirección en  $z$ , prácticamente, todos los vectores pueden ser considerados, ya que no existen valores disímiles en esta dirección.

En la siguiente subsección, detallaremos el pseudocódigo desarrollado para la implementación general de la técnica, la cual es nombrada Detector de Puntos de interés Dispersos o Sparse Keypoint Detection (SKD, por sus siglas en inglés).

### 3.3.3. Pseudocódigo de la técnica SKD

A fin de describir el pseudocódigo para detectar puntos de interés dispersos, se tomará en cuenta la siguiente nomenclatura, sea  $I_R$  la imagen de rango,  $H_m$  la imagen de homogeneidad,  $N_D$  vectores normales densos de  $I_R$ ,

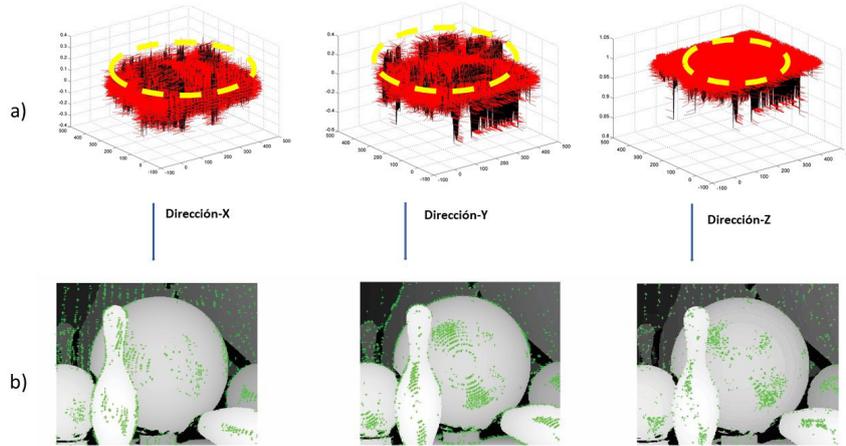


Figura 3.9: Diagrama representativo a) de los vectores más sobresalientes en cada dirección, b) para obtener puntos de interés.

$N_S$  normales dispersas:

---

**Algoritmo 1: Técnica para detectar puntos de interés dispersos.**

---

**Inicio**

**Entrada:**  $I_R$

**Salida:**  $N_S$

1. **Para** todos los pixels en  $I_R$  **Hacer**
2. Calcular  $H_m$
3. **Fin del Para**
4. **Para** todos los pixels en  $H_m$  **Hacer**
5. Calcular las normales densas  $\mathbf{N}_D = [\mathbf{N}_{Dx}, \mathbf{N}_{Dy}, \mathbf{N}_{Dz}]$
6. **Fin del Para**
7. Calcular umbrales  $t_x$  y  $t_y$  desde  $\mu(\mathbf{N}_D) \pm \sigma(\mathbf{N}_D)$ , en dirección **x**- y **y**-
8. Calcular umbral  $t_z$ , desde  $\mu(\mathbf{N}_{Dz})$  o  $\min(\mathbf{N}_{Dz})$ .
9. **Para** todas las normales en  $\mathbf{N}_D$  **Hacer**

10. Calcular vector normal disperso  $N_S$ :
  - a) Generar vectores normales dispersos  $[N_{S_x}, N_{S_y}, N_{S_z}]$  usando  $\mathbf{t} = [t_x, t_y, t_z]$
  - b) **Sí**  $[N_{D_x}, N_{D_y}, N_{D_z}] > 0$  y  $N_D > \mathbf{t}$  **Entonces**  
 $N_D \rightarrow N_{St}$
  - c) **Fin del Para**
11. Ubicar las posiciones de los vectores  $N_{St}$  en  $M_x, M_y$  y  $M_z$
12. Detectar puntos característicos contenidos en  $M_x, M_y$  y  $M_z$ 
  - a) Calcular umbral  $t_k$
  - b) **Sí**  $N_{St} > t_k$  **Entonces**  
 $N_{St} \rightarrow N_S$

**Fin**

---

La figura 3.10 ilustra los pasos 1 a 6 pasos correspondientes a dicho pseudocódigo. Estos hacen referencia al proceso denso que se revisó en la sección 3.2. Como se mencionó en dicha sección, el vector  $N_D$  contiene los vectores densos de la superficie formada por la imagen de textura (ver imagen 3.10). Los valores contenidos por el vector  $N_D$  se encuentran dentro del rango de  $[-1, 1]$ .

Por otro lado, si las superficies formadas por las características de textura corresponden a una semiesfera positiva, tal como se presentó en 3.1; entonces, solamente los vectores que se encuentran “hacia afuera” de dicha superficie son los que serán seleccionados para formar el *vector normal disperso* aquí, nombrado  $N_S$ . Esta primera discriminación da inicio al proceso del mismo nombre. Por lo tanto,  $N_S = [N_{S_x}, N_{S_y}, N_{S_z}]$  es definido por la ecuación 3.4, en la cual se considera que el rango de los valores para  $N_S$  se encuentra entre  $(0, 1]$ .

$$N_S = \{N_D > t | 0 < t < 1\} \quad (3.4)$$

En los pasos 7 a 10 se lleva a cabo un proceso de selección. Este, es realizado por medio de un análisis de la distribución Gaussiana de los vectores densos, en la superficie formada por la imagen de la característica de textura. Esta selección es descrita por medio de la desigualdad  $N_{Di} > \mathbf{t}$  donde,  $i$  es el  $i$ -ésimo componente de  $N_D$ . El vector  $\mathbf{t}$  contiene los tres niveles de umbral en el rango entre  $0 < \mathbf{t} < 1$ ; los cuales son llamados  $t_x, t_y$  y  $t_z$ , respectivamente. Por tanto, dichos umbrales son calculados haciendo uso de la información estadística del vector  $N_D$ ,

$$\left. \begin{array}{l} \mu(N_{D_x}) + \sigma(N_{D_x}) \quad \text{si} \quad 0 < N_{D_x} < t_u \\ \mu(N_{D_x}) - \sigma(N_{D_x}) \quad \text{si} \quad t_u \leq N_{D_x} < 1 \end{array} \right\} = t_x \quad (3.5)$$

$$\left. \begin{array}{l} \mu(N_{D_y}) + \sigma(N_{D_y}) \quad \text{si } 0 < N_{D_y} < \mathbf{t}_u \\ \mu(N_{D_y}) - \sigma(N_{D_y}) \quad \text{si } \mathbf{t}_u \leq N_{D_y} < 1 \end{array} \right\} = \mathbf{t}_y \quad (3.6)$$

$$\left. \begin{array}{l} \mu(N_{D_z}) \quad \text{si } \max(N_{D_z}) < 1 \\ \mu(N_{D_z}) - \sigma(N_{D_z}) \quad \text{si } \max(N_{D_z}) = 1 \end{array} \right\} = \mathbf{t}_z \quad (3.7)$$

donde,  $\mathbf{t}_u$  es un nivel de umbral definido dentro del 95 % del intervalo de confianza establecido a través de la ecuación 3.8, concerniente a los componentes de  $N_D$  (ecuación 3.2).

$$\begin{aligned} P(-z \leq Z \leq z) &= 1 - \alpha = 0,95 \\ 0,95 &= P(-1,96 \leq \frac{\overline{N_D} - \mu}{\sigma/\sqrt{n}} \leq 1,96) \end{aligned} \quad (3.8)$$

De la ecuación 3.8 se considera a  $n$  como el número de elementos contenidos en  $N_D$ . Tal que,  $\mathbf{t}_u = [\mathbf{t}_{ux}, \mathbf{t}_{uy}]$ , donde

$$\left. \begin{array}{l} \overline{N_D} + 1,96 \frac{\sigma}{\sqrt{n}} = \text{extremo superior} \\ \overline{N_D} - 1,96 \frac{\sigma}{\sqrt{n}} = \text{extremo inferior} \end{array} \right\} = \mathbf{t}_u \quad (3.9)$$

por lo tanto, la ecuación 3.9 debe ser aplicada en cada una de las direcciones correspondientes. Tal que, el cómputo de los umbrales estimados, por medio de las ecuaciones 3.5 y 3.6 están en función del valor correspondiente a  $\mathbf{t}_u$ . Por lo anterior, se asegura que la información contenida por  $N_S$  sea la que represente de forma cualificada, la superficie propia a la imagen de la característica de textura. Con respecto de la dirección en  $z$ , en la generalidad todos los valores son muy cercanos a 1, por lo que, el proceso de selección en relación con las otras dos direcciones es casi imperceptible. Más aún, una de las contribuciones de este trabajo, es el número de elementos contenidos en el vector  $N_S$ , el cual está por debajo del 80 % en promedio, con respecto de la cardinalidad (*card*) calculada para el vector  $N_D$ .

$$\text{card}(N_S) \lll \text{card}(N_D) \quad (3.10)$$

por lo tanto,

$$\begin{aligned} \text{card}(N_{Sx}) &\lll \text{card}(N_{Dx}) \\ \text{card}(N_{Sy}) &\lll \text{card}(N_{Dy}) \\ \text{card}(N_{Sz}) &\ll \text{card}(N_{Dz}) \end{aligned} \quad (3.11)$$

La ecuación 3.10 aplicada en las direcciones  $x$  y  $y$  correspondientes a la ecuación 3.11 nos presenta una reducción en el contenido de elementos en los vectores cercano al 90 %, mientras que en la dirección  $z$ , esta será cercana al 30 %.

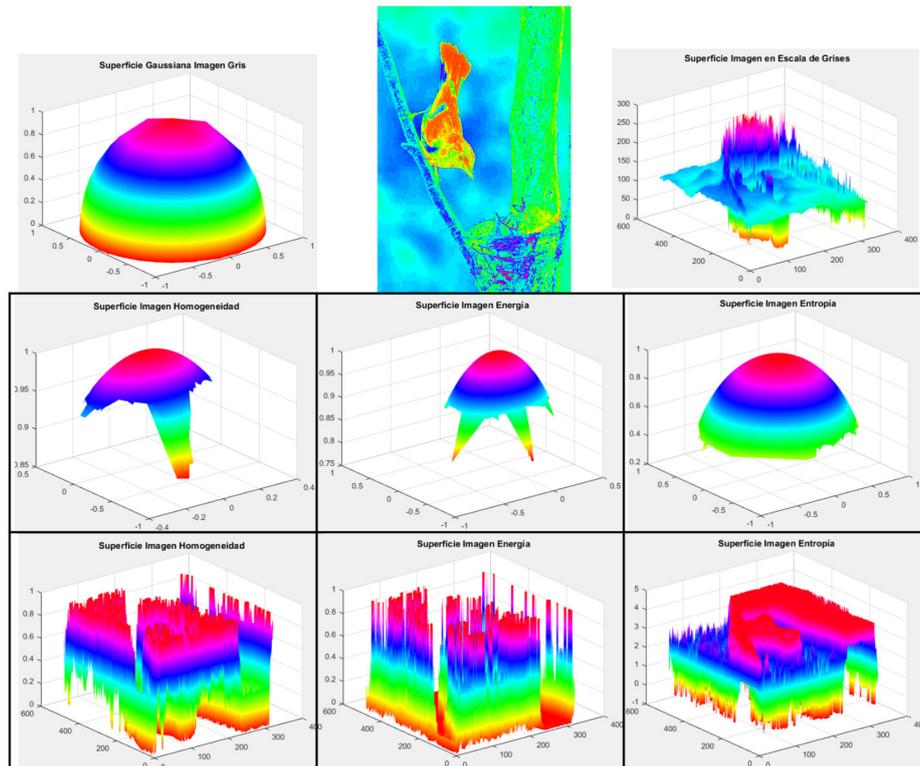


Figura 3.10: Ilustración de los pasos 1 a 6 del algoritmo 1 de la técnica SKD aplicada a la imagen bird (en niveles de gris representada en hsv) de base de datos conocida. A su izquierda su representación en su forma Gaussiana y a la derecha como superficie. Segundo renglón representación de la forma Gaussiana de diferentes descriptores de textura. Tercero, superficie densa de la que se seleccionara la información para el proceso disperso.

La diferencia significativa entre los diversos valores de la cardinalidad, en las respectivas direcciones, se debe principalmente a las estimaciones de las características de textura, con las cuales se forma la superficie densa. Por lo que, las superficies direccionales  $M_x$ ,  $M_y$  y  $M_z$  son representadas a través de los vectores dispersos contenidos como se mencionó, en  $N_S$ .

De lo anterior, en el paso 11 se determinan las posiciones de los datos contenidos en  $N_{Sx}$ ,  $N_{Sy}$  y  $N_{Sz}$ ; con respecto de las superficies direccionales  $M_x$ ,  $M_y$  y  $M_z$ . Esto es, con el fin de preservar la información contenida, con respecto a las características de textura, en la superficie direccional correspondiente. Esto implica que, por posición se cuenta tanto con el vector normal que describe la superficie, así como su correspondiente valor de la característica de textura. Por lo que, se conforma un vector de descriptores  $[\vec{N}_S, M_{[x,y,z]}]$  del cual, nos referiremos como el vector  $N_S$  por el hecho, de que es a través de la posición de este, que ubicamos los otros descriptores.

Por otro lado, en el paso 12, el cómputo de los puntos de interés se realiza mediante un análisis de una distribución discreta uniforme de probabilidad. Esto es porque, se asume que todos los descriptores contenidos en las superficies direccionales tienen la misma probabilidad de ser un punto de interés. Partiendo de este supuesto, es entonces que se debe de establecer un parámetro de discriminación, para llevar a cabo un proceso de selección. Por lo que estará sujeto al cálculo de

$$\mu = \frac{1}{C_{S[x,y,z]}} \sum_{i=1}^{C_{S[x,y,z]}} N_{Si} \quad (3.12)$$

$$\sigma^2 = \frac{1}{C_{S[x,y,z]}} \sum_{i=1}^{C_{S[x,y,z]}} (N_{Si} - \mu)^2 \quad (3.13)$$

donde  $C_{S[x,y,z]}$  hace referencia a la cardinalidad de los descriptores contenidos en cada una de las superficies direccionales; mientras que  $N_{Si}$  corresponde al  $i$ -ésimo elemento del vector  $N_S$ .

Es importante hacer notar que cada estimación propuesta por las ecuaciones 3.12 y 3.13 es llevada a cabo en cada una de las superficies direccionales. Por lo que, haciendo uso de dichas ecuaciones podemos establecer un umbral en relación con el valor de los descriptores. Este umbral llamado  $t_k$  a diferencia del primer umbral  $t_u$ , calculado en 3.9 (calculado para la superficie densa), permite seleccionar los descriptores dispersos contenidos en cada superficie direccional. Por lo tanto, como lo mencionan los pasos 12.a y 12.b

$$\mathbf{t}_{\mathbf{k}[x,y,z]} = \mu + \sigma^2 \quad (3.14)$$

$$\text{Puntos de interés} = N_S > \mathbf{t}_{\mathbf{k}[x,y,z]} \quad (3.15)$$

a través de estos últimos pasos marcados por medio de las ecuaciones 3.14 y 3.15, es como finalmente se obtienen los puntos de interés contenidos en las

superficies direccionales, los cuales apuntan a los objetos contenidos en la escena. Una representación de éstos, lo podemos observar de forma cualitativa en la figura 5.5.

### 3.3.4. Algoritmo para la detección de un objeto en la ROI

Una modificación propuesta al algoritmo 1, es la que presentamos a continuación. El propósito de la modificación es para llevar a cabo una serie de procesos que eficienten los tiempos de cómputo. Además, deberá ser posible identificar una región de interés conocida como ROI (por sus siglas en inglés) para determinar la ubicación del objeto. La figura 3.11 resume el algoritmo propuesto.

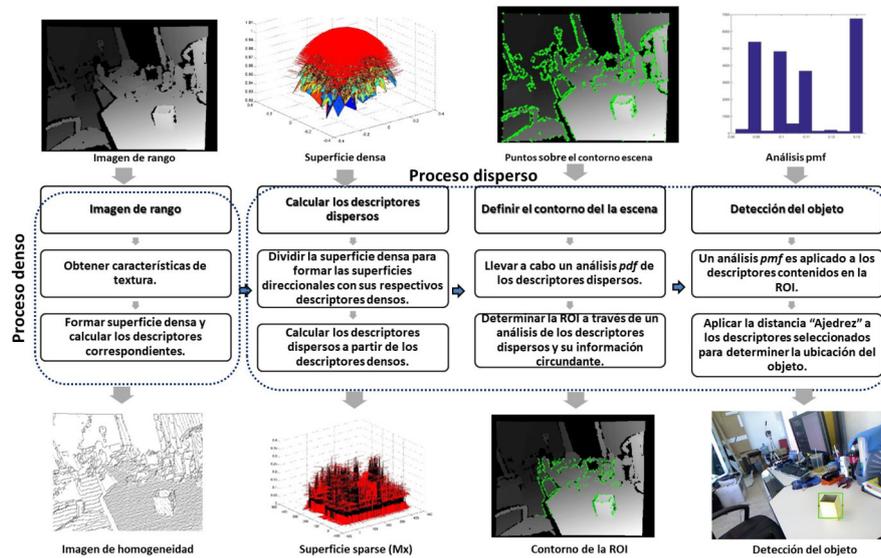


Figura 3.11: Diagrama general de la técnica SKD aplicada a la detección de un objeto ubicado en una ROI.

La figura 3.11 resume el algoritmo propuesto. La siguiente nomenclatura es utilizada para describir el algoritmo:  $I_R$  la imagen de rango,  $H_m$  la imagen de homogeneidad,  $N_D$  vectores normales densos de  $I_R$ ,  $N_S$  normales dispersas:

---

**Algoritmo 2. Pseudocódigo de la técnica SKD para la detección de un objeto en una ROI.**

---

#### Inicio

1. Proceso disperso

- a) Calcular los descriptores dispersos ( $N_S$ ):
    - 1)  $N_S = [N_D \geq \mathbf{t}_u | 0 < \mathbf{t}_u \leq 1]$  donde,  $\mathbf{t}_u$  es definido por 3.9
    - 2)  $N_S = [N_{S_x}, N_{S_y}, N_{S_z}]$  donde, 3.10
  - b) Calcular posiciones de los vectores  $\vec{N}_S$  en  $\mathbf{M}_x, \mathbf{M}_y$  y  $\mathbf{M}_z$
2. Detectar los puntos de interés del contorno de la escena:
- a) **Si**  $\text{card}(M_x \text{ descriptores}) < \text{card}(M_y \text{ descriptores})$   
seleccionar  $M_x$   
**Si No Entonces** seleccionar  $M_y$ ;  
**Fin del Sí**
  - b)  $S_K = N_S > \mathbf{t}_k$ , donde  $\mathbf{t}_k = \text{pdf}(N_S)$
3. Detectar objeto
- a) **Para** cada  $(k, l)$  en  $\mathbf{S}_K$ , **Hacer**
    - 1) calcular la posición de la **ROI** usando una pdf
  - b) **Fin del Para**
    - 1) Calcular las adyacencias de los  $\mathbf{S}_K \rightarrow \text{adyacencia}(S_k)$
    - 2) Definir la **ROI** con respecto de una:  $\min(\text{adyacencia}(S_k))$  y  $\max(\text{adyacencia}(S_k))$
  - c) Puntos de interés en el contorno de la **ROI** =  $\text{jpmf}([\vec{N}_S, M_{[xoy]}])$  de  $\mathbf{S}_K$ .
  - d) Calcular la distancia *Chess* ( $\mathbf{D}_{\text{Chess}}$ ) de los descriptores de la **ROI**.
  - e) Posición de objeto:  $\min(\mathbf{D}_{\text{Chess}})$

**Fin**

---

Los pasos 1 y 2 son los correspondientes al algoritmo 1, presentados en la sección 3.3.3. Por lo que, en el paso 3, se lleva a cabo el proceso de selección de la superficie direccional. Esta elección es una comparación de la cardinalidad (*card*), de los descriptores contenidos en cada una de las superficies  $M_x$  y  $M_y$ . Posteriormente, a través de un análisis PDF y haciendo uso de las ecuaciones 3.14 y 3.15 aplicado a los descriptores contenidos en la superficie elegida ( $M_x$  o  $M_y$ ), se determina los puntos de interés  $\mathbf{S}_K$  donde,  $\mathbf{S}_K \in \mathbb{R}^2$ .

$$\mathbf{S}_K = N_{S_x} > \mathbf{t}_k, \text{ para } M_x \quad || \quad \mathbf{S}_K = N_{S_y} > \mathbf{t}_k, \text{ para } M_y \quad (3.16)$$

La ecuación 3.16 generalizada en el paso 3.b, nos permite considerar los descriptores correspondientes a todos los puntos  $\mathbf{S}_K$ , como variables aleatorias discretas. Más aún, la representación en la superficie dispersa para cada punto  $\mathbf{S}_K$ , se lleva a cabo mediante la posición  $(k_i, l_i)$ , correspondiente a cada vector normal. Dicha posición es descrita como:

$$\{(k_i, l_i) | k_i \in K, l_i \in L\} \quad (3.17)$$

dónde,  $K = \{k_1, k_2, \dots\}$  y  $L = \{l_1, l_2, \dots\}$  e  $i = \{1, \dots, \text{card}(M_{[x,y]})\}$ . Dónde, la cardinalidad marcada para  $i$  corresponde a la calculada por la ecuación 3.11.

Para dar inicio con la detección del objeto contenido en la ROI, lo primero es definir esta. En el paso 4.a se realiza el proceso de selección de cada una de las coordenadas  $(k_i, l_i)$  pertenecientes a la posición de los puntos de interés. Para ello, aplicamos primero un análisis estadístico de densidad a la información contenida en  $K$  y otro similar para  $L$ . Esto con el fin de elegir las posiciones de los descriptores que conformarán la región de interés.

Adicionalmente, de los resultados obtenidos del análisis PDF se establece la conectividad existente entre las coordenadas con el fin de validar la pertenencia o no, con respecto de los diferentes objetos o áreas en la región de interés. En la figura 3.12 se ilustra una máscara de  $3 \times 3$  con la que se realiza el recorrido de la ROI definida en primer instancia con todos los puntos seleccionados; con el fin de evaluar la conectividad de cada coordenada con respecto de las demás. Por tanto, cada posición  $(k_i, l_i)$  pasa a ser el punto central de la máscara. Esta conectividad o *conectividad-m* (mixed-connectivity, en inglés), nos permite determinar si un descriptor en una posición  $(k_i, l_i)$  en particular pertenece o no, a una determinada región dentro de la ROI. Entonces, se establece que, una *mínima adyacencia* indica descriptores en una misma área; mientras que, una *máxima adyacencia* indicará una región diferente. Finalmente, dicha conectividad, permite determinar el radio ( $r$ ) correspondiente a la ROI. Esta distancia en lugar de formar un círculo con este radio, lo que se forma será un cuadrado de lado  $2r$ .

Posteriormente, en el paso 4.b de la información resultante del paso 4.a se aplica un análisis JPMF a los descriptores que corresponden a los puntos de interés  $\mathbf{S}_K$  contenidos únicamente en la ROI. La importancia de esta evaluación es determinar los puntos de interés que pertenecen o están muy próximos del contorno de los objetos contenidos en esta área.

Finalmente, para llevar a cabo los pasos 4.c – d el procesamiento se inicia aplicando la distancia de *Chebyshev* discreta llamada  $D_{Chess}$ . La distancia de Chebyshev discreta se define como

$$D_{Chess} = \max(|k_i - k_j|, |l_i - l_j|) \quad (3.18)$$

la razón por la cual se hace uso de esta distancia es porque tiene un comportamiento similar, al de calcular el trayecto de los puntos de interés con conectividad-m. Por lo tanto, aplicando la ecuación 3.18 se calcula primero, la distancia máxima entre los puntos  $\mathbf{S}_K$  obtenidos en el paso 4.b. El objetivo es llevar a cabo dentro de la ROI una separación de los diferentes objetos contenidos en esta área. Por otro lado, habiendo realizado esta segmentación de objetos o áreas se propone que, al aplicar  $\min(D_{Chess})$  a las posiciones de los descriptores más destacados en esta región, se podrá detectar el objeto de interés. Este puede ser definido a través de los puntos que se ubican como

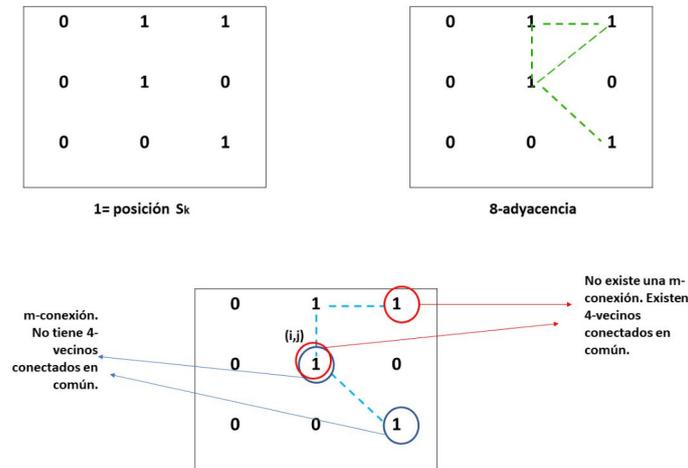


Figura 3.12: Ejemplo ilustrativo de la representación de la *conectividad-m* entre los puntos de interés ( $S_K$ ).

más próximos o, en el contorno del objeto.

Adicionalmente, la identificación de los posibles puntos de sujeción se puede plantear cómo el cálculo del centroide de la figura geométrica. Este centroide se “extiende” tanto al lado derecho como al izquierdo, para que dicho punto pueda ser relacionado con la posición más cercana a un punto de interés en cada lado. Aún cuando esto, no garantiza que dichas posiciones sean las óptimas se puede considerar como una primera aproximación.

### 3.3.5. Proceso de Validación

Es importante hacer notar que el número de puntos de interés computados no puede determinar la eficiencia de nuestro detector; para validar el desempeño de este con respecto de las técnicas ya conocidas se decidió hacer uso del proceso de validación presentado por Lowe y Muja en [106] y [91]. Aun cuando dicho proceso es bastante complejo, a través del esquema presentado en la figura 3.13 se presenta una simplificación de este.

De acuerdo con el esquema propuesto en 3.13, el cómputo de las características de entrada se realiza de acuerdo con el método que se haya seleccionado para realizar la detección y extracción de puntos.

Finalmente, en la última etapa se tiene que llevar a cabo una validación de acuerdo con el “peso probabilístico” asignado a cada punto determinado por el cómputo de sus características y análisis de su vecindad (circunscrito a un radio) lo que es realizado en la etapa de proceso. Lo anterior permite buscar

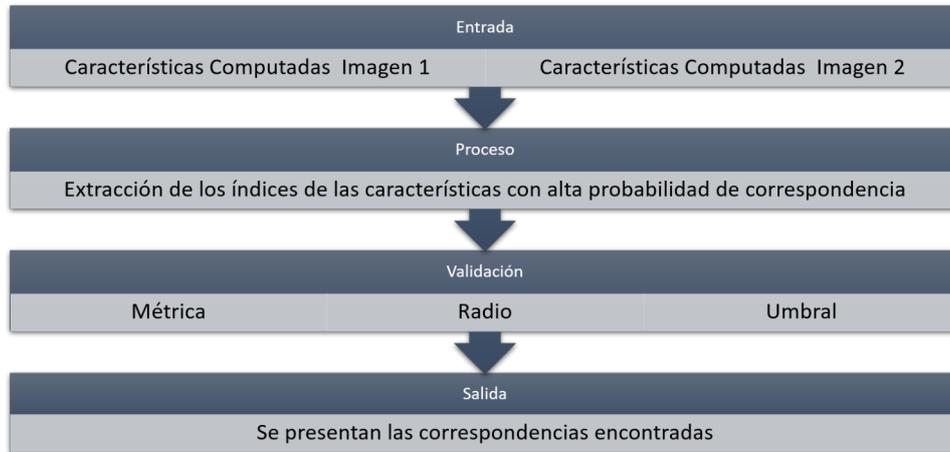


Figura 3.13: Diagrama esquemático para validar puntos de interés (con base en la propuesta de Muja y Lowe en [106] y [91]).

“pesos” similares en ambas imágenes con lo cual se determina el seguimiento o correspondencia entre los puntos de interés encontrados en estas. El umbral en este caso, es más un parámetro de sensibilidad para considerar si, el punto detectado con las características detectadas puede ser el mismo punto en ambas imágenes, esto con base en los descriptores analizados. La métrica empleada para establecer la distancia entre las características contenidas en la vecindad determinada es la suma de la diferencia de cuadrados (SSD, por sus siglas en inglés).

Es importante hacer notar que, una parte fundamental de este proceso es la localización de las características en la imagen; proceso que llevamos a cabo con nuestra técnica, tal como se presenta en los algoritmos descritos en las secciones 3.3.3 y 3.3.4. Puesto que, a partir de esta se establece un radio y un umbral con los que se indica el *tamaño* de la vecindad que será considerado para analizar y seleccionar un punto, como punto de interés. En nuestro caso, empleamos la métrica conocida como distancia de ajedrez.

Por otro lado, la precisión es calculada por la fórmula ya establecida,

$$\text{precisión} = \frac{\text{verdaderos positivos}}{(\text{verdaderos positivos} + \text{falsos positivos})}$$

En relación con el número de puntos calculados para cada una de las técnicas, se establece una relación sencilla,

$$Num_{puntos} = \frac{\#Puntos_{original} - \#Puntos_{ruido}}{\#Puntos_{original}} \times 100\% \quad (3.19)$$

Los resultados de la aplicación de lo aquí descrito es presentado en la subsección 5.2 del capítulo correspondiente.

## Capítulo 4

# Análisis de Entropía en las Imágenes de Rango

*“Life is a local order (spatial and temporal) within the process of increasing entropy, just as are the formation of galaxies and planets. Although they are not so much because energy goes in and out continuously.”*

Murray Gell-Mann

### 4.1. Entropía para Imágenes de Rango

Una imagen de rango es una muestra de la variabilidad en las medidas de distancia o disparidad obtenidas de un sistema visión 3D; donde dichas variaciones son expresadas a través de intensidades de gris. Así, una aproximación a la medida de entropía se considera como la incertidumbre a la distribución de dichas intensidades. Donde, una alta variabilidad significa un valor de entropía alto. Esta aproximación de la entropía solamente es considerada cuando las variables son independientes, como es el caso si se establece que cada medida es una variable aleatoria discreta.

La información contenida en la imagen de rango  $I_{Range}$  es considerada como una variable aleatoria discreta. En la cual, los valores o niveles de gris ( $\mathcal{G}$ ) son asociados, a un cierto valor de probabilidad  $p_i$ . Así pues, bajo condiciones de adquisición de la imagen similares se entenderá que, la probabilidad de que  $\mathcal{G}$  tome el valor de  $g_i$ , se encuentra dado por  $P\{\mathcal{G} = g_i\} = p_i$ . Por lo tanto, se establece que dicha probabilidad corresponde a la función de masa de probabilidad (*pmf*, por sus siglas en inglés), para la cual se determina que todos los eventos  $\{\mathcal{G} = g\}$  son mutuamente excluyentes. Así, la sumatoria de

todos los  $p'_i$ s es igual a 1 para todo el espacio muestral, esto es,  $\sum_{i=0}^n p_i = 1$ . Es de hacer notar que, al estar trabajando con una variable aleatoria discreta, esta sumatoria tenderá a 1, ya que, los valores de la  $I_{Range}$  son representados como una función discreta y no una continua. Donde,  $i = \{0, \dots, n\}$  y  $n = 255$ , este último valor corresponde al número máximo de intensidades o niveles de gris, que pueden presentarse en la  $I_{Range}$ .

Ahora bien, a fin de estimar una aproximación de la entropía se toma como referente la definición de Shannon [131], tal que,

$$H(\mathcal{G}) = -p_0 \log p_0 - \dots - p_n \log p_n \quad (4.1)$$

de acuerdo con la ecuación 4.1, una secuencia indexada de las variables aleatorias  $\{\mathcal{G}\} = \{g_1, \dots, g_n\}$ , tiene sus correspondientes valores de entropía definidos como

$$H(\mathcal{G}_0, \dots, \mathcal{G}_n) = - \sum_{i=0}^n p(g_0, \dots, g_n) \log p(g_0, \dots, g_n) \quad (4.2)$$

donde,  $p(\mathcal{G}_0, \dots, \mathcal{G}_n) = P(\mathcal{G}_0 = g_0, \mathcal{G}_1 = g_1, \dots, \mathcal{G}_n = g_n)$  es la probabilidad conjunta de todas las variables aleatorias discretas. Por lo que, al definir la entropía para una  $I_{Range}$  tenemos que,

$$H(I_{Range}) = - \sum_{i=0}^n p_i \log p_i \quad (4.3)$$

Es importante hacer notar que, para que sea válida la ecuación 4.3 existe una medida de similaridad correspondiente a una distribución de probabilidad con respecto a las variables aleatorias discretas, donde,  $F(g)$  es la función de distribución de una función masa de probabilidad (mpf, por sus siglas en inglés). Por tanto,  $F(g)$  es la probabilidad de que  $\mathcal{G}$  tome valores menores o iguales a  $g$ . Esto es,

$$F(g) = P\{\mathcal{G} \leq g\} = \sum_{g_i \leq g} P\{\mathcal{G} = g_i\} = \sum_{g_i \leq g} p_i \quad (4.4)$$

tal que, de la ecuación 4.4 se establece que  $P[\mathcal{G} = g] = F(g) - F(g-1)$ . Por tanto,  $p_i$  es presentada en la sección 4.2, como una aproximación de  $P(g)$ . Esta medida, se ha relacionado con el concepto de *entropía relativa* descrito por la divergencia de Kullback-Leibler en [74].

## 4.2. Entropía como Característica de Textura

Al considerar que las ecuaciones 2.12 y 2.13 definen la probabilidad de las sumas y diferencias como eventos independientes discretos; esta misma probabilidad puede ser considerada como una función de masa de probabilidad conjunta una *jmpf*. Por lo que, partiendo de la ecuación 2.11, tendremos que:

$$\begin{aligned} P_{sum}(p_0) &= \text{jmpf}\{S_{sum}(\varsigma_1)\} \\ \text{y,} & \\ P_{dif}(p_d) &= \text{jmpf}\{S_{dif}(\varsigma_2)\} \end{aligned} \quad (4.5)$$

Por lo tanto, si consideramos las funciones de distribución correspondientes para las sumas y diferencias, tendremos que, la  $p_i$  corresponde a  $p_0$  para las sumas, mientras que  $p_d$  para las diferencias. Donde,  $\{p_0, p_d\}$  corresponde a la probabilidad para las intensidades en los puntos  $(S_{k,l})$  y  $(S_{k+d_1, l+d_2})$  respectivamente, ver 2.7 de la sección 2.2.

$$\begin{aligned} F_{sum}(g) &= P_{sum}\{\mathcal{G} \leq g\} = \sum_{g_i \leq g} P_{sum}\{\mathcal{G} = g_i\} = \sum_{g_i \leq g} p_0 \\ \text{y,} & \\ F_{dif}(g) &= P_{dif}\{\mathcal{G} \leq g\} = \sum_{g_i \leq g} P_{dif}\{\mathcal{G} = g_i\} = \sum_{g_i \leq g} p_d \end{aligned} \quad (4.6)$$

De tal forma que, al reescribir la ecuación 4.3, pero en términos de las distribuciones de probabilidad consideradas para  $S_{sum}$  y  $S_{dif}$  se pueden expresar como:

$$H(I_{Range}) = H(S_{sum}) + H(S_{dif})$$

tal que,

$$H(I_{Range}) = - \underbrace{\sum_{p_0} P_{sum}(p_0) \cdot \log(P_{sum}(p_0))}_{H(S_{sum})} - \underbrace{\sum_{p_d} P_{dif}(p_d) \cdot \log(P_{dif}(p_d))}_{H(S_{dif})} \quad (4.7)$$

Esta última ecuación 4.7 representa una aproximación de la entropía considerando mutua independencia entre la suma y diferencia expresadas por medio de las ecuaciones 4.5 y 4.6, estas relaciones se establecen solamente si la *mpf* y las distribuciones correspondientes a  $S_{sum}$  y  $S_{dif}$  toman valores enteros.

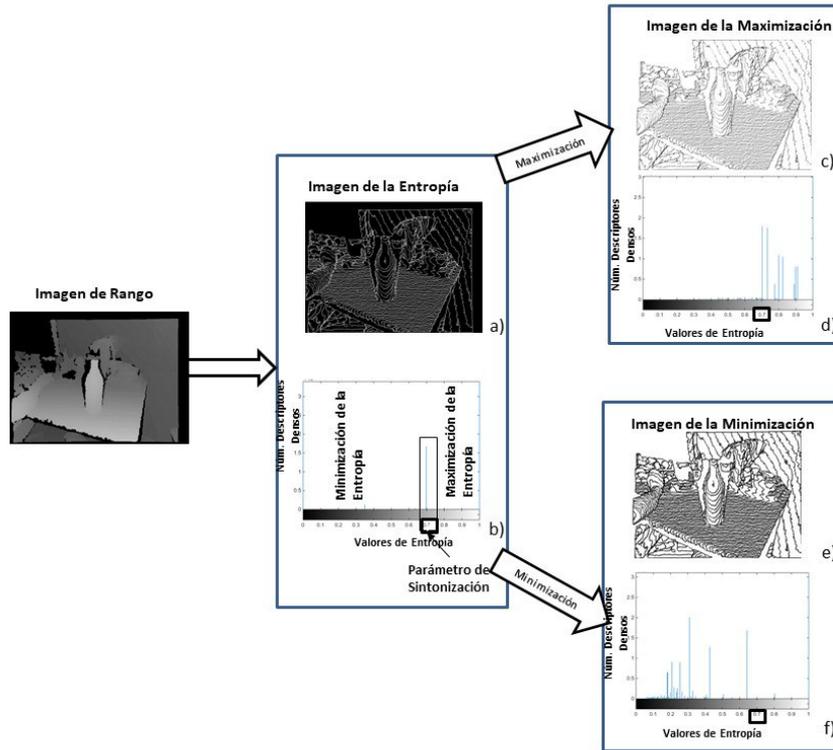


Figura 4.1: Cálculo de la entropía para una imagen de rango (salsita-DSPLab dataset). El valor de entropía calculado es de 0.7. Este se encuentra marcado en los respectivos histogramas por un recuadro negro. a) Representación de la entropía como imagen, así como, b) su respectivo histograma de frecuencias. c) Representación de la maximización de la entropía como imagen y d) su correspondiente histograma. e) Representación de la minimización de la entropía como imagen, así como, f) su respectivo histograma (más detalles en la sección 4.2.1).

#### 4.2.1. Maximización y Minimización de la Entropía

El valor de la entropía determinado por la ecuación 4.7, en la sección anterior puede ser considerado como un *parámetro de sintonización*  $t_{tun}$ .

$$t_{tun} = -H_{sum} - H_{dif}$$

Para dar una interpretación del cálculo de la entropía como característica

de textura de la  $I_{Range}$  se considera que, cada valor obtenido a través de  $H_{sum}$  y  $H_{dif}$  nos permite determinar que para cambios pequeños de probabilidad, el valor de entropía será bajo, por lo que, se tomará como una textura suave, o bien, una escena con pocos objetos o cambios en el ambiente. Por el contrario, una escena con una aglomeración de objetos o un ambiente poco controlado, tenderá a un alto índice de entropía. Más aún, si todos los elementos son equiprobables entonces, la entropía será máxima.

Para ejemplificarlo, presentamos la imagen (figura 4.1a), la cual es generada con base en los resultados obtenidos de la matriz de aproximación de la entropía calculada con la ecuación 4.7. Adicionalmente, se muestra la distribución de probabilidad (ver imagen 4.1b). Con el fin de establecer el rango y la división en los que se deberá de encontrar la maximización, así como, la minimización de la entropía. De tal forma que, todos los valores calculados por encima de  $t_{tun}$  se consideran una *maximización* (ver, figura 4.1c-d). Mientras que, de manera inversa, todos los valores estimados por debajo de  $t_{tun}$  son considerados como una *minimización* de la entropía (ver, figura 4.1e-f).

El hecho de poder realizar una separación de estos dos procesos ( maximización y minimización de la entropía), nos permite establecer que en ambos lados de  $t_{tun}$ , la información de la imagen de rango, puede manejarse como sistemas en equilibrio independientes. Esto es, del lado izquierdo de  $t_{tun}$  los valores correspondientes a las  $p_i$ 's son más cercanos a cero. A fin de generar un sistema en equilibrio, los valores correspondientes a la entropía son los más bajos mientras que, los valores correspondientes a la energía son los más altos. Esto nos permitirá posteriormente (ver, subsección 4.2.2), establecer puntos de interés que se encuentren en regiones de intensidades similares; pero, al mismo tiempo, determinando una separación entre los diferentes objetos que conforman una escena.

De manera inversa, sobre el lado derecho de  $t_{tun}$  los valores de las  $p_i$ 's son cercanos a 1, así que, estos apuntan a un alto nivel en los valores de la entropía, con la cual se puede representar la diagonal principal correspondiente a la ecuación 2.7; dando como resultado una cercanía espacial de dichos valores ( $p_i$ 's) hacia la diagonal. En este caso, al identificar los puntos de interés que presentan una alta entropía, podemos caracterizar la imagen a través de una simple representación por medio de dichos puntos, que conforman el contorno de los objetos contenidos en la escena (ver, subsección 4.2.3).

#### 4.2.2. Minimización de la Entropía y su Relación con la Energía

Law en [81] propone medir la energía de la textura de imágenes digitales usando filtros de tipo Laplacianos y Gausianos. Las imágenes son procesadas por lo tanto, con un filtro conocido como *energía de textura local no lineal*, para detectar los puntos que contengan un alto contenido de energía, y que representen los descriptores de los objetos contenidos en la escena. Más aún,

el autor presenta cómo, para un *campo de medida cero*, la **varianza** es una medida de **energía**. De tal forma que, al ser la desviación estándar la raíz cuadrada de esta medida, dicha desviación se torna a su vez, en una medida *local de energía* calculada para la textura. Así, para mantener el sistema en equilibrio, tendremos que para, valores altos de energía (texturas suaves), la entropía tenderá a valores mínimos.

Por lo tanto, la minimización de la entropía para este caso es considerada como la estimación de las distribuciones de probabilidad  $p_i$ , con los pequeños valores promedio de los cambios de energía ( $dE$ ). Asumimos, entonces, que los valores promedio de alguna  $E_i$  es asociado con la distribución de probabilidad, para la cual el valor esperado de la energía es descrito como

$$E = \sum_i^n p_i E_i$$

tal que, la distribución de la probabilidad se expresa solamente como variaciones de primer orden, esto es  $\sum_i^n dp_i = 0$ . Así, la ecuación 4.3 es reescrita en términos de  $H'(I_{Range})$  como sigue:

$$H'(I_{Range}) = \sum_i^n dp_i + \sum_i^n \log p_i dp_i \quad (4.8)$$

Adicionalmente, este proceso es llevado a cabo usando la matriz de Lagrange [121] en su forma más simple como multiplicador de Lagrange por tanto, la ecuación 4.8 se reescribe como,

$$H'(I_{Range}) = \sum_i^n (\gamma + \beta E_i) dp_i \quad (4.9)$$

El término  $\log p_i$  en la ecuación 4.8 es expresado en función del multiplicador de Lagrange, donde el valor de  $\beta$  es considerado como una variable independiente, tal que,  $\gamma$  dependa de  $\beta$ . En el proceso de minimización, se debe de asumir que  $0 \leq \beta \leq t_{tun}$ , tal que si  $\beta = 0$ , entonces todas las probabilidades en un número finito de estados ( $n$ ) son similares. En este caso, el proceso de minimización de entropía fue llevado a cabo con un valor fijo de  $t_{tun}$ , tal que, si  $\log P_i = t_{tun}$  y  $\beta = 0$  entonces  $\gamma = t_{tun}$ . Más aún, es posible asumir que  $dp_i$  involucra un pequeño cambio de energía, el cual es  $dp_i = t_{tun} - E_i$  donde,  $E_i$  corresponde al nivel de energía presente en el punto  $(k, l)$ . Por lo tanto, el parámetro de sintonización tiende a moverse hacia el lado izquierdo en el rango de  $(0, t_{tun}]$  a fin de mantener el sistema en equilibrio a través de la ecuación 4.9.

Por otro lado, los puntos de interés (ver sección 4.9) son calculados de acuerdo a los valores más bajos de la entropía, los cuales pertenecen al lado izquierdo de  $t_{tun}$  (ver histograma de la figura 4.1b). Estos resultados son asociados con el *principio de simplicidad* [70], el cual, asocia los patrones

subyacentes más simples. En este caso, valores similares de baja entropía pueden ser considerados como un patrón descrito a través de líneas de isoelevación en la imagen resultante, como se muestra en 4.1e-f.

### 4.2.3. Maximización de la Entropía y su Relación con la Homogeneidad

La maximización de la entropía es comparable con obtener una superficie con un cambio mínimo de energía [31]. Consecuentemente, este proceso es bastante similar a extraer la homogeneidad (como característica de textura) a partir de una imagen de rango. En efecto, los valores que pertenecen al proceso de maximización de la entropía son considerados valores disimilares presentes en la superficie homogénea. Así, la información obtenida para este proceso representa objetos que pueden ser separados uno de otro de acuerdo con un nivel máximo de entropía, presente en el contorno de los objetos contenidos en la escena.

De tal forma que, los valores de una máxima entropía se encuentran en el rango de  $[t_{tun}, 1]$  correspondiente a la ecuación 4.9 donde  $\beta \geq t_{tun}$ , tal como se muestra en la figura 4.1d. Esto significa que el sistema tenderá al equilibrio siempre y cuando el cambio de energía sea mínimo en las regiones que definen una de otra. Como resultado, se tiene que en la generalidad la superficie será homogénea (ver, la figura 4.1c donde se observa, que las líneas de isoelevación definen esta superficie). Pero, justo en las zonas de disimilaridad (observar los cambios de nivel de gris, en los “bordes” de cada objeto o área de la escena 4.1c) se presentan los diferentes cambios de profundidad o disparidad; por lo que, será cuando se presente una alta entropía.

Un punto interesante de la metodología propuesta es la aproximación por medio de las características de la textura como entropía, y su correspondiente a la maximización (homogeneidad) y la minimización (energía) de la entropía. Además, la técnica muestra cómo se calcula la entropía, utilizando una aproximación que describe la aleatoriedad en las imágenes de rango, considerando la información de los niveles de gris como textura. Los resultados obtenidos presentan puntos claves con los que es posible detectar el contorno del objeto en la escena. Estos puntos se obtienen con menos del 10% de la información total resultante del proceso denso. Además, esta técnica nos permite definir una región interesante (espacio de trabajo), en la que existe un objeto particular a detectar. La evaluación del rendimiento de esta propuesta se lleva a cabo utilizando una base de datos genérica (Middlebury) y nuestra propia base de datos.

## Capítulo 5

# Resultados

*“I’ve always believed that if you put in  
the work, the results will come.”*

Michael Jordan

### 5.1. Resultados de la Metodología I

El conjunto de datos de mamografía utilizado en este documento se recopiló en el Hospital General de Irapuato, Guanajuato. El conjunto de datos consta de 74 imágenes: 22 diagnosticadas con calcificaciones y 52 diagnosticadas como tejido normal. Esto se lleva a cabo tomando como referencia el sistema de clasificación para mamografías BI-RADS (Breast Imaging Reporting and Data Systems, en inglés). La base de datos que se consideró para probar la técnica son imágenes almacenadas en formato *DICOM 3.0* con un tamaño de  $4784 \times 3517$  píxeles. La figura 5.1a muestra una mamografía en el formato DICOM; mientras que, en 5.1b se muestra el atributo o característica CP de la mamografía.

Los resultados obtenidos de la interpolación global del histograma de  $h_{cp}$  se ilustran en la figura 5.2. Es de hacer notar, que esta función resalta masas, conductos, calcificaciones, así como tejido sano al mismo tiempo. Es por ello que, como se revisó en la subsección 2.1.2, se requiere aplicar nuevamente una interpolación en la zona de alta frecuencia del histograma CP, que produce la función  $g(cp)$ . La figura 5.2 ilustra la calcificación detectada por la función  $g(cp)$ . Sin embargo, este segundo análisis aún necesita el validar cuáles de estos puntos encontrados pertenecen al crecimiento o masa anormal del tejido, es decir, si estos puntos son calcificaciones o no. Para lograr esto, se calculan y analizan cinco atributos o características desde la función  $g(cp)$  y se usan como entrada para el clasificador *KNN*.

La figura 5.3a ilustra los resultados de clasificación obtenidos por un ex-

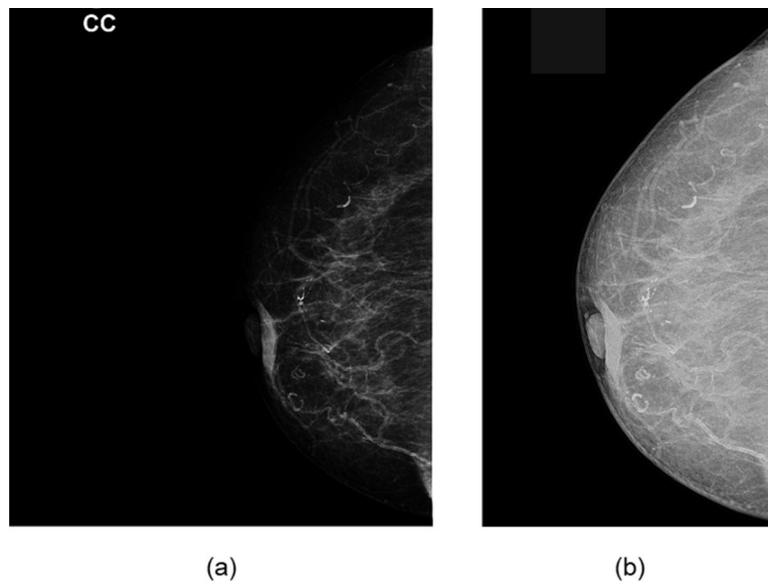


Figura 5.1: a) Mamografía en formato *DICOM* 3,0. b) Imagen resultante de la mamografía procesada para estimar CP.

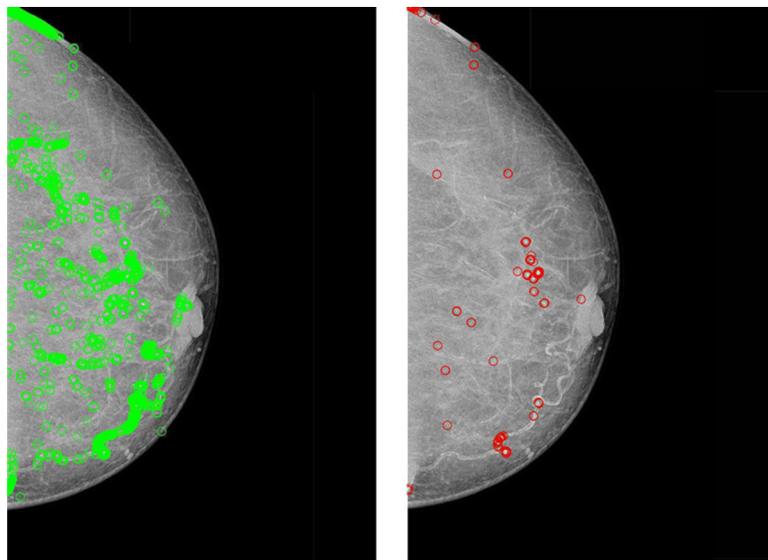


Figura 5.2: Ejemplo de los resultados de calcificaciones detectadas. a) Calcificaciones y tejido normal detectado por el experto. b) Calcificaciones detectadas con la interpolación local.

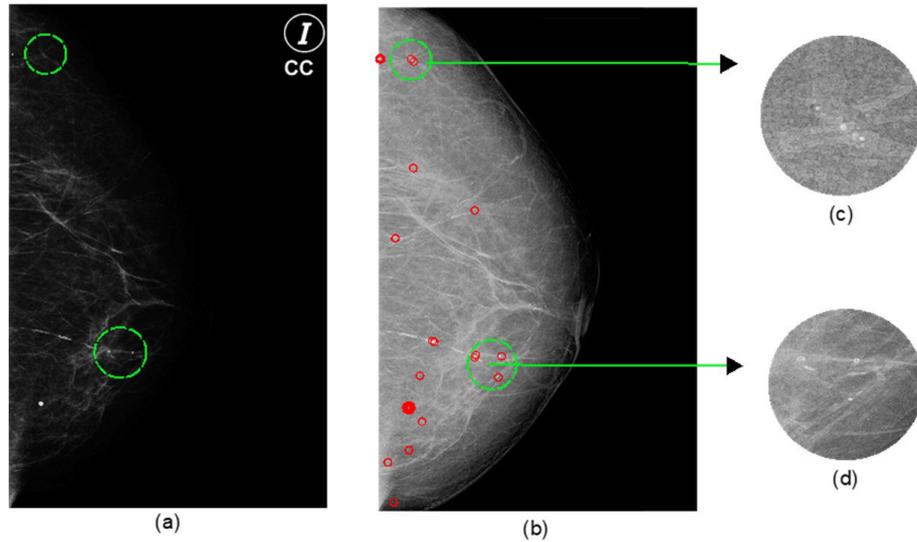


Figura 5.3: Ejemplo de los resultados de calcificaciones detectadas. a) Calcificaciones y tejido normal detectado. b) Calcificaciones detectadas con la interpolación local.

perto; mientras que, en la figura 5.3b se presentan los resultados utilizando nuestro método. Como se puede apreciar cualitativamente, la técnica propuesta detecta las dos calcificaciones en la zona marcada por el experto en 5.3a. La función CP detecta cambios de intensidad en la mamografía; por lo que, es posible detectar calcificaciones en 5.3b, las cuales son zonas marcadas por el experto. La 5.3c y 5.3d muestran los grupos de microcalcificaciones encontradas. Por otro lado, un segundo resultado se ilustra en la figura 5.4. Más aún, existen algunos agrupamientos de píxeles que por su comportamiento matemático pueden ser consideradas como microcalcificaciones.

#### **Resultados cuantitativos.**

Los resultados cuantitativos obtenidos de las pruebas experimentales se resumen en la matriz de confusión que se muestra en la tabla 5.1. El peor de los casos de un clasificador es la puntuación de Falsos Negativos ( $FN$ ), que se refiere a los casos en que el clasificador no detecta la calcificación y esta existe. De la tabla 5.1, se obtienen dos falsos negativos ( $FN$ ) de las pruebas experimentales. Por otro lado, se obtuvieron cuatro Falsos Positivos ( $FP$ ) de las pruebas experimentales, este resultado indica la presencia de calcificaciones que no existen, de acuerdo con el diagnóstico del médico experto.

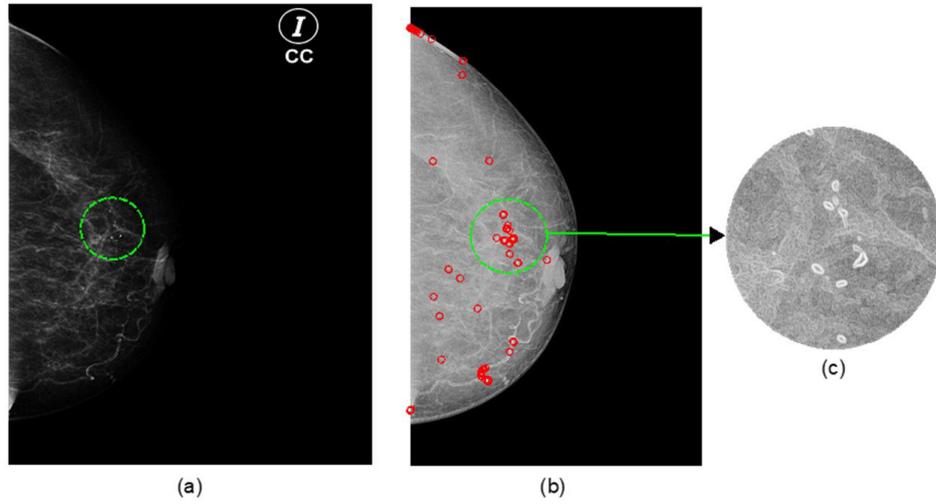


Figura 5.4: Resultados de la clasificación. (a) Mamografía en formato DICOM, los círculos resaltan las zonas con calcificaciones encontradas por un experto, (b) atributo de CP con las calcificaciones encontradas por el método propuesto, (c) acercamiento a las microcalcificaciones detectadas.

Tabla 5.1: Matriz de confusión correspondiente a la función  $g(cp)$ , para  $K = 3$  con una distancia Euclidiana.

Clases	Calcificación	Tejido normal	Especificidad
<b>Calcificación</b>	20( $TP$ )	2( $FN$ )	0.9090
<b>Tejido normal</b>	4( $FP$ )	48( $TN$ )	0.9230

Como se informó en la matriz de confusión, la precisión obtenida de los resultados experimentales es 0.9090 para la detección de calcificaciones y 0.9230 de especificidad para la detección de tejido normal. Los resultados obtenidos a nivel de exactitud serán comparados con las propuestas encontradas en trabajos previos a fin de obtener un mejor análisis. Tal que, en el peor de los casos durante la etapa de clasificación se debe a la presencia de valores conocidos como “falsos negativos” ( $FN$ ). Al analizarse las posibles causas de este resultado se concluyó que, en ambos casos, las imágenes de mamografía originales muestran bajo brillo.

Tabla 5.2: Resultados del rendimiento presentado por las métricas con respecto del método propuesto.

Pol. Grado	Exactitud	Precisión	Especificidad	Sensitividad	%FA
9	0.8648	0.8333	0.9423	0.6818	0.0576
10	0.9054	0.8571	0.9423	0.8181	0.0576
11	0.8513	0.8333	0.9411	0.6818	0.0588
12	0.8378	0.9166	0.9807	0.5000	0.0192
13	0.9189	0.8333	0.9230	0.9090	0.0769
14	0.9189	0.8333	0.9230	0.9090	0.0769
15	0.9054	0.8226	0.9230	0.8636	0.0769

La tabla 5.2 muestra la evaluación llevada a cabo, con respecto del rendimiento de nuestro enfoque. Tales como exactitud, precisión, especificidad, sensibilidad y porcentaje de “falsas alarmas” (FA). Si se observa, el valor de exactitud es 0.9189, por lo tanto, el porcentaje de reconocimiento es alto, al ser este cercano a 1, esto es, el error de detección es menor al 10 %. Mientras que, el valor de precisión calculado es 0.8333, que no es tan alto como cabría de esperar, debido a los falsos positivos. Esto significa que el error más común de nuestra estrategia ocurre cuando se detecta una calcificación, pero no existe. Dichos resultados también son validados por la especificidad y los valores de sensibilidad: 0.9230 y 0.9090, respectivamente. El porcentaje de detecciones positivas y negativas incorrectas (falsas alarmas, FA) es 0.0384, un bajo porcentaje de falsos negativos es el resultado más esperado obtenido.

Tabla 5.3: Tabla comparativa de nuestra método con respecto de otros trabajos previos.

Métodos	Base de Datos	Modo de procesamiento	Exactitud
<b>Estrategia propuesta</b>	Hospital Irapuato	Automatizado	<b>0.9189</b>
AlexNet+Sparse MIL [169]	INbreast	Automatizado	0.90 ± 0.02
AlexNet+Label MIL [169]	INbreast	Automatizado	0.86 ± 0.02
RFF on CNN w/pre-training [34]	INbreast	Automatizado	0.91 ± 0.02
RFF w/pre-training [34]	INbreast	Automatizado	0.84 ± 0.02
Shape and texture+LDA [36]	INbreast	Manual	0.8900
Level set [132]	Univ. of Michigan	Semi-Automatizado	0.84 ± 0.02
Active contourshi2008characterization	Univ. of Michigan	Semi-Automatizado	0.84 ± 0.02
Texture feature + SLDA [9]	INbreast	Semi-Automatizado	0.8700
Contour + ANN + LDA [152]	INbreast	Semi-Automatizado	0.8700

El análisis comparativo entre los diferentes métodos encontrados en trabajos previos y nuestra técnica se muestra en la tabla 5.3. La primera columna presenta las técnicas utilizadas por diferentes autores, incluida la estrategia propuesta. La segunda columna muestra la base de datos utilizada para los resultados experimentales. La tercera columna indica cómo se realiza el procesamiento de la imagen, el cual puede ser: manual, semiautomático o automatizado. La última columna muestra la precisión alcanzada por cada método, en el reconocimiento y clasificación. Es de hacer notar que, nuestro método propuesto tiene resultados similares a los de AlexNet + Sparse MIL[169] y Random Forest en CNN con entrenamiento previo [34], con un valor de precisión superior a 0.9.

Además, en estos métodos el procesamiento de algoritmos se automatiza permitiendo el uso de tales estrategias para aplicaciones reales. Sin embargo, queremos señalar que en el caso de los métodos que calculan características de textura, como [36], [9] y nuestro enfoque, nuestra estrategia muestra el mejor rendimiento. Además, entre estos tres métodos, nuestro enfoque es el único método que procesa imágenes de forma automatizada.

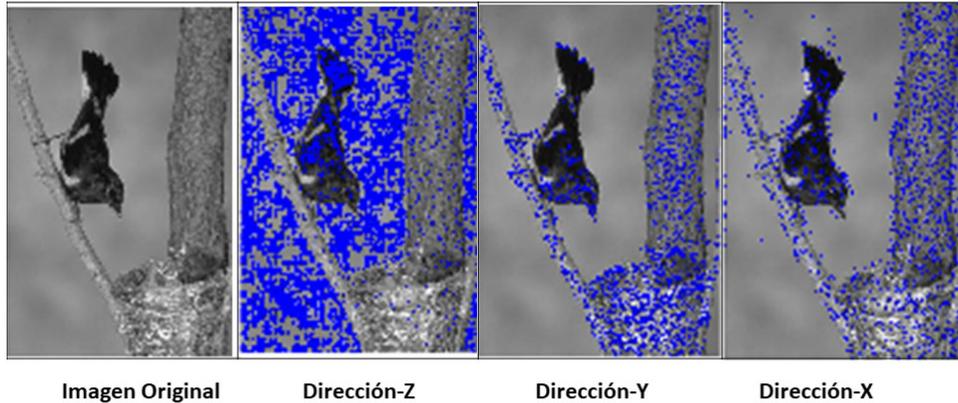


Figura 5.5: Ejemplo ilustrativo de los puntos de interés obtenidos usando la técnica SKD calculando la homogeneidad como característica de textura, aplicados a la imagen bird de base genérica conocida.

## 5.2. Resultados de la Metodología II

En esta sección, la exposición de los resultados generados de aplicar la técnica SKD se muestra usando imágenes en nivel de gris de una base de datos conocida, nuestra base de datos DSPLab [57]; así como, los resultados alcanzados, usando las imágenes de disparidad de la base genérica Middlebury database. Estas últimas se obtuvieron de un sistema de estereovisión, bajo condiciones ambientales controladas (un pormenor al respecto del sistema de adquisición, favor de dirigirse a las referencias [58] y [112]).

La base de datos DSPLab contiene imágenes VGA obtenidas directamente del sensor Kinect. A través de su sensor de profundidad [103], es como se adquieren las imágenes, tal como se revisó en la sección 1.5.5. La adquisición de estas imágenes fue bajo un ambiente de interior no controlado, donde imperaba una gran cantidad de luz exterior que incidía a través de los diversos ventanales; la propia iluminación del laboratorio, así como la disposición de los objetos aglomerados o no, en el escenario.

### 5.2.1. Resultados Previos de Imágenes en Niveles de Gris

En esta sección se exponen algunos de los resultados obtenidos al aplicar nuestra técnica SKD, desarrollada en la sección 3.3.3. Primero, se ofrece un análisis descriptivo haciendo uso de imágenes en niveles de gris, tal como se muestra en la figura 5.5.

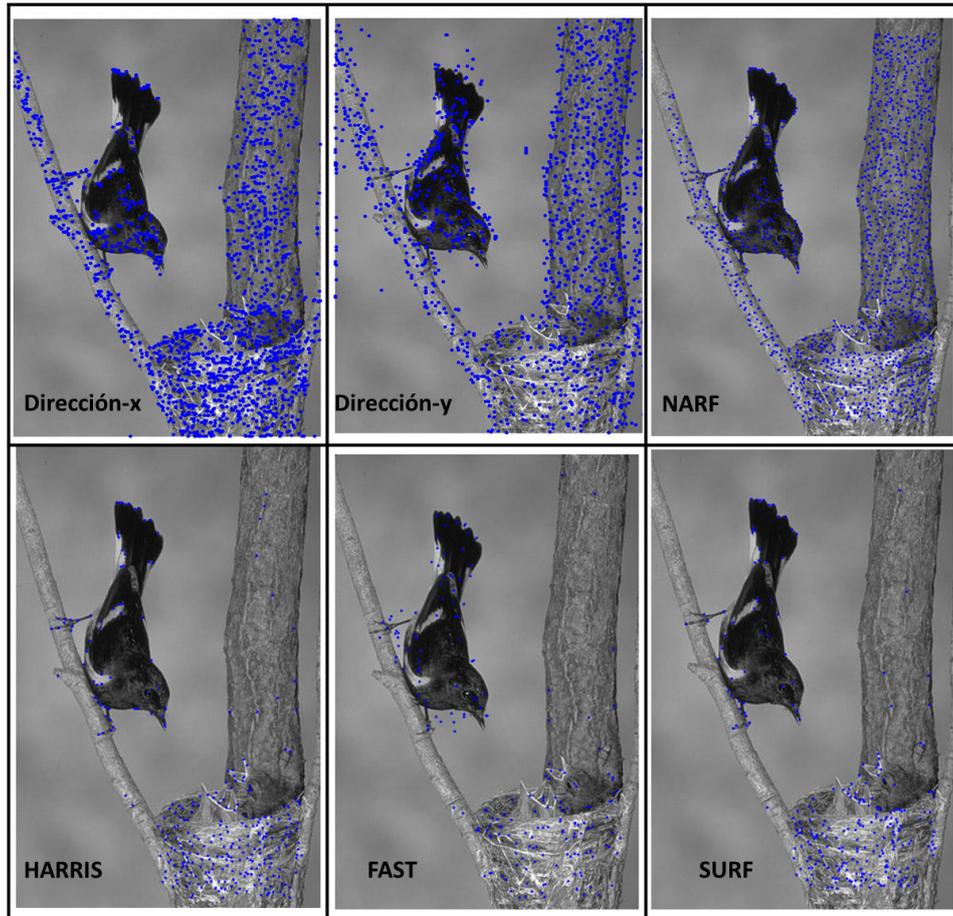


Figura 5.6: Ilustración comparativa de los puntos de interés generados con técnicas de trabajos previos como son NARF; Harris; FAST y SURF. Las primeras dos imágenes corresponden a los puntos de interés obtenidos en las direcciones  $x$  y  $y$  con la técnica propuesta.

El objetivo de computar los puntos de interés en cada superficie direccional tiene la finalidad de separar los objetos contenidos en la escena del fondo de esta; tal como se muestra en la ilustración 5.5 para las direcciones  $x$ ,  $y$  y  $z$ , respectivamente. En el caso de la dirección  $z$  (o  $M_z$ ), parecería que dichos puntos se encuentran concentrados en la parte correspondiente al

“fondo” de la imagen. La razón de esto, es debido a que, prácticamente son indistinguibles las diferencias existentes entre los valores de los diversos descriptores; esto es, la información se encuentra homogenizada en la superficie correspondiente a  $M_z$ .

Un ejemplo comparativo visual de la presente propuesta con respecto de otras técnicas conocidas para la detección y extracción de puntos en imágenes en niveles de gris se ejemplifica en la figura 5.6.

Es importante hacer notar que los resultados que se lograron aplicando la técnica SKD, con respecto de otras técnicas presentadas en la subsección 1.5.7 como son SURF, FAST, NARF, S & T y Harris visual y cuantitativamente presentan un buen desempeño en relación con el número de puntos de interés detectados. La validación de esta afirmación se presenta en la siguiente subsección.

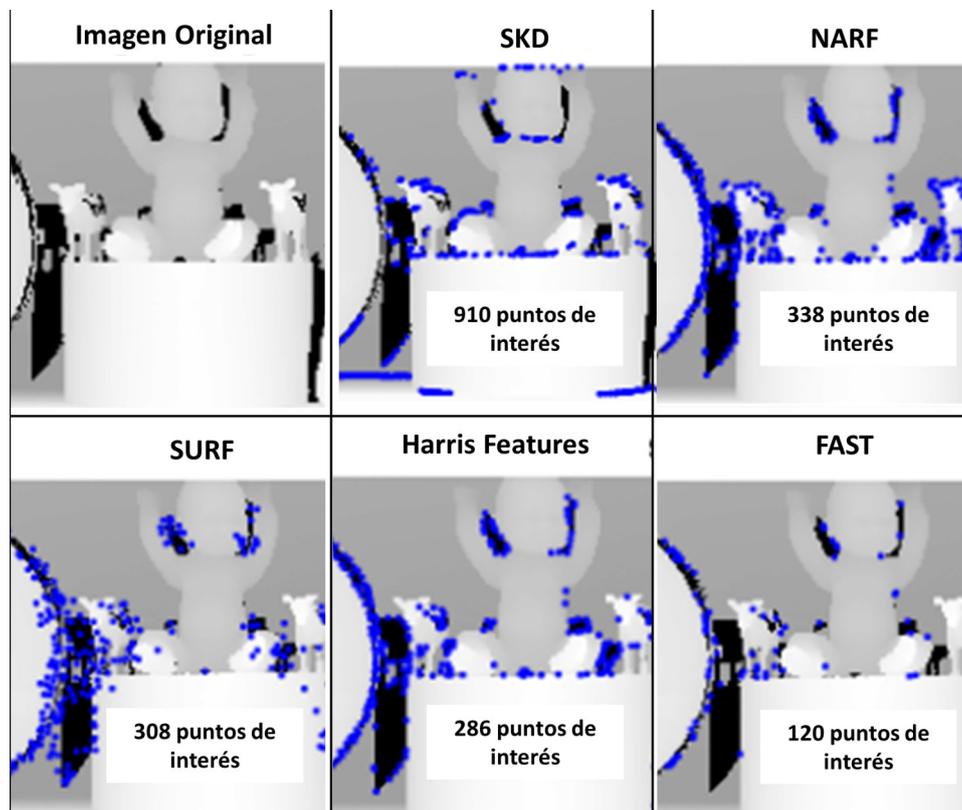


Figura 5.7: Ilustración de los puntos de interés generados con la técnica SKD en la dirección  $y$ ; así como otras técnicas ya conocidas y revisadas en la subsección 1.5.7.

### 5.2.2. Resultados Previos de Imágenes de Rango

La técnica propuesta SKD nos proporciona una fácil implementación en relación con otras técnicas ya probadas (ver, subsecciones 1.5.7 y 3.3.3) para la detección y extracción de puntos en imágenes en niveles de gris. Aunado a ello, cuando dicha técnica es aplicada a imágenes de rango como en este caso son las de disparidad o profundidad el desempeño de nuestro detector es similar al mostrado en trabajos que son referentes en esta área.

En relación con las imágenes de disparidad, comenzaremos mostrando en la figura 5.7 los resultados obtenidos con respecto del número de puntos detectados por nuestra técnica en una comparativa con respecto de las técnicas antes mencionadas. El número de puntos de interés estimados, por cada una de las técnicas empleadas, es presentado al pie de cada imagen correspondiente. La técnica SKD estima al menos un 60% más, de puntos de interés con respecto de los calculados por las otras técnicas probadas. En general, estos resultados, se presentan en la mayoría de las pruebas realizadas para imágenes de rango.

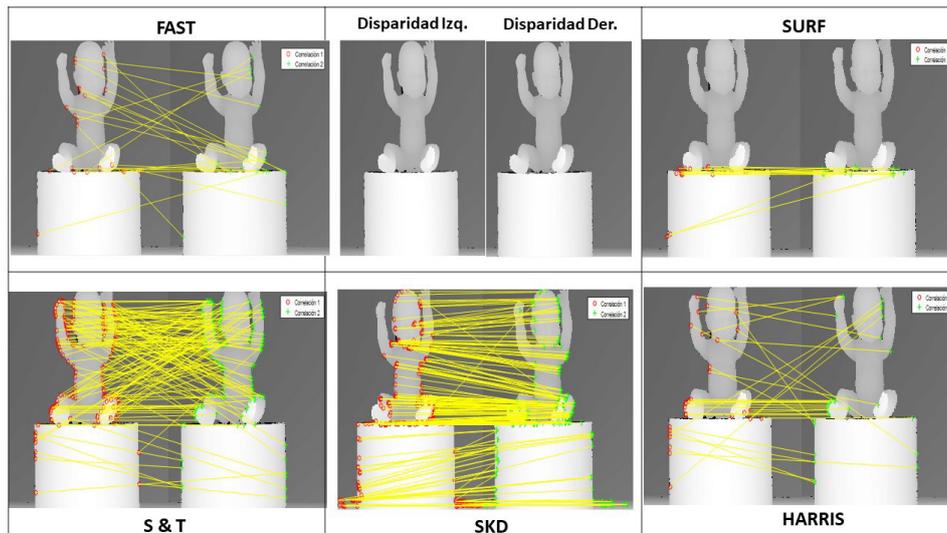


Figura 5.8: Ilustración que muestra el seguimiento de los puntos de interés (rojos imagen disparidad izquierda y verdes para la derecha) computados de las imágenes de disparidad de baby1 [58]. Estos son evaluados de acuerdo con lo presentado en 3.3.5.

Por otro lado, a pesar de que nuestra propuesta presenta un buen número de puntos detectados, esto no determina la eficiencia de nuestro detector; por lo tanto, para validar el desempeño de este con respecto de las técnicas ya conocidas, se decidió hacer uso del proceso de validación presentado en la subsección 3.3.5.

Aun cuando este método tiene una eficacia de correlación que oscilala en

un rango [70, 85] % de acuerdo con lo reportado en [91]; para su implementación es necesario hacer uso de al menos tres imágenes. La adecuación que se aplica con dicho método nos permitió trabajar solamente con dos imágenes de disparidad utilizando la base de datos de Middlebury [58] y [112]; la cual es una base genérica como se ha mencionado anteriormente. Un ejemplo descriptivo visualmente de los resultados obtenidos los tenemos en la figura 5.8. En esta imagen se presentan las correlaciones consideradas como verdaderos positivos y falsos positivos. La discriminación de estos últimos todavía es un tema pendiente para desarrollar en trabajos futuros. Pero a pesar de ello, en promedio nuestra técnica presenta una precisión en relación con la correlación que está por encima del 80 % de correlaciones verdaderas positivas.

En la Tabla 5.4 se presenta en promedio los datos de entrada con los que se alimenta este proceso. Las imágenes computadas en este caso son de tamaño  $370 \times 443$  [58].

Tabla 5.4: Tabla comparativa de nuestra método con respecto de otros trabajos previos.

	<b>HARRIS</b>	<b>SURF</b>	<b>FAST</b>	<b>S &amp; T</b>	<b>SKD</b>
<b>Carac. Imagen Izq.</b>	$85 \times 1$	$16 \times 64$	$24 \times 1$	$205 \times 1$	$402 \times 1$
<b>Carac. Imagen Der.</b>	$88 \times 1$	$23 \times 64$	$38 \times 1$	$213 \times 1$	$435 \times 1$
<b>Puntos Detec. Izq.</b>	74	16	24	181	375
<b>Puntos Detec. Der.</b>	80	21	38	191	402

La Tabla 5.5 presenta un concentrado en promedio de la aplicación del proceso de validación de los puntos detectados como puntos de interés. El número de correlaciones que se muestra corresponde a las reales, esto es, a correlaciones que son verdadero positivo y falso positivo. Nuestra propuesta puede ser equiparable al desempeño que se tiene con respecto de otras técnicas ya ampliamente conocidas y probadas.

Tabla 5.5: Tabla comparativa del número de correlaciones reales en promedio y la precisión obtenida aplicando el proceso de validación referido en 3.3.5 a la base de datos Middlebury.

	<b>HARRIS</b>	<b>SURF</b>	<b>FAST</b>	<b>S &amp; T</b>	<b>SKD</b>
<b>Núm. de Correlaciones</b>	30	14	22	134	228
<b>Precisión</b>	0.8333	<b>0.8571</b>	0.4554	0.7089	<i>0.8245</i>

Por otro lado, la generación de imágenes sintéticas partiendo de la in-

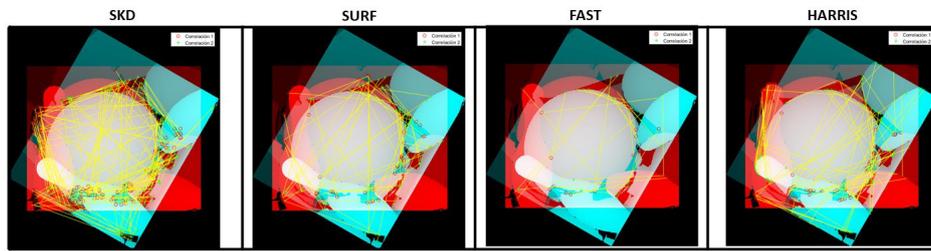


Figura 5.9: Ilustración de los puntos de interés “calculados” con la técnica SKD, así como otras técnicas de trabajos previos. La imagen original bowling2 [58], es rotada  $60^\circ$  para producir la imagen sintética.

formación contenida en las imágenes de disparidad de la base de datos de Middlebury ([112] y [58]); nos ayuda a ejemplificar la obtención de puntos de interés en la escena. En este sentido, debe existir una “correspondencia” (seguimiento) entre los puntos de interés obtenidos partiendo de la imagen original (puntos en color rojo), así como los calculados en la imagen sintética (puntos en color verde), tal como se muestra en la figura 5.9.

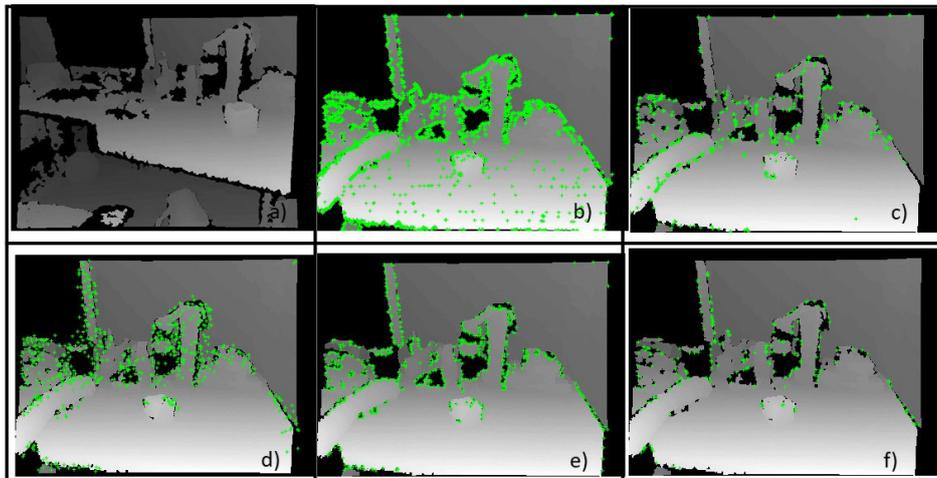


Figura 5.10: Ilustración comparativa de los puntos de interés que se obtienen aplicando las técnicas b) SKD (dirección en  $x$ ), c) Harris; d) NARF; e) SURF; f) FAST; a) imagen de profundidad correspondiente a nuestra base de datos.

Por otro lado, haciendo uso de nuestra base de datos generada del sensor de profundidad Kinect tenemos que, los resultados estimados son mostrados en la figura 5.10. En esta se muestra que, el descriptor SKD (ver figura 5.11-c) presenta una mejor definición del contorno de la escena (más detalle, referirse a sección 3.3.4); que el presentado por las otras técnicas presentadas

en la sección 1.5.

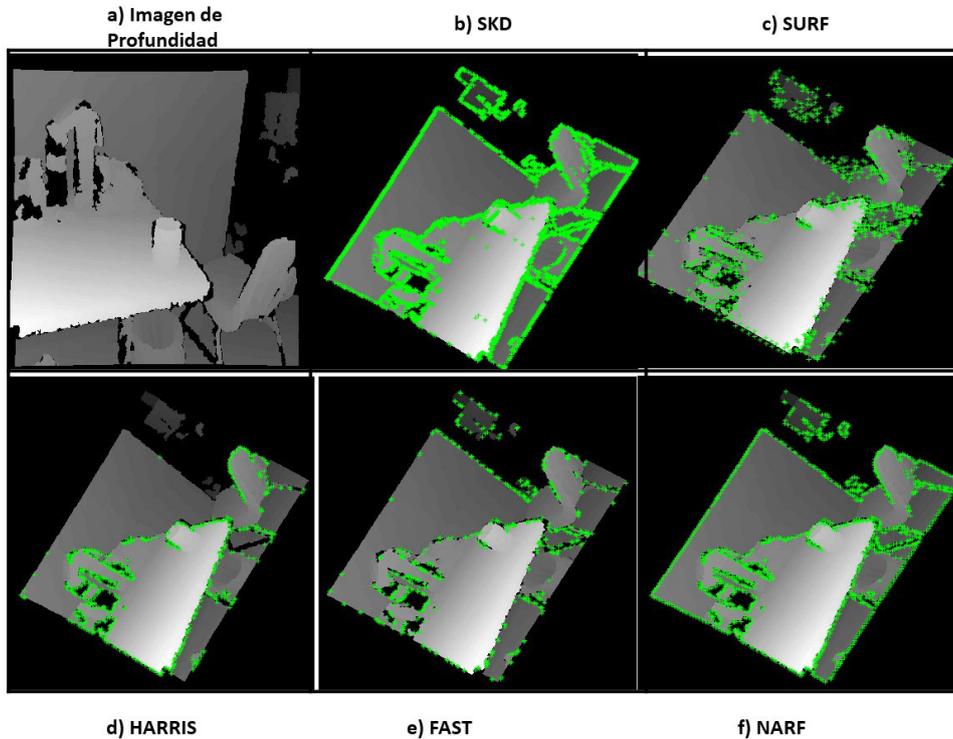


Figura 5.11: La figura presenta los resultados correspondientes a la rotación de la información contenida en la imagen de rango ( $60^\circ$ ). Las pruebas realizadas con respecto de otras bases de datos son: b) nuestra técnica; c) SURF; d) HARRIS; e) FAST; f) NARF. a) Es la imagen de profundidad, solo como referencia.

En la generalidad, en todas las pruebas realizadas para imágenes de rango, nuestra técnica supera por más del 60 % el número de puntos extraídos, en relación con las otras técnicas presentadas en esta sección, con una correlación en promedio mayor al 50 % con respecto de los puntos detectados y con una precisión similar a las de las técnicas ya reconocidas en esta área.

### 5.2.3. Análisis Cuantitativo

La figura 5.12 en la primera columna, presenta las imágenes VGA sólo como referencia. Mientras que, en la segunda columna se muestran las imágenes de profundidad, con las cuales se lleva a cabo el procedimiento previamente descrito. Los resultados obtenidos de aplicar la técnica SKD, son los presentados en la figura 5.12 de la tercera a la quinta columna. Esto

es, en la tercera columna, se muestran los puntos de interés, los cuales en su mayoría se encuentran sobre el contorno de los objetos contenidos en la escena. Dicho contorno es representado por medio de los puntos de interés  $S_K$ , calculados en el paso 3. En este punto, es importante destacar que, estos puntos de interés se obtuvieron con menos del 15% en promedio del total de los descriptores  $N_D$ .

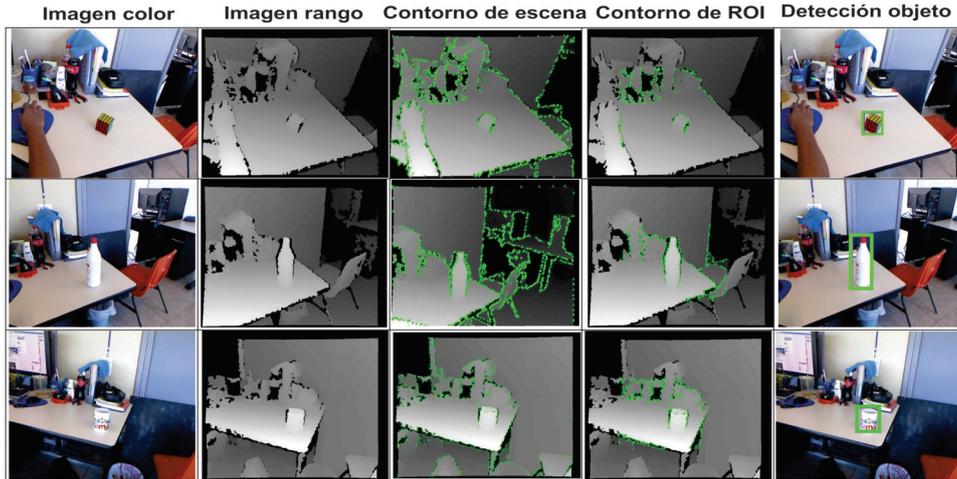


Figura 5.12: Resultados obtenidos de aplicar la técnica SKD, a las imágenes de profundidad de la base de datos DSSLab. Primer renglón cubo de Rubik, segundo renglón salsa, tercer renglón taza.

Los resultados de llevar a cabo el proceso marcado en el paso 4.a, se observan en la cuarta columna de la figura 5.12. Aquí se presenta el contorno de los objetos, contenidos de manera parcial o total en dicha región. Más aún, el objeto de interés a detectar puede ser fácilmente observable, ya que se encuentra bien definido, a través de sus puntos  $S_K$ . Por lo que, en los subsecuentes pasos 4.b a d, se procesan los descriptores correspondientes a los  $S_K$  de la ROI, para aislar el objeto. Por lo tanto, la detección del objeto se lleva a cabo, con menos del 7% de la información contenida en  $N_S$ . Los resultados correspondientes a este proceso se presentan en la quinta columna, de la figura 5.12 en la cual, finalmente, se muestra el objeto de interés detectado.

Por otro lado, la figura 5.13, presenta los resultados de la comparación cualitativa, llevada a cabo con la base de datos genérica Middlebury, para obtener puntos de interés. La primera columna de la figura 5.13, presenta las imágenes RGB sólo como referencia. Algunos resultados representativos de esta comparación se muestran de la segunda a la sexta columna. En particular, la segunda columna muestra los resultados obtenidos aplicando nuestra técnica, a las imágenes de disparidad de la base de datos Middlebury. Más

aún, en el primer renglón de la figura 5.13, se presentan los resultados, para un escenario que tiene un alto contenido de oclusión y saturación de objetos. Mientras que, en el segundo renglón los objetos contenidos en la escena se encuentran en primer plano esto es, la discrepancia de disparidades es mínima.

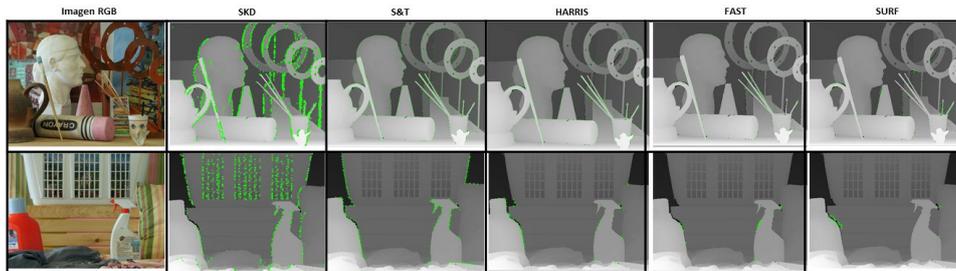


Figura 5.13: Comparación cualitativa con respecto a la base de datos Middlebury ([112] y [58]), de los resultados obtenidos para la detección de puntos de interés. Nuestra técnica SKD (segunda columna), tercera columna *S&T*, cuarta Harris, quinta FAST y sexta SURF. Primer renglón Art, segundo renglón Laundry.

La diferencia existente entre la cantidad de puntos de interés obtenidos usando nuestra técnica con respecto de las otras técnicas previamente revisadas, consideramos que se debe al número de descriptores calculados; aún cuando, esto no implica un parámetro de calidad, si es importante hacer notar, que la cantidad de puntos de interés va en decremento de la segunda a la sexta columna. Más aún, con nuestra técnica es posible distinguir el contorno de los objetos contenidos en la escena, independientemente, si estos están en un primer plano o en varios planos.

Adicionalmente, también podemos marcar contornos que se puedan presentar al interior de algún objeto (ver, segunda columna, segundo renglón de la figura 5.13). Mientras que, técnicas como la de *S&T* (tercera columna) o la de Harris Edge detector (cuarta columna) definen el contorno de los objetos, cuando estos están en diferentes planos. Estas mismas técnicas tienen problemas para detectar bordes o contornos de objetos, que se encuentran en un mismo plano. Por otro lado, las técnicas SURF y FAST presentan cualitativamente, una detección insuficiente de puntos de interés en ambos casos (imágenes con y sin oclusión). Finalmente, en la siguiente sección se presenta un análisis comparativo del costo computacional haciendo uso de la técnica SKD y otras propuestas previamente revisadas.

Para evaluar el desempeño de la técnica SKD con respecto de otras técnicas de trabajos previos en relación a la detección de puntos clave, primero se muestra en la figura 5.14, un ejemplo de las imágenes sintéticas generadas con la base de datos Middlebury, para llevar a cabo la evaluación. Estas

imágenes fueron generadas con dos tipos de ruido y sus combinaciones. La imagen difusa ( $D$ ) se obtiene a través de un filtro Gaussiano con una máscara de  $3 \times 3$  píxeles. Mientras que, las imágenes sintéticas sal y pimienta ( $S\&P$ ) se generaron con un 5% de nivel de ruido en promedio. Los otros dos tipos son una combinación de  $S\&P + D$  y  $D + S\&P$ . La evaluación que se presenta en la figura 5.15, muestra el desempeño del índice de repetibilidad de nuestra técnica, con respecto de algunas de las técnicas presentadas en la sección anterior.

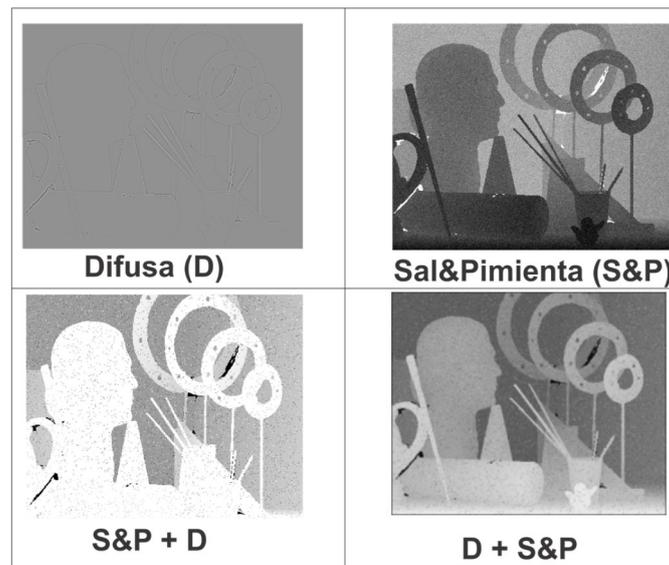


Figura 5.14: Ejemplo de imágenes sintéticas generadas con ruido Gaussiano ( $D$ ) y sal y pimienta ( $S\&P$ ) y algunas combinaciones entre estas.

El índice de repetibilidad se establece, como la cantidad de puntos de interés encontrados en las diferentes imágenes sintéticas, bajo un mismo proceso con respecto del total de puntos de interés contenidos en la imagen original (ver, por ejemplo, la figura 5.13). De tal forma que, cada punto de interés calculado (en la imagen sintética) por la técnica propuesta y las otras técnicas del trabajos previos, son asociados a cada uno de los puntos de interés contenidos en la imagen original como se presentó en la sección 5.2 ejemplificado en 5.9. Finalmente, estos puntos son validados por medio de la técnica SSD (Sum of Squares Difference technique, en inglés), para evitar los falsos positivos.

En 5.15, presenta un buen desempeño ante ruido  $S\&P$  en general para todas las técnicas aplicada, mientras que, se observa un deficiente desempeño para imágenes sintéticas con ruido tipo  $D$  y sus diferentes combinaciones. Este se presenta con un índice de repetibilidad por debajo de 0,5. Donde, 1

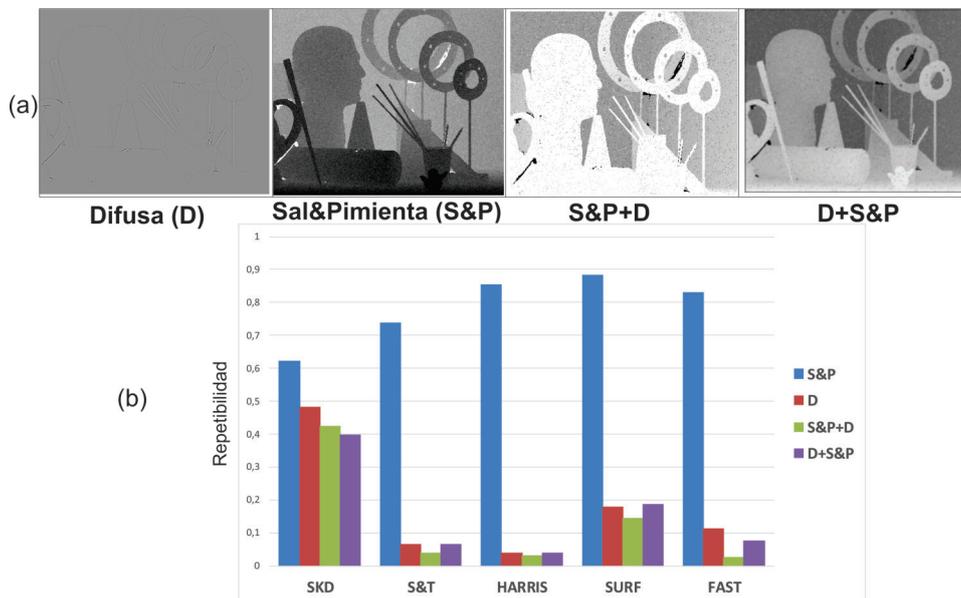


Figura 5.15: a) Ejemplo de la generación de imágenes sintéticas usando la base de datos Middlebury b) Comparativa del índice de repetibilidad usando la técnica SKD con respecto de las técnicas *S&T*, Harris, SURF y FAST.

corresponde a un 100 % de la correspondencia entre los datos. Por otro lado, cabe destacar que, se eligió trabajar con muestras no uniformes, debido a que, para las técnicas SURF y FAST, el número de puntos de interés obtenidos era menor al 1 %, con respecto de las otras técnicas. Este hecho, lo podemos observar cualitativamente desde 5.13.

Las imágenes mostradas en 5.16 son un ejemplo del comportamiento de las diferentes técnicas en relación con el número de puntos de interés detectados, así como la correspondencia que se lleva a cabo, entre estos puntos y los calculados usando imágenes sintéticas. Esto es, la información contenida en la imagen de profundidad o disparidad es rotada a  $\theta = a)30^\circ, b)60^\circ, c)150^\circ, d)180^\circ$ . Algunos ejemplos los podemos revisar en la sección 5.2. En la mayoría de los casos, para cada técnica la información se encuentra más concentrada hacia el tercer cuartil. Lo que nos indica una mayor dispersión de los puntos obtenidos al llevar a cabo el proceso de correspondencia. Por otro lado, los puntos que se ubican fuera de las cajas (ver figura 5.16a-d), son considerados valores atípicos.

De forma similar, se realizó la evaluación de desempeño de las diferentes técnicas, cuando se presentan cambios o transformaciones en la escala  $\alpha$ . Tal como se muestra en la figura 5.16-e, con factores de escalamiento iguales a  $\alpha = 0,5, 0,8, y 1,25$ . Cualitativamente, en dicha figura, se muestra el desempeño que se tiene de aplicar la técnica SKD a través de las superficies  $M_x$

y  $M_y$  como se presentó en el algoritmo 1, de la sección 3.3.3. Cabe hacer notar, que, el número de descriptores calculados en la técnica propuesta, en la generalidad siempre es mayor al número de descriptores presentados por las demás técnicas en orden un 60 % (ver sección 5.2).

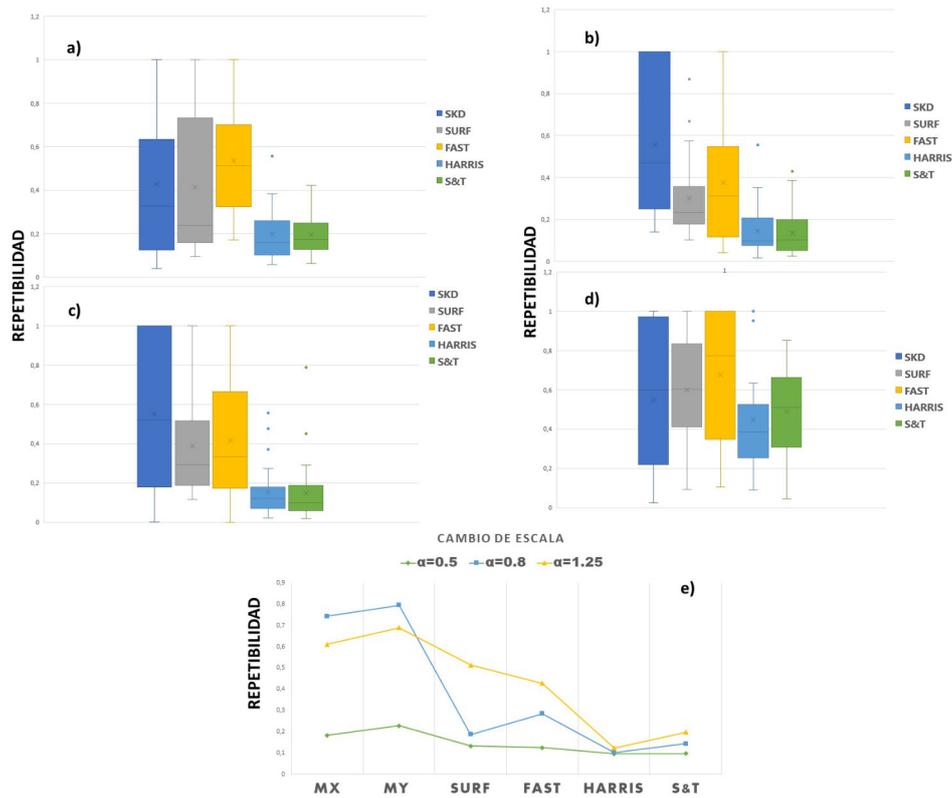


Figura 5.16: Resultados obtenidos de imágenes sintéticas con  $\alpha = 1$  y a)  $\theta = 30^\circ$ ; b)  $\theta = 60^\circ$ ; c)  $\theta = 150^\circ$  d)  $\theta = 180^\circ$ . e)  $\theta = 0^\circ$  y diversas escalas  $\alpha$ .

Finalmente, presentamos un comparativo del tiempo de procesamiento que presentan las diferentes técnicas, con respecto a la detección de puntos de interés. Estos resultados se muestran en la figura 5.17. De la gráfica, es posible observar que la técnica SURF, es la que presenta el mejor desempeño para cualquier base de datos. En seguida tenemos la técnica FAST y posteriormente, nuestra técnica SKD. Más aún, si observamos dicho desempeño con respecto a técnicas que llevan a cabo la detección de bordes o contornos (S&T y Harris Edge detector), la técnica SKD presenta un menor tiempo de procesamiento. Todos nuestros procesos se llevaron a cabo, a través de código Matlab, con un procesador Intel®Core2™Duo 2,66GHz CPU con 8GB de RAM.

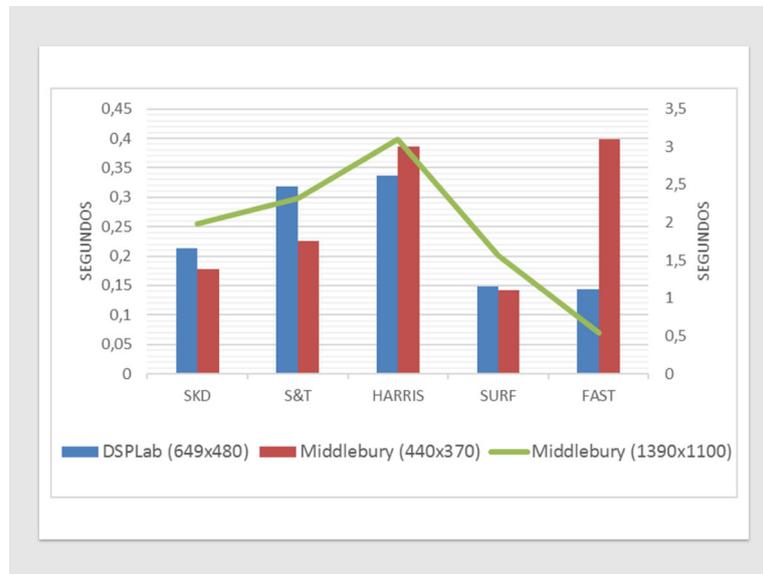


Figura 5.17: Gráfica comparativa del tiempo de procesamiento. El eje vertical izquierdo corresponde al tiempo en s, entre bases de datos de dimensiones similares. El de la derecha, corresponde a una base de datos de una dimensión mayor. La dimensión de cada base se expresa en  $K \times L$  píxeles.

### **Resultados: Detección de Objetos**

En esta sección presentamos algunos ejemplos del desempeño de la técnica aquí propuesta. Por ejemplo, en la figura 5.18 se muestra una comparativa de los descriptores calculados a través de la metodología propuesta usando la base de datos Middlebury con imágenes de tamaño  $440 \times 370$ .

La figura 5.18 muestra gráficamente, un comparativo de la relación en cantidad, de los descriptores en las pruebas realizadas, donde se estima en promedio un índice menor al 15%, entre los descriptores correspondientes a los puntos  $S_K$ , con respecto de los contenidos en  $N_S$ . Cabe aclarar que, el total de descriptores de  $S_K$ , sólo corresponden al contorno de los objetos contenidos en el total de la escena (ver, imagen 5.13). Por lo que, la relación con respecto a los descriptores de los puntos  $S_K$ , que pertenecen a la ROI es menor, en promedio un 5%. Esta última información no se presenta en la gráfica, ya que no es visualmente distinguible. Por otro lado, es justo de esa cantidad de descriptores de donde se lleva a cabo el proceso de detección del objeto, de acuerdo a como se revisó en el algoritmo 2 paso 4. En este sentido, en la tabla 5.6 se muestra una evaluación en relación con el desempeño de la técnica propuesta para la detección de los objetos de interés, considerando 100 pruebas realizadas con imágenes de profundidad (ver ejemplo ilustrativo en 5.12).

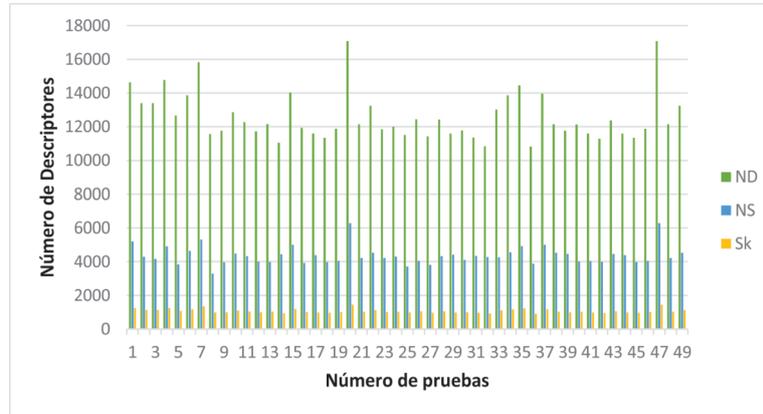


Figura 5.18: Los descriptores  $N_S$  corresponden a la superficie direccional  $M_x$ . Estos han sido seleccionados del vector  $N_D$ . Mientras que los  $S_K$  se obtienen de  $N_S$ .

Tabla 5.6: Relación presentada de los objetos detectados haciendo uso de la técnica propuesta.

Objetos Detectados		
Correctamente	Parcialmente	No Detectados
90	7	3

En relación con la información proporcionada por la tabla 5.6 se establece que un objeto es detectado de forma correcta cuando la ubicación del mismo es identificado en su totalidad, como se ilustra en el primer renglón última columna de la figura 5.12. Mientras que, una detección parcial es como se presenta en la quinta columna en el segundo y tercer renglón de la misma figura. Como se observa cualitativamente, el recuadro que enmarca la botella queda “parcialmente alejado” del objeto, en su lado izquierdo. En el caso de la taza, el recuadro no contiene al objeto en su totalidad, con respecto de la parte asentada en la mesa. En relación con la primera situación esta es dada debido a que existe en la imagen de profundidad una falta de información generada como “pseudo sombra” en la botella. Por el contrario, en la imagen correspondiente a la taza, el problema apunta a una uniformidad en las intensidades de gris, que hace no distinguible la separación entre la mesa y el objeto. Los resultados obtenidos con respecto a la no detección de objetos, de acuerdo con las imágenes analizadas, es porque se fusiona la “sombra” generada por la falta de información alrededor del objeto junto con una área, sin información. Al no existir información, entonces no es posible realizar ningún tipo de análisis que involucre estimaciones probabilísticas o de alguna

otra índole.

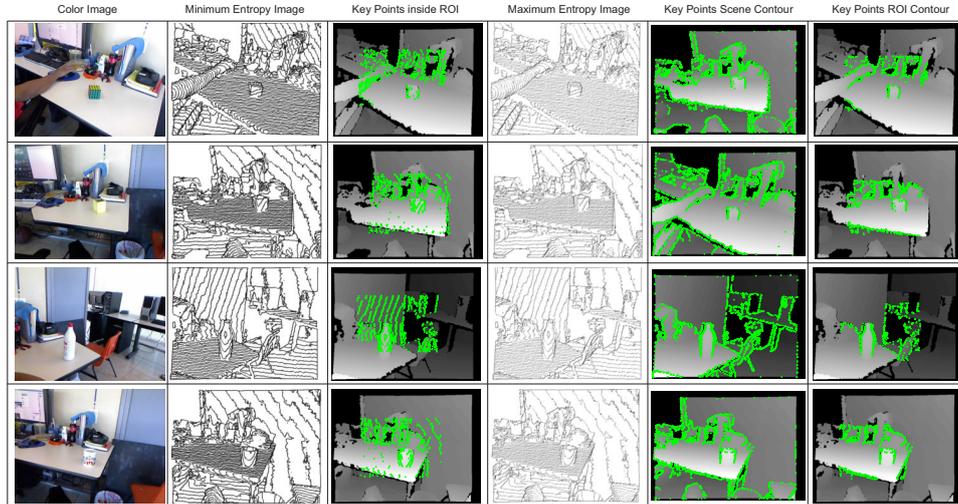


Figura 5.19: Resultados comparativos de la maximización y minimización de la entropía. Primera columna imágenes a color sólo como referencia. De la imagen superior a la inferior: cubo de Rubik, cubo amarillo, bote blanco, tacita, de la base de datos DSPLab. Segunda y tercer columna minimización de la entropía y sus resultados. De la cuarta a la sexta columna los resultados de la maximización.

### 5.3. Resultados del Análisis de la Entropía

La fig. 5.19 presenta los resultados obtenidos de comparar la minimización y maximización de la entropía aplicada a la base de datos DSPLab. En particular, la representación como imagen, de la minimización y maximización (procesos revisados en las subsecciones 4.2.2 y 4.2.3) se muestran en la segunda y cuarta columna. Estas imágenes describen de forma visual a la  $I_{Range}$ , por medio de líneas de iso-elevación, con las cuales se caracteriza un alto o bajo nivel de energía en la imagen [128].

De la imagen 5.19 cualitativamente, los resultados obtenidos correspondientes a la segunda y tercera columna muestran la imagen de energía (minimización de la entropía) y sus correspondientes puntos de interés, respectivamente. El proceso de maximización es mostrado en la cuarta columna a través de su representación como imagen de homogeneidad. Los puntos de interés obtenidos a partir de esta superficie son mostrados en la quinta columna. Finalmente, la sexta columna muestra el contorno correspondiente a los objetos contenidos en la región de interés (ROI), en la cual se encuentra el objeto de interés.

Por lo tanto, a partir de la ec. 4.9 cuando  $\gamma = t_{tun}$  para ambos casos maximización y minimización, el valor de  $t_{tun}$  se establece como una guía de

referencia. Esto es, los descriptores contenidos en la superficie correspondiente a la imagen de energía (ver, segunda columna de la fig. 5.19) presentan bajos niveles de entropía,  $0 \leq N_S < t_{tun}$ . De este modo, los puntos de interés calculados para una región de interés en particular, nos permite obtener puntos característicos al interior de los objetos contenidos en dicha región, pero sin describir el contorno de los mismos (ver, tercer columna correspondiente a la figura 5.19).

Por el contrario, los descriptores calculados a partir de la superficie correspondiente a la imagen de homogeneidad (ver cuarta columna, en la fig. 5.19) tiene un alto valor de la entropía  $t_{tun} < N_S < 1$ . donde,  $N_S$  corresponderá a los descriptores dispersos descritos en la sección 3.3. Tal que, los puntos de interés son obtenidos y definen el contorno de la escena (quinta columna), así como, el contorno de dichos objetos contenidos en la ROI (sexta columna).

Además, es importante hacer notar que, el valor de umbral  $t_{tun}$  calculado a partir de la ec. 4.7 se encuentra en el rango de  $0,6 \leq t_{tun} \leq 0,8$ , siendo que, en la generalidad,  $t_{tun}$  es cercano a 0,7. Este resultado ha sido obtenido, en la mayoría de los casos probados con las dos bases de datos. Adicionalmente, es importante comentar que el pseudocódigo presentado en la sección 3.3.3 es desarrollado con base en las consideraciones hechas para la maximización de la entropía. Apuntando en este caso los pasos 3 y 4 a considerar la función de probabilidad como una *jmpf*, para fusionar la información correspondiente a las ecuaciones revisadas en la sección 4.2.

Ahora bien, los resultados presentados en la tercera columna muestran los puntos de interés más destacados, “al interior” de los objetos contenidos en la ROI. Aquí es importante el mencionar, que se lleva a cabo una sencilla modificación al algoritmo 2 correspondiente a la sección 3.3.4. La cual consiste en colocar en el paso 3.b,

$$S_K = N_S > t_k$$

donde

$$t_k = jmpf(N_S)$$

y

$$t_k \leq t_{tun}$$

mientras que, en la sexta columna se presentan los puntos característicos, pero en este caso como resultado de la maximización de la entropía ( $t_k \geq t_{tun}$ ), definiendo el contorno de los objetos contenidos en esta misma región.

### 5.3.1. Análisis Cuantitativo

La técnica SKD permite llevar a cabo la detección de puntos de interés en una región particular. Esto con el fin de detectar, para posteriormente ubicar la posición del objeto de interés dentro de dicha región. Por lo que,

es necesario primero definir la región de interés o ROI como se muestra en la propuesta hecha a través del algoritmo 2 de la sección 3.3.4. Así pues, si se recapitula en relación con los atributos calculados, primero revisamos que del 100 % de los descriptores estimados cerca del 70 % corresponden a los atributos obtenidos en el proceso denso (3.2) tal como se ilustra 5.20. Mientras que, los descriptores que fueron seleccionados haciendo uso de la técnica propuesta y que se encuentran contenidos en la superficie direccional elegida, forman aproximadamente el 20 % del total de estos. El otro 10 % corresponderá a los descriptores con los cuales se llevará a cabo el cómputo de la región de interés (3.3.4).

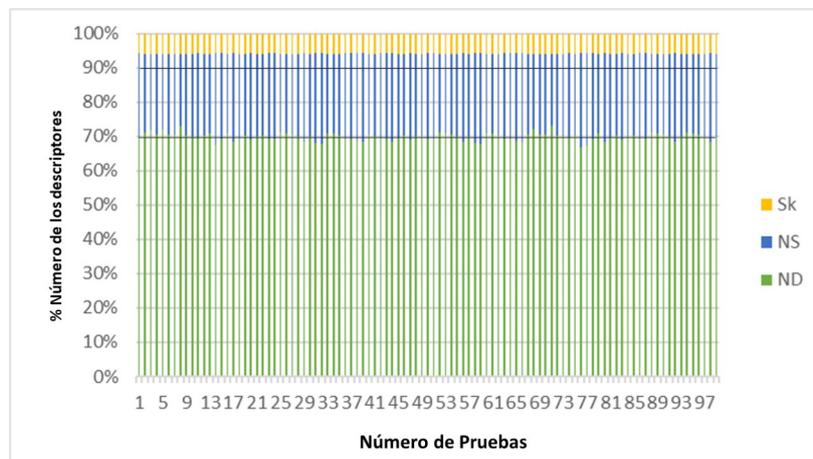


Figura 5.20: Ejemplo comparativo de la cantidad de descriptores contenidos en cada vector de atributos obtenidos en el proceso denso ( $N_D$ ) y disperso  $N_S$  y  $S_K$ .

Con respecto de este último 30 %, en la figura 5.21 se presenta una separación del total de atributos contenidos en la superficie  $M_y$ , como ejemplo. De dicho gráfico los elementos en verde corresponden al 100 % del total de los atributos contenidos en  $M_y$ . Mientras que, la información en azul corresponde, al número de atributos con los que se formó el contorno de cada escena. En amarillo se presenta el total de atributos empleados para obtener el contorno de la ROI. Algunos valores atípicos se han presentado en al menos 2 de 100 pruebas realizadas con la base de datos propia y de Middlebury; los cuales hacen referencia a imágenes con demasiados objetos aglomerados en primer plano.

En la figura 5.22 se encuentra una ampliación de los datos correspondientes a la figura 5.21; los cuales proporcionan el número de atributos con los que se determinaron los puntos de interés para formar los contornos de la escena, así como de la ROI.

Las relaciones que se establecen entre el número de descriptores calcula-

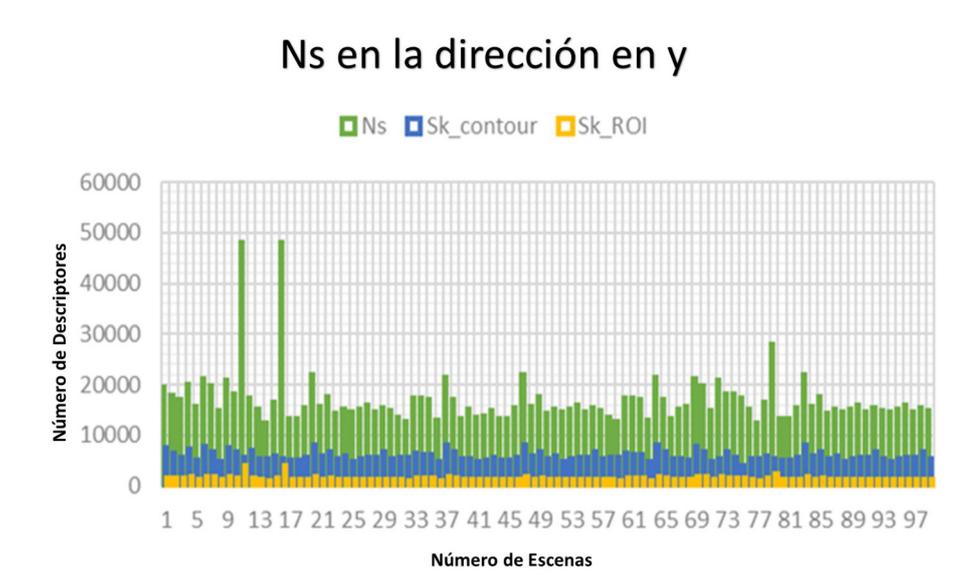


Figura 5.21: Ejemplo ilustrativo del total de atributos contenidos en  $M_y$  en  $N_S$  la segmentación de estos, en  $S_{K_{contour}}$  para el contorno de la escena y de la ROI ( $S_{K_{ROI}}$ ).

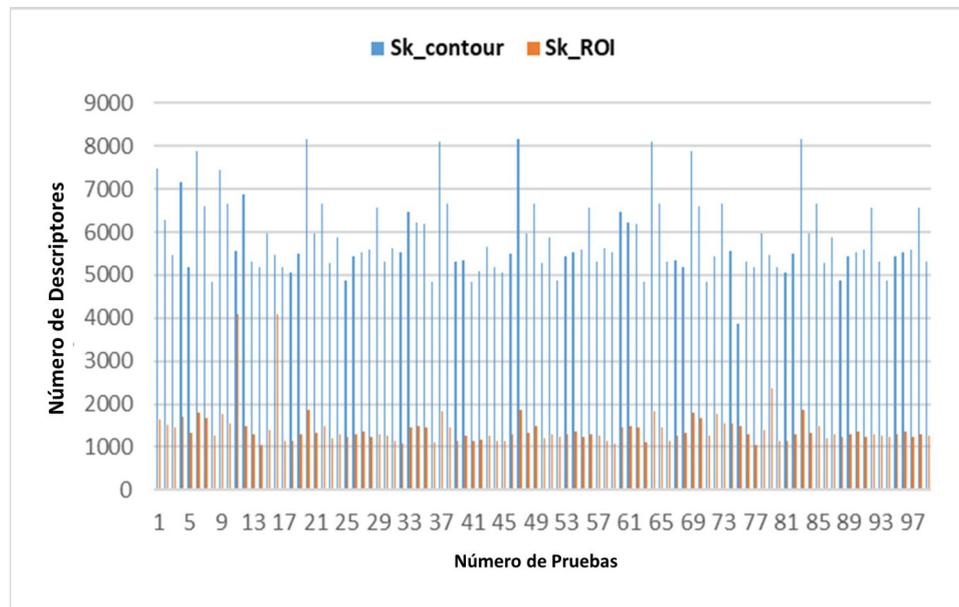


Figura 5.22: Ejemplo ilustrativo de la comparación de los  $S_k$  obtenidos para el contorno de la escena con respecto a los de la ROI.

dos y los seleccionados para la formación de los correspondientes contornos, se pueden observar en la figura 5.24. En este sentido es importante hacer notar, que una imagen de tamaño  $640 \times 480$  píxeles puede tener un total de 307,200 atributos considerando, un sólo atributo por cada posición  $(k, l)$  en  $I_{Range}$ . Si se estima que solamente el 30% de esta información es la que se procesa en la dirección elegida a través del algoritmo 2, se puede presentar en la tabla 5.7 el concentrado de los atributos estimados para obtener los contornos; así como el tiempo de procesamiento con respecto de otras técnicas del trabajos previos el cual es presentado en la tabla 5.8.

Teniendo como referencia la figura 5.24 es posible esquematizar puntualmente el impacto que presenta la disminución en el cómputo y selección de los puntos de interés; con los cuales se pueda definir el contorno de los objetos representados en la escena así como, los contenidos en la región de interés tal como se ilustra en la figura 5.23.

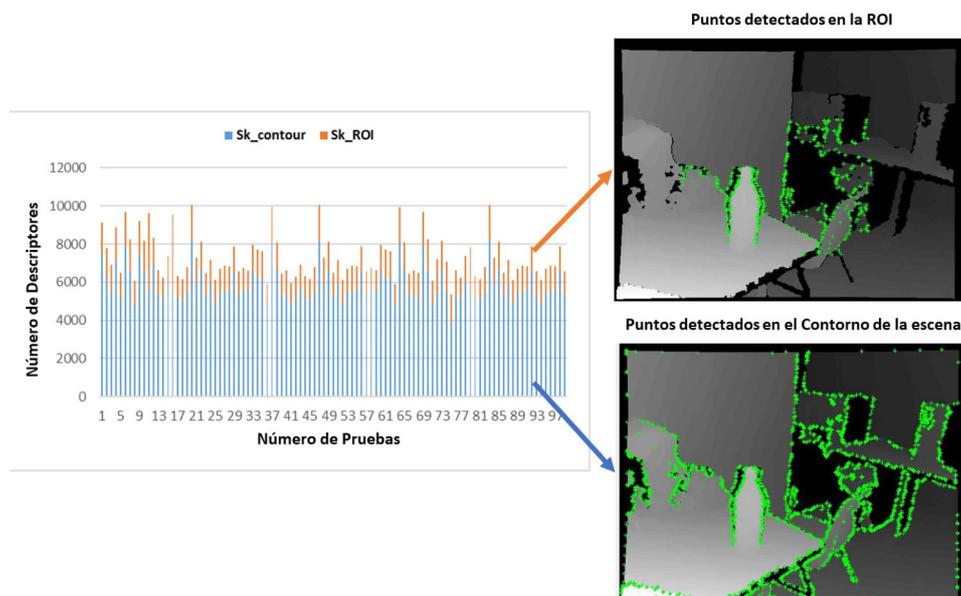


Figura 5.23: Ejemplo esquemático de la cantidad de descriptores requeridos para el cómputo de los puntos de interés que definan el contorno de la escena y la ROI.

Tabla 5.7: Valores promedio del número de descriptores procesados por la técnica propuesta.

Número de Descriptores			
$(N_D)$	$(N_S)$	(Num. $S_{K_{contour}}$ )	(Num. $S_{K_{ROI}}$ )
307,200	16,764	5,854	1,425

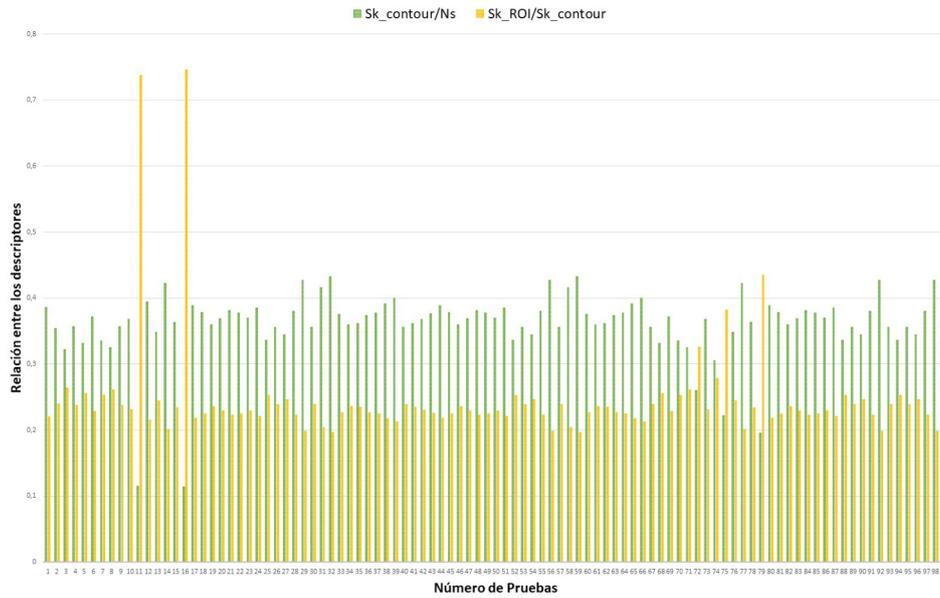
Figura 5.24: Relación establecida entre los descriptores correspondientes a los diferentes contornos obtenidos partiendo de la información contenida en  $M_y$ .

Tabla 5.8: Estimación del tiempo promedio (segundos) de ejecución de la técnica propuesta con respecto de otras técnicas revisadas en la introducción a trabajos previos.

Base de datos ( $K \times L$ )	Tiempo de procesamiento (segundos)				
	SKD	S&T	Harris	SURF	FAST
DSPLab( $640 \times 480$ )	0.213	0.318	0.337	1.084	<b>0.154</b>
Middlebury( $440 \times 370$ )	0.177	0.225	0.386	<b>0.152</b>	0.398
Middlebury( $1390 \times 1100$ )	1.987	2.318	3.097	1.563	<b>0.545</b>

Como se aprecia de la tabla 5.8 el mejor desempeño en relación a tiempos lo presentan las técnicas FAST y SURF. Nuestra técnica por otro lado se encuentra en valores muy cercanos al mejor desempeño, para las dos primeras bases de datos; más aún, se ubica por encima de la otras técnicas ocupando el segundo lugar. Mientras que, para la última, el tiempo prácticamente se triplica esto es,  $0.545 \rightarrow 1.987$  segundos. Aún, cuando la cantidad de puntos de interés calculados no es un factor determinante con respecto de la calidad de los detectores y descriptores, consideramos que, es importante el destacar que nuestra técnica computa una mayor cantidad de puntos de interés sobre toda la imagen, pero, con una menor cantidad de información, tal como se revisó en 3.3.3. En promedio la relación (*razón*) del número de descriptores dispersos calculados con respecto del tiempo de ejecución se presenta en la tabla 5.9. Para ello se toma el número máximo de puntos de interés computado como el 100% en cada caso.

Tabla 5.9: Razón establecida entre el promedio del tiempo de ejecución y el número de descriptores calculados  $\Delta(t)/\Delta(\#descriptores)$ .

	SKD	S&T	Harris	SURF	FAST
<b>Base de datos</b>					
$K \times L$					
	$\Delta(t)/\Delta(\#descriptores)$				
	(segundos/#descriptores)				
<b>DSPLab</b>	0.213/14,064]	0.318/11,194	0.337/8645	1.084/3456	0.154/1056
(640 × 480)					
<b>Middlebury</b>	0.177/1,090	0.225/619	0.386/304	0.152/371	0.398/187
(440 × 370)					
<b>Middlebury</b>	1.987/1,3	2.318/0,6	3.097/0,058	1.563/0,03	0.545/0,0084
(1390 × 1100)					
#descrip × (1 × 10 <sup>6</sup> )					

De acuerdo con la tabla 5.9 el rendimiento de nuestra técnica en relación al tiempo de ejecución es al menos dos veces mejor en promedio, en correlación al promedio del número de descriptores calculados.

## Capítulo 6

# Conclusiones y Trabajos Futuros

*“For a research worker the unforgotten moments of his life are those rare ones which come after years of plodding work, when the veil over nature's secret seems suddenly to lift and when what was dark and chaotic appears in a clear and beautiful light and pattern.”*

Gerty Cori

### 6.1. Conclusiones

En este trabajo de Tesis se presentan dos técnicas de procesamiento de imágenes en niveles de gris, que coadyuvarán en el proceso de enseñanza-aprendizaje de técnicas para el desarrollo de nuevos proyectos prácticos y de divulgación entre los estudiantes y profesores del área afiliados a la carrera de Ingeniería Robótica. Es importante que a través de la aplicación de las técnicas aquí desarrolladas se pueda dar consecución a trabajos de investigación que les permita desarrollar a nuestros jóvenes el espíritu de investigación bajo una metodología científica.

Por lo anterior, se desarrolló una técnica que nos permite ofrecer una herramienta en el área médica, para analizar imágenes de mamografías haciendo uso de los descriptores de textura. Los resultados obtenidos del análisis de mamografías a través de nuestra propuesta nos permite ser un instrumento de apoyo para el radiólogo durante el proceso de un diagnóstico temprano del cáncer de mama. La fiabilidad de esta metodología se ha demostrado al ser comparados sus resultados obtenidos con propuestas similares, estableciendo que su eficiencia sobrepasa el 90% en detecciones correctas de mamografías que contienen microcalcificaciones definidas por el experto como tejido

canceroso. En relación con el error que presenta, este corresponde a falsas detecciones esto es dado en los falsos positivos (detección de calcificaciones benignas como malignas); por lo que nos permite evitar un escenario de un diagnóstico médico menos crítico, que si se obtuvieran resultados de falsos negativos (presencia de tejido maligno no detectado).

En otro orden de ideas, en este trabajo de Tesis se presenta la técnica SKD, la cual nos permite llevar a cabo la detección de puntos de interés a través de una implementación computacional simple. Este punto es fundamental, ya que, nos permite que dicha codificación sea fácil de entender e implementar por los usuarios finales que en este caso son las y los estudiantes de la carrera.

El hacer uso de imágenes de rango permite hacer uso de sistemas de adquisición de bajo costo, como son los sistemas par estéreo o bien el sensor Kinect. Se ha mostrado que, a través del cómputo de sus características de textura de la información adquirida, se obtienen descriptores locales correspondientes a la superficie. Las superficies formadas por los descriptores son divididas en superficies direccionales y la información contenida en estas es tratada como matrices de tipo disperso, lo que decremента de forma significativa la cantidad de información a procesar. Esto implica por lo tanto, una disminución en el tiempo de ejecución.

Al ser esta detección una primera etapa en la localización de objetos 3D, es importante tener puntos de interés validados. En este sentido, nuestra técnica es capaz de detectar puntos con una precisión similar a las propuestas ya establecidas y ampliamente probadas en el área. Analizando estos puntos con nuestra metodología es posible definir el contorno de los objetos contenidos en una escena, haciendo uso de una sola imagen de rango. Más aún, la técnica permite determinar una región que puede considerarse un espacio de trabajo; en el cual, la detección de un objeto 3D se vuelve una tarea posible de resolver, sin importar si es una escena se encuentra bajo condiciones controladas o no controladas. Esto se mostró, con las imágenes correspondientes a nuestra base de datos DSPLab, en cuyo caso la sobreexposición de luz implicaba falta de información en las imágenes de rango (en este caso, imágenes de profundidad).

Con el fin de lograr la detección del objeto, se presentaron tres posibles casos, detección total del objeto, parcial o no detección. Una de las causas de una detección parcial son falta de información generada como “pseudo sombra” en el objeto de interés. Por otra parte, la detección parcial puede darse debido a una homogeneidad en las intensidades de gris; lo que conlleva, a que no sea distinguible la separación del objeto y el medio en el que se encuentre. Los resultados obtenidos con respecto a la no detección de objetos, de acuerdo con las imágenes analizadas, es debido a que se fusiona tanto la “sombra” generada alrededor del objeto como un área sin información. Al no existir datos, entonces no es posible realizar ningún tipo de análisis

que involucre estimaciones probabilísticas o de alguna otra índole. Bajo esta misma perspectiva, es que la detección de los puntos de sujeción para la manipulación del objeto no ha sido la idónea.

Adicionalmente, esta técnica ha sido probada con base de datos genéricas y propias, presentando un desempeño satisfactorio para el uso en aplicaciones didácticas con respecto a otras técnicas clásicas de métodos propuestos.

## 6.2. Trabajos Futuros

En relación con los trabajos futuros, con respecto de la primera propuesta desarrollada en la detección de cáncer de mama, se incluyen más pruebas experimentales de este enfoque utilizando el conjunto de datos INbreast y una nueva caracterización de la función para la obtención de agrupamiento de prominencias, para mejorar los resultados de rendimiento en el tiempo.

En relación con el método propuesto para la detección de objetos de interés, es necesario fortalecer la técnica, tal que se logren mejorar los resultados con respecto a las variaciones en la escala; así como en los tiempos de ejecución para el proceso denso (cálculo de los atributos de textura). Por otro lado, como se mencionó, es necesario llevar a cabo un preprocesamiento de la imagen de profundidad o disparidad que presente áreas o regiones sin información, esto con el fin de mejorar la detección del objeto para su posterior manipulación. En ese sentido, al tener los puntos de sujeción del objeto, será viable la generación de trayectorias del efector final del manipulador, considerando sólo como área de trabajo la región de interés ya determinada por nuestra técnica.

# Bibliografía

*Y así, del mucho leer y del poco dormir,  
se le secó el cerebro de manera que vino  
a perder el juicio.*

-Miguel de Cervantes Saavedra

- [1] *Handbook of Pattern Recognition and Computer Vision*. World Scientific, 1993.
- [2] ABADI, M. S. E. y BANIHASHEMI, N. Automatic classification of travertine stones based on sum and difference histograms algorithm. En *Machine Vision and Image Processing (MVIP), 2015 9th Iranian Conference on*, páginas 132–137. IEEE, 2015.
- [3] AHMAD, S. y CHEONG, L.-F. Robust detection and affine rectification of planar homogeneous texture for scene understanding. *International Journal of Computer Vision*, páginas 1–33, 2018.
- [4] ALAM, N., OLIVER, A., DENTON, E. R. y ZWIGGELAAR, R. Automatic segmentation of microcalcification clusters. En *Annual Conference on Medical Image Understanding and Analysis*, páginas 251–261. Springer, 2018.
- [5] ALI, Y. y HAMED, S. Early breast cancer detection using mammogram images: a review of image processing techniques. *Biosci., Biotech. Res. Asia*, vol. 12, páginas 225–234, 2015.
- [6] ALMANZA, D.-L. y IBARRA, M.-A. *Recent Advances in Mobile Robotics*, capítulo 3D visual information for dynamic object detection and tracking during robot mobile navigation, páginas 71–88. ed. Dr. Andon Venelinov Topalov, InTech, NY, USA, 2011. ISBN 978-953- 307-909-7.
- [7] ANDREARCZYK, V. y WHELAN, P. F. Deep learning for biomedical texture image analysis. En *Proceedings of the Irish Machine Vision & Image Processing Conference*. Irish Pattern Recognition & Classification Society *IPRCS*, 2017.

- 
- [8] ARMSTRONG, M. A. *Basic topology*. Springer Science & Business Media, 2013.
- [9] BALL, J. E. y BRUCE, L. M. Digital mammographic computer aided diagnosis (cad) using adaptive level set segmentation. En *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, páginas 4973–4978. IEEE, 2007.
- [10] BARIYA, P., NOVATNACK, J., SCHWARTZ, G. y NISHINO, K. 3d geometric scale variability in range images: Features and descriptors. *International journal of computer vision*, vol. 99(2), páginas 232–255, 2012.
- [11] BATURONE, A. O. *Robótica: manipuladores y robots móviles*. Marcombo, 2005.
- [12] BELONGIE, S., MALIK, J. y PUZICHA, J. Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, vol. 24(4), páginas 509–522, 2002.
- [13] BENNAMOUN, M., GUO, Y. y SOHEL, F. Feature selection for 2d and 3d face recognition. *Wiley Encyclopedia of Electrical and Electronics Engineering*, páginas 1–28, 2015.
- [14] BENNANI-BAITI, B. y BALTZER, P. A. Mr imaging for diagnosis of malignancy in mammographic microcalcifications: a systematic review and meta-analysis. *Radiology*, vol. 283(3), páginas 692–701, 2016.
- [15] BRANDÃO, S., COSTEIRA, J. P. y VELOSO, M. Combining 3d shape and color for 3d object recognition. En *International Conference Image Analysis and Recognition*, páginas 481–489. Springer, 2016.
- [16] BRONSTEIN, A. M., BRONSTEIN, M. M., GUIBAS, L. J. y OVSJANIKOV, M. Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics (TOG)*, vol. 30(1), página 1, 2011.
- [17] BRONSTEIN, A. M., BRONSTEIN, M. M. y KIMMEL, R. *Numerical geometry of non-rigid shapes*. Springer Science & Business Media, 2008.
- [18] BRONSTEIN, M. M., BRUNA, J., LECUN, Y., SZLAM, A. y VANDERGRHEYNST, P. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, vol. 34(4), páginas 18–42, 2017.
- [19] BURNSIDE, E. S., OCHSNER, J. E., FOWLER, K. J., FINE, J. P., SALKOWSKI, L. R., RUBIN, D. L. y SISNEY, G. A. Use of microcalcification descriptors in bi-rads 4th edition to stratify risk of malignancy. *Radiology*, vol. 242(2), páginas 388–395, 2007.

- [20] BUSTOS, B., KEIM, D. A., SAUPE, D., SCHRECK, T. y VRANIĆ, D. V. Feature-based similarity search in 3d object databases. *ACM Computing Surveys (CSUR)*, vol. 37(4), páginas 345–387, 2005.
- [21] CAI, M., KITANI, K. M. y SATO, Y. Understanding hand-object manipulation with grasp types and object attributes. En *Robotics: Science and Systems*, vol. 3. Ann Arbor, Michigan,; 2016.
- [22] CAI, Z., HAN, J., LIU, L. y SHAO, L. Rgb-d datasets using microsoft kinect or similar sensors: a survey. *Multimedia Tools and Applications*, vol. 76(3), páginas 4313–4355, 2017.
- [23] CHENG, H.-D., SHI, X., MIN, R., HU, L., CAI, X. y DU, H. Approaches for automated detection and classification of masses in mammograms. *Pattern recognition*, vol. 39(4), páginas 646–668, 2006.
- [24] CHEUNG, Y.-C., TSAI, H.-P., LO, Y.-F., UENG, S.-H., HUANG, P.-C. y CHEN, S.-C. Clinical utility of dual-energy contrast-enhanced spectral mammography for breast microcalcifications without associated mass: a preliminary analysis. *European radiology*, vol. 26(4), páginas 1082–1089, 2016.
- [25] CHUANG, G.-H. y KUO, C.-C. Wavelet descriptor of planar curves: Theory and applications. *IEEE Transactions on Image Processing*, vol. 5(1), páginas 56–70, 1996.
- [26] CIDÓN, A. B. U., DE LA TORRE, I. y CIDÓN, E. U. El estándar dicom y su nivel de implantación en europa. *RevistaeSalud. com*, vol. 7(27), páginas 8–11, 2011.
- [27] CITRON, C. Stereo vision system module for low-cost fpgas for autonomous mobile robots. 2014.
- [28] COPAROMAN. 1.1 estructura de los robots manipuladores. 2019. <https://coparoman.blogspot.com/2013/05/estructura-de-los-robots-manipuladores.html>.
- [29] CREUSOT, C., PEARS, N. y AUSTIN, J. A machine-learning approach to keypoint detection and landmarking on 3d meshes. *International Journal of Computer Vision*, vol. 102(1–3), páginas 146–179, 2013.
- [30] CRUZ, L., LUCIO, D. y VELHO, L. Kinect and rgb-d images: Challenges and applications. En *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials*, páginas 36–49. IEEE, 2012.
- [31] CRUZ-BERNAL, A., ALAMANZA-OJEDA, D.-L. y IBARRA-MANZANO, M.-A. Contour detection at range images using sparse normal detector. En *Mexican Conference on Pattern Recognition*, páginas 115–124. Springer, 2016.

- [32] DAVIES, E. R. *Machine vision: theory, algorithms, practicalities*. Elsevier, 2004.
- [33] DHEEBA, J., SINGH, N. A. y SELVI, S. T. Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of biomedical informatics*, vol. 49, páginas 45–52, 2014.
- [34] DHUNGEL, N., CARNEIRO, G. y BRADLEY, A. P. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical image analysis*, vol. 37, páginas 114–128, 2017.
- [35] DOLLÁR, P., RABAUD, V., COTTRELL, G. y BELONGIE, S. Behavior recognition via sparse spatio-temporal features. En *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*, páginas 65–72. IEEE, 2005.
- [36] DOMINGUES, I., SALES, E., CARDOSO, J. y PEREIRA, W. Inbreast-database masses characterization. *XXIII CBEB*, 2012.
- [37] DROESCHEL, D., SCHWARZ, M. y BEHNKE, S. Continuous mapping and localization for autonomous navigation in rough terrain using a 3d laser scanner. *Robotics and Autonomous Systems*, vol. 88, páginas 104–115, 2017.
- [38] FANIZZI, A., BASILE, T., LOSURDO, L., AMOROSO, N., BELLOTTI, R., BOTTIGLI, U., DENTAMARO, R., DIDONNA, V., FAUSTO, A., MASSAFRA, R. ET AL. Hough transform for clustered microcalcifications detection in full-field digital mammograms. En *Applications of Digital Image Processing XL*, vol. 10396, página 1039616. International Society for Optics and Photonics, 2017.
- [39] FENG, J., LIU, X., DONG, Y., LIANG, L. y PU, J. Structural difference histogram representation for texture image classification. *IET Image Processing*, vol. 11(2), páginas 118–125, 2016.
- [40] FROME, A., HUBER, D., KOLLURI, R., BÜLOW, T. y MALIK, J. Recognizing objects in range data using regional point descriptors. En *European conference on computer vision*, páginas 224–237. Springer, 2004.
- [41] GAO, Y. y DAI, Q. View-based 3d object retrieval: challenges and approaches. *IEEE multimedia*, vol. 21(3), páginas 52–57, 2014.
- [42] GARCIA, J. y ZALEVSKY, Z. Range mapping using speckle decorrelation. 2008. US Patent 7,433,024.

- 
- [43] GATYS, L., ECKER, A. S. y BETHGE, M. Texture synthesis using convolutional neural networks. En *Advances in Neural Information Processing Systems*, páginas 262–270. 2015.
- [44] GIBSON, J. J. The perception of the visual world. 1950.
- [45] GOMEZ-BARRERO, M., RATHGEB, C., SCHERHAG, U. y BUSCH, C. Is your biometric system robust to morphing attacks? En *Biometrics and Forensics (IWBF), 2017 5th International Workshop on*, páginas 1–6. IEEE, 2017.
- [46] GONZÁLEZ, R. C. y WOODS, R. E. *Digital Image Processing*, página 954. Pearson, NJ, USA, 2008. ISBN 0-13-168728-x978-0-13-168728-8.
- [47] GUO, Y., BENNAMOUN, M., SOHEL, F., LU, M. y WAN, J. 3d object recognition in cluttered scenes with local surface features: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36(11), páginas 2270–2287, 2014.
- [48] GUO, Y., BENNAMOUN, M., SOHEL, F., LU, M., WAN, J. y KWOK, N. M. A comprehensive performance evaluation of 3d local feature descriptors. *International Journal of Computer Vision*, vol. 116(1), páginas 66–89, 2016.
- [49] GUO, Y., SOHEL, F., BENNAMOUN, M., LU, M. y WAN, J. Rotational projection statistics for 3d local surface description and object recognition. *International journal of computer vision*, vol. 105(1), páginas 63–86, 2013.
- [50] GUO, Z., ZHANG, L. y ZHANG, D. A completed modeling of local binary pattern operator for texture classification. *IEEE transactions on image processing*, vol. 19(6), páginas 1657–1663, 2010.
- [51] HAN, J., L.SHAO, D.XU y SHOTTON, J. Enhanced computer vision with microsoft kinect sensor: A review. *Cybernetics, IEEE Transactions on*, vol. 43(5), páginas 1318–1334, 2013.
- [52] HARALICK, R. M. Statistical and structural approaches to texture. *Proceedings of the IEEE*, vol. 67(5), páginas 786–804, 1979.
- [53] HARALICK, R. M., SHANMUGAM, K., DINSTEIN, I. ET AL. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, vol. 3(6), páginas 610–621, 1973.
- [54] HARALICK, R. M. y SHAPIRO, L. G. *Computer and robot vision*, vol. 1. Addison-wesley Reading, 1992.

- [55] HE, C., LI, S., LIAO, Z. y LIAO, M. Texture classification of polsar data based on sparse coding of wavelet polarization textons. *IEEE Trans. Geoscience and Remote Sensing*, vol. 51(8), páginas 4576–4590, 2013.
- [56] HEDRICH, J., PAULUS, D., GENOIS, F. y GRZEGORZEK, M. Enhanced surface normal computation by exploiting rgb-d sensory information. En *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*, páginas 26–29. IEEE, 2015.
- [57] HERNÁNDEZ-LÓPEZ, J.-J., QUINTANILLA-OLVERA, A.-L., LÓPEZ-RAMÍREZ, J.-L., RANGEL-BUTANDA, F.-J., IBARRA-MANZANO, M.-A. y ALMANZA-OJEDA, D.-L. Detecting objects using color and depth segmentation with kinect sensor. *Procedia Technology*, vol. 3, páginas 196–204, 2012. ISSN 2212-0173.
- [58] HIRSCHMULLER, H. y SCHARSTEIN, D. Evaluation of cost functions for stereo matching. En *Computer Vision and Pattern Recognition, 2007. CVPR07. IEEE Conference on*, páginas 1–8. IEEE, 2007.
- [59] HU, J. y HUA, J. Salient spectral geometric features for shape matching and retrieval. *The visual computer*, vol. 25(5-7), páginas 667–675, 2009.
- [60] IBARRA, M.-A. y ALMANZA, D.-L. *Recent Advances in Stereo Vision*, capítulo 3D visual information for dynamic object detection and tracking during robot mobile navigation, páginas 71–88. ed. Prof. Jose R.A. Torrealo, InTech, NY, USA, 2011. ISBN ISBN: 978-953-307-837-3.
- [61] IBARRA-MANZANO, M.-A., DEVY, M. y BOIZARD, J.-L. Real-time classification based on color and texture attributes on an fpga-based architecture. En *2010 Conference on Design and Architectures for Signal and Image Processing (DASIP)*, páginas 250–257. IEEE, 2010.
- [62] JAIN, A. K., DUIN, R. P. W. y MAO, J. Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22(1), páginas 4–37, 2000.
- [63] JALALIAN, A., MASHOHOR, S., MAHMUD, R., KARASFI, B., SARIPAN, M. I. y RAMLI, A. R. Computer-assisted diagnosis system for breast cancer in computed tomography laser mammography (ctlm). *Journal of digital imaging*, vol. 30(6), páginas 796–811, 2017.
- [64] JALALIAN, A., MASHOHOR, S. B., MAHMUD, H. R., SARIPAN, M. I. B., RAMLI, A. R. B. y KARASFI, B. Computer-aided detection/-diagnosis of breast cancer in mammography and ultrasound: a review. *Clinical imaging*, vol. 37(3), páginas 420–426, 2013.

- [65] JARAMILLO, J. D., MACIÁ, A. M. V. y ZABALA, F. J. C. Métodos directos para la solución de sistemas de ecuaciones lineales simétricos, indefinidos, dispersos y de gran dimensión. *Universidad Eafit*, 2006.
- [66] JAYASUMANA, S., HARTLEY, R., SALZMANN, M., LI, H. y HARANDI, M. Kernel methods on the riemannian manifold of symmetric positive definite matrices. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, páginas 73–80. 2013.
- [67] JEON, J., CHO, S., TONG, X. y LEE, S. Intrinsic image decomposition using structure-texture separation and surface normals. En *European Conference on Computer Vision*, páginas 218–233. Springer, 2014.
- [68] JOHNSON, A. E. y HEBERT, M. Using spin images for efficient object recognition in cluttered 3d scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21(5), páginas 433–449, 1999.
- [69] JULESZ, B. Visual pattern discrimination. *IRE transactions on Information Theory*, vol. 8(2), páginas 84–92, 1962.
- [70] KAPUR, J. N. *Maximum-entropy models in science and engineering*. John Wiley & Sons, 1989.
- [71] KIM, J. K. y PARK, H. W. Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE transactions on medical imaging*, vol. 18(3), páginas 231–238, 1999.
- [72] KNOPP, J., PRASAD, M., WILLEMS, G., TIMOFTE, R. y VAN GOOL, L. Hough transform and 3d surf for robust three dimensional classification. En *Computer Vision-ECCV 2010*, páginas 589–602. Springer, 2010.
- [73] KOPETZ, H. *Real-time systems: design principles for distributed embedded applications*, página 376. Springer Science & Business Media, NY, USA, 2011. ISBN 978-1-4419-8237-7.
- [74] KULLBACK, S. y LEIBLER, R. Letters to the editor. *The American Statistician*, vol. 41(4), páginas 338–341, 1987.
- [75] KUO, H.-Y., SU, H.-R., LAI, S.-H. y WU, C.-C. 3d object detection and pose estimation from depth image for robotic bin picking. En *Automation Science and Engineering (CASE), 2014 IEEE International Conference on*, páginas 1264–1269. IEEE, 2014.
- [76] KURANI, A. S., XU, D.-H., FURST, J. y RAICU, D. S. Co-occurrence matrices for volumetric data. *Heart*, vol. 27, página 25, 2004.
- [77] LACHENBRUCH, P. A. y MICKEY, M. R. Estimation of error rates in discriminant analysis. *Technometrics*, vol. 10(1), páginas 1–11, 1968.

- [78] LAPTEV, I. On space-time interest points. *International journal of computer vision*, vol. 64(2-3), páginas 107–123, 2005.
- [79] LATEGAHN, H., GROSS, S., STEHLE, T. y AACH, T. Texture classification by modeling joint distributions of local patterns with gaussian mixtures. *IEEE Transactions on Image Processing*, vol. 19(6), páginas 1548–1557, 2010.
- [80] LAURIA, A., PALMIERO, R., FORNI, G., CERELLO, P., GOLOSIO, B., FAUCI, F., MAGRO, R., RASO, G., TANGARO, S. y INDOVINA, P. L. The calma system: an artificial neural network method for detecting masses and microcalcifications in digitized mammograms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 518(1-2), páginas 391–393, 2004.
- [81] LAWS, K. I. Textured image segmentation. Informe técnico, DTIC Document, 1980.
- [82] LAZEBNIK, S., SCHMID, C. y PONCE, J. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27(8), páginas 1265–1278, 2005.
- [83] LIM, J. y LEE, K. 3d object recognition using scale-invariant features. *The Visual Computer*, páginas 1–14, 2017.
- [84] LIN, B., WANG, F., SUN, Y., QU, W., CHEN, Z. y ZHANG, S. Boundary points based scale invariant 3d point feature. *Journal of Visual Communication and Image Representation*, vol. 48, páginas 136–148, 2017.
- [85] LIU, L., CHEN, J., FIEGUTH, P., ZHAO, G., CHELLAPPA, R. y PIETIKAINEN, M. A survey of recent advances in texture representation. *arXiv preprint arXiv:1801.10324*, 2018.
- [86] LIU, L. y FIEGUTH, P. Texture classification from random features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34(3), páginas 574–586, 2012.
- [87] LLOYD, K., ROSIN, P. L., MARSHALL, D. y MOORE, S. C. Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (glcm)-based texture measures. *Machine Vision and Applications*, vol. 28(3-4), páginas 361–371, 2017.
- [88] LO, T.-W. R. y SIEBERT, J. P. Local feature extraction and matching on range images: 2.5 d sift. *Computer Vision and Image Understanding*, vol. 113(12), páginas 1235–1250, 2009.

- [89] LONCOMILLA, P., RUIZ-DEL SOLAR, J. y MARTÍNEZ, L. Object recognition using local invariant features for robotic applications: A survey. *Pattern Recognition*, vol. 60, páginas 499–514, 2016.
- [90] LOPEZ-ALIGUE, F. J., POVEDA-PIEROLA, A., ACEVEDO-SOTOCA, I. y GARCIA-URRA, F. Detection of microcalcifications in digital mammograms. En *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, páginas 3906–3909. IEEE, 2007.
- [91] LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, vol. 60(2), páginas 91–110, 2004.
- [92] LUO, S., YUAN, W., ADELSON, E., COHN, A. G. y FUENTES, R. Vitac: Feature sharing between vision and tactile sensing for cloth texture recognition. *arXiv preprint arXiv:1802.07490*, 2018.
- [93] MA, L., TAN, T., WANG, Y. y ZHANG, D. Personal identification based on iris texture analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12), páginas 1519–1533, 2003.
- [94] MAHMOUD-GHONEIM, D., TOUSSAINT, G., CONSTANS, J.-M. y JACQUES, D. Three dimensional texture analysis in mri: a preliminary evaluation in gliomas. *Magnetic resonance imaging*, vol. 21(9), páginas 983–987, 2003.
- [95] MALASSIOTIS, S. y STRINTZIS, M. G. Snapshots: A novel local surface descriptor and matching algorithm for robust 3d surface alignment. *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29(7), páginas 1285–1290, 2007.
- [96] MARTON, Z.-C., BALINT-BENCZEDI, F., SEIDEL, F., GORON, L. C. y BEETZ, M. Object categorization in clutter using additive features and hashing of part-graph descriptors. En *International Conference on Spatial Cognition*, páginas 17–33. Springer, 2012.
- [97] MATHWORKS. Image processign toolbox. 2019. <https://la.mathworks.com/products/image.html>.
- [98] MEEKS, W., PÉREZ, J. y PÉREZ, J. *A survey on classical minimal surface theory*, vol. 60. American Mathematical Society, 2012.
- [99] MEL, B. W. Seemore: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural computation*, vol. 9(4), páginas 777–804, 1997.

- [100] MEZIED, A. A. y ALATTAR, A. Medical image retrieval based on gray cluster co-occurrence matrix and edge strength levels. En *Promising Electronic Technologies ICPET, 2017 International Conference on*, páginas 71–76. IEEE, 2017.
- [101] MIAN, A., BENNAMOUN, M. y OWENS, R. On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *International Journal of Computer Vision*, vol. 89(2-3), páginas 348–361, 2010.
- [102] MIAN, A. S., BENNAMOUN, M. y OWENS, R. Three-dimensional model-based object recognition and segmentation in cluttered scenes. *IEEE transactions on pattern analysis and machine intelligence*, vol. 28(10), páginas 1584–1601, 2006.
- [103] MICROSOFT. The kinect effect- how the world is using kinect. 2016. <http://www.xbox.com/en-GB/kinect>.
- [104] MORALES, L. A., SOTOMAYOR, N. S. y BOADA, Y. Control de seguimiento de trayectoria y paletización de un robot de tres grados de libertad tipo scara (selective compliance assembly robot arm). *Revista Politécnica*, vol. 33(1), 2014.
- [105] MORELAND, K. Is black a color? 2013. <http://www.drmoron.org/is-black-a-color/>.
- [106] MUJA, M. y LOWE, D. G. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP (1)*, vol. 2(2), páginas 331–340, 2009.
- [107] DO NASCIMENTO, E. R., OLIVEIRA, G. L., VIEIRA, A. W. y CAMPOS, M. F. On the development of a robust, fast and lightweight keypoint descriptor. *Neurocomputing*, vol. 120, páginas 141–155, 2013.
- [108] NEX, F. y REMONDINO, F. Uav for 3d mapping applications: a review. *Applied geomatics*, vol. 6(1), páginas 1–15, 2014.
- [109] OJALA, T., PIETIKAINEN, M. y MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24(7), páginas 971–987, 2002.
- [110] OTHMEN, E. B., SAYADI, M. y FNIAECH, F. 3d gray level co-occurrence matrices for volumetric texture classification. En *Systems and Control (ICSC), 2013 3rd International Conference on*, páginas 833–837. IEEE, 2013.

- [111] OYALLON, E. y MALLAT, S. Deep roto-translation scattering for object classification. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, páginas 2865–2873. 2015.
- [112] PAL, C. J., WEINMAN, J. J., TRAN, L. C. y SCHARSTEIN, D. On learning conditional random fields for stereo. *International Journal of Computer Vision*, vol. 99(3), páginas 319–337, 2012.
- [113] RAAD, L., DAVY, A., DESOLNEUX, A. y MOREL, J.-M. A survey of exemplar-based texture synthesis. *arXiv preprint arXiv:1707.07184*, 2017.
- [114] REED, T. R. y WECHSLER, H. Segmentation of textured images and gestalt organization using spatial/spatial-frequency representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12(1), páginas 1–12, 1990.
- [115] ROBTICS, R. Robot manipulador para 3 grados de libertad. 2019. <https://rivasrobotics.com/robot-manipulador-secuencial/>.
- [116] RODOLÀ, E., ALBARELLI, A., CREMERS, D. y TORSELLO, A. A simple and effective relevance-based point sampling for 3d shapes. *Pattern Recognition Letters*, vol. 59, páginas 41–47, 2015.
- [117] ROG. Asus,xtion pro live. 2016. [https://www.asus.com/3D-Sensor/Xtion\\_PRO\\_LIVE/](https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/).
- [118] ROMERO, A., GOUIFFÈS, M. y LACASSAGNE, L. Enhanced local binary covariance matrices (elbcm) for texture analysis and object tracking. En *Proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*, página 10. ACM, 2013.
- [119] RUSS, J. C. *The image processing handbook*. CRC press, 2016.
- [120] SAAD, Y. *Iterative methods for sparse linear systems*, vol. 82, página 460. siam, 2003.
- [121] SALAMON, P., NULTON, J. y IHRIG, E. On the relation between entropy and energy versions of thermodynamic length. *The Journal of chemical physics*, vol. 80(1), páginas 436–437, 1984.
- [122] SAMALA, R. K., CHAN, H.-P., HADJIISKI, L. M., CHA, K. y HELVIE, M. A. Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis. En *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785, página 97850Y. International Society for Optics and Photonics, 2016.

- [123] SAMALA, R. K., CHAN, H.-P., HADJIISKI, L. M. y HELVIE, M. A. Analysis of computer-aided detection techniques and signal characteristics for clustered microcalcifications on digital mammography and digital breast tomosynthesis. *Physics in Medicine & Biology*, vol. 61(19), página 7092, 2016.
- [124] SAMPSON, D. J., CHANG, Y. K., RUPASINGHE, H. V. y ZAMAN, Q. U. A dual-view computer-vision system for volume and image texture analysis in multiple apple slices drying. *Journal of Food Engineering*, vol. 127, páginas 49–57, 2014.
- [125] SÁNCHEZ, J. A. S. *Avances en robótica y visión por computador*, vol. 38. Univ de Castilla La Mancha, 2002.
- [126] SANDI, F. y DOUIK, A. Dominant and minor sum and difference histograms for texture description. En *Image Processing, Applications and Systems (IPAS), 2016 International*, páginas 1–5. IEEE, 2016.
- [127] SANDID, F. y DOUIK, A. Robust color texture descriptor for material recognition. *Pattern Recognition Letters*, vol. 80, páginas 15–23, 2016.
- [128] SAPIRO, G. *Geometric partial differential equations and image analysis*. Cambridge university press, 2006.
- [129] SAXENA, A., DRIEMEYER, J. y NG, A. Y. Robotic grasping of novel objects using vision. *The International Journal of Robotics Research*, vol. 27(2), páginas 157–173, 2008.
- [130] SERNA, W. S., LEMUS, J. P. T. y PIEDRAHITA, J. H. R. Descripción del estándar dicom para un acceso confiable a la información de las imágenes médicas. *Scientia et Technica*, vol. 16(45), páginas 289–294, 2010.
- [131] SHANNON, C. E. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5(1), páginas 3–55, 2001.
- [132] SHI, J., SAHINER, B., CHAN, H.-P., GE, J., HADJIISKI, L., HELVIE, M. A., NEES, A., WU, Y.-T., WEI, J., ZHOU, C. ET AL. Characterization of mammographic masses based on level set segmentation with new image features and patient information. *Medical physics*, vol. 35(1), páginas 280–290, 2008.
- [133] SHOTTON, J., WINN, J., ROTHER, C. y CRIMINISI, A. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, vol. 81(1), páginas 2–23, 2009.

- [134] SIMONYAN, K., PARKHI, O. M., VEDALDI, A. y ZISSERMAN, A. Fisher vector faces in the wild. En *BMVC*, vol. 2, página 4. 2013.
- [135] SMISEK, J., JANCOSEK, M. y PAJDLA, T. 3d with kinect. En *Consumer depth cameras for computer vision*, páginas 3–25. Springer, 2013.
- [136] STEDER, B., RUSU, R. B., KONOLIGE, K. y BURGARD, W. Narf: 3d range image features for object recognition. En *Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics at the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, vol. 44. 2010.
- [137] STEDER, B., RUSU, R. B., KONOLIGE, K. y BURGARD, W. Point feature extraction on 3d range scans taking into account object boundaries. En *Robotics and automation (icra), 2011 ieee international conference on*, páginas 2601–2608. IEEE, 2011.
- [138] STEVENS, K. A. Surface perception from local analysis of texture. En *AI-Tech. Rep.-512*. AI Laboratory, Massachusetts Institute of Technology Cambridge, 1980.
- [139] *Sting*, D. E. G. normal vector tangent. 2008. <https://commons.wikimedia.org/w/index.php?curid=1325452>.
- [140] SUN, J., OVSJANIKOV, M. y GUIBAS, L. A concise and provably informative multi-scale signature based on heat diffusion. En *Computer graphics forum*, vol. 28, páginas 1383–1392. Wiley Online Library, 2009.
- [141] SZCZYPIŃSKI, P. M. y KLEPACZKO, A. Mazda—a framework for biomedical image texture analysis and data exploration. En *Biomedical Texture Analysis*, páginas 315–347. Elsevier, 2018.
- [142] TABIA, H., LAGA, H., PICARD, D. y GOSSELIN, P.-H. Covariance descriptors for 3d shape matching and retrieval. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, páginas 4185–4192. 2014.
- [143] TEAGUE, M. R. Image analysis via the general theory of moments. *JOSA*, vol. 70(8), páginas 920–930, 1980.
- [144] TOMBARI, F. y DI STEFANO, L. Hough voting for 3d object recognition under occlusion and clutter. *IPSI Transactions on Computer Vision and Applications*, vol. 4, páginas 20–29, 2012.
- [145] TOMBARI, F., SALTI, S. y DI STEFANO, L. Unique signatures of histograms for local surface description. En *European conference on computer vision*, páginas 356–369. Springer, 2010.

- [146] TOMBARI, F., SALTI, S. y DI STEFANO, L. A combined texture-shape descriptor for enhanced 3d feature matching. En *Image Processing (ICIP), 2011 18th IEEE International Conference on*, páginas 809–812. IEEE, 2011.
- [147] TOMBARI, F., SALTI, S. y DI STEFANO, L. Performance evaluation of 3d keypoint detectors. *International Journal of Computer Vision*, vol. 102(1-3), páginas 198–220, 2013.
- [148] TUYTELAARS, T., MIKOLAJCZYK, K. ET AL. Local invariant feature detectors: a survey. *Foundations and trends® in computer graphics and vision*, vol. 3(3), páginas 177–280, 2008.
- [149] UNSER, M. Local linear transforms for texture measurements. *Signal processing*, vol. 11(1), páginas 61–79, 1986.
- [150] UNSER, M. Sum and difference histograms for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1), páginas 118–125, 1986.
- [151] VAN ZELST, J., TAN, T., PLATEL, B., DE JONG, M., STEENBAKERS, A., MOURITS, M., GRIVEGNEE, A., BORELLI, C., KARSSEMEIJER, N. y MANN, R. Improved cancer detection in automated breast ultrasound by radiologists using computer aided detection. *European journal of radiology*, vol. 89, páginas 54–59, 2017.
- [152] VARELA, C., TIMP, S. y KARSSEMEIJER, N. Use of border information in the classification of mammographic masses. *Physics in medicine & biology*, vol. 51(2), página 425, 2006.
- [153] VARMA, M. y ZISSERMAN, A. A statistical approach to texture classification from single images. *International journal of computer vision*, vol. 62(1-2), páginas 61–81, 2005.
- [154] VIVAS, A. Robótica paralela: aplicaciones industriales, modelado y control. *Universidad del Cauca. Popayán*, 2005.
- [155] WANG, H., HU, J. y DENG, W. Face feature extraction: A complete review. *IEEE Access*, vol. 6, páginas 6001–6039, 2018.
- [156] WANG, J., YANG, X., CAI, H., TAN, W., JIN, C. y LI, L. Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Scientific reports*, vol. 6, página 27327, 2016.
- [157] WANG, J., YANG, Y. y NISHIKAWA, R. M. Quantitative study of image features of clustered microcalcifications in for-presentation mammograms. En *2016 IEEE International Conference on Image Processing (ICIP)*, páginas 3404–3408. IEEE, 2016.

- [158] WU, S., OERLEMANS, A., BAKKER, E. M. y LEW, M. S. A comprehensive evaluation of local detectors and descriptors. *Signal Processing: Image Communication*, vol. 59, páginas 150–167, 2017.
- [159] XIE, X. y MIRMEHDI, M. Texems: Texture exemplars for defect detection on random textured surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29(8), páginas 1454–1464, 2007.
- [160] YAHIA, S., SALEM, Y. B. y ABDELKRIM, M. N. 3d face recognition using local binary pattern and grey level co-occurrence matrix. En *Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2016 17th International Conference on*, páginas 328–338. IEEE, 2016.
- [161] YANG, J., ZHANG, Q., XIAN, K., XIAO, Y. y CAO, Z. Rotational contour signatures for both real-valued and binary feature representations of 3d local shape. *Computer Vision and Image Understanding*, vol. 160, páginas 133–147, 2017.
- [162] YANG, X., TRIDANDAPANI, S., BEITLER, J. J., DAVID, S. Y., YOSHIDA, E. J., CURRAN, W. J. y LIU, T. Ultrasound glcm texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity. *Medical physics*, vol. 39(9), páginas 5732–5739, 2012.
- [163] YOU, S. y HUANG, J. Recognizing objects in 3d data with distinctive self-similarity features. En *Automatic Target Recognition XXVIII*, vol. 10648, página 106480D. International Society for Optics and Photonics, 2018.
- [164] ZHANG, J., MARSZALEK, M., LAZEBNIK, S. y SCHMID, C. Local features and kernels for classification of texture and object categories: A comprehensive study. *International journal of computer vision*, vol. 73(2), páginas 213–238, 2007.
- [165] ZHAO, G. y PIETIKAINEN, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE transactions on pattern analysis and machine intelligence*, vol. 29(6), páginas 915–928, 2007.
- [166] ZHOU, K., RICHTSFELD, A., ZILlich, M., VINCZE, M., VRECKO, A. y SKOCAJ, D. Visual information abstraction for interactive robot learning. En *Advanced Robotics (ICAR), 2011 15th International Conference on*, páginas 328–334. IEEE, Tallinn, 2011. ISBN 978-1-4577-1158-9.

- 
- [167] ZHOU, W., MA, C., YAO, T., CHANG, P., ZHANG, Q. y KUIJPER, A. Histograms of gaussian normal distribution for 3d feature matching in cluttered scenes. *The Visual Computer*, páginas 1–17, 2018.
- [168] ZHU, S. C., WU, Y. y MUMFORD, D. Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. *International Journal of Computer Vision*, vol. 27(2), páginas 107–126, 1998.
- [169] ZHU, W., LOU, Q., VANG, Y. S. y XIE, X. Deep multi-instance networks with sparse label assignment for whole mammogram classification. En *International Conference on Medical Image Computing and Computer-Assisted Intervention*, páginas 603–611. Springer, 2017.
- [170] ZOU, G., HUA, J., LAI, Z., GU, X. y DONG, M. Intrinsic geometric scale space by shape diffusion. *IEEE transactions on visualization and computer graphics*, vol. 15(6), 2009.



