

***“ALGORITMO PARA DETERMINAR LA
COMPLEJIDAD DE IMÁGENES APLICADO A
SEGMENTACIÓN”***

TESIS

QUE PARA OBTENER EL GRADO DE

***DOCTOR EN CIENCIAS EN INGENIERÍA
ELECTRÓNICA***

PRESENTA:

M.C. LUIS MADRID HERRERA

**DIRECTOR DE LA TESIS:
*DR. MARIO IGNACIO CHACÓN MURGUÍA***

CHIHUAHUA, CHIH., NOVIEMBRE 2022



Chihuahua, Chih., 17 de noviembre de 2022

**LUIS MADRID HERRERA
PRESENTE**

En cumplimiento con los requerimientos para la obtención del grado de Doctor en Ciencias en Ingeniería Electrónica y a propuesta de su Comité Tutorial, la División de Estudios de Posgrado e Investigación le concede la autorización para imprimir la tesis titulada **"Algoritmo para determinar la complejidad de imágenes aplicado a segmentación"**, dirigida por el Dr. Mario Ignacio Chacón Murguía con el siguiente contenido de capítulos:

- I Antecedentes
- II Complejidad de imagen y análisis del criterio humano
- III Determinación de la complejidad de imagen
- IV Complejidad de imagen para determinar el número de grupos en algoritmos de segmentación de imágenes
- V Enfoque difuso para la evaluación de la segmentación de imágenes basado en la complejidad de imagen
- VI Resultados y conclusiones generales

ATENTAMENTE
Excelencia en Educación Tecnológica®
"La técnica por el Engrandecimiento de México"



ROGELIO ENRIQUE PARAY ARANA
JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN





Chihuahua, Chih., 15 de noviembre de 2022

ROGELIO ENRIQUE BARAY ARANA
JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN
PRESENTE

En cumplimiento con los requerimientos para la obtención del grado de Doctor en Ciencias en Ingeniería Electrónica, le notificamos que el documento de tesis del alumno **LUIS MADRID HERRERA** titulado **"Algoritmo para determinar la complejidad de imágenes aplicado a segmentación"**, dirigido por el Dr. Mario Ignacio Chacón Murguía, ha sido aprobado y aceptado para su impresión.

Por lo anterior, proponemos le sea concedida la autorización de impresión correspondiente.

Agradeciendo la atención a la presente, quedamos de usted:

ATENTAMENTE

Excelencia en Educación Tecnológica®
"La técnica por el Engrandecimiento de México"

DR. MARIO IGNACIO CHACÓN MURGUÍA
DIRECTOR DE TESIS

DRA. GRACIELA MARÍA DE JESÚS RAMÍREZ ALONSO
MIEMBRO DEL COMITÉ TUTORIAL

DR. JUAN ALBERTO RAMÍREZ QUINTANA
MIEMBRO DEL COMITÉ TUTORIAL

DR. LUIS FRANCISCO CORRAL MARTÍNEZ
MIEMBRO DEL COMITÉ TUTORIAL

DR. JESÚS ALFONSO MEDRANO HERMOSILLO
MIEMBRO DEL COMITÉ TUTORIAL





CARTA CESIÓN DE DERECHOS

En la ciudad de Chihuahua el día 17 de noviembre de 2022, la que suscribe, LUIS MADRID HERRERA con número de control D18060575, del Doctorado en Ciencias en Ingeniería Electrónica adscrito a la División de Estudios de Posgrado e Investigación del Instituto Tecnológico de Chihuahua, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección de el Dr. Mario Ignacio Chacón Murguía y cede los derechos del trabajo titulado **"Algoritmo para determinar la complejidad de imágenes aplicado a segmentación"**, al Tecnológico Nacional de México y/o Instituto Tecnológico de Chihuahua para su difusión, divulgación, transmisión, reproducción, así como su digitalización con fines académicos y de investigación.

LUIS MADRID HERRERA





DECLARACIÓN DE ORIGINALIDAD

En la ciudad de Chihuahua el día 17 de noviembre de 2022, la que suscribe, LUIS MADRID HERRERA con número de control D18060575, del Doctorado en Ciencias en Ingeniería Electrónica adscrito a la División de Estudios de Posgrado e Investigación del Instituto Tecnológico de Chihuahua, manifiesta ser autor intelectual del presente trabajo de Tesis titulado **"Algoritmo para determinar la complejidad de imágenes aplicado a segmentación"**, bajo la dirección de el Dr. Mario Ignacio Chacón Murguía; y manifiesta que el contenido es original y que las fuentes de información consultadas para su fundamentación están debidamente citadas y referenciadas.

LUIS MADRID HERRERA



Chihuahua, Chihuahua, noviembre de 2022

Dra. María Elena Álvarez-Buylla Roces
Directora de Conacyt

Sirva la presente para saludarla e informarle que a la fecha he obtenido el grado de **Doctor en Ciencias en Ingeniería Electrónica** en la División de Estudios de Posgrado e Investigación del Instituto Tecnológico de Chihuahua, motivo por el cual agradezco todo el apoyo brindado por este Consejo que usted representa, con el otorgamiento de una beca que permitió dedicarme de manera exclusiva a la realización de mis estudios de posgrado y de esta manera lograr el objetivo principal del convenio establecido.

Sin otro particular quedo a sus órdenes no sin antes reiterar mi agradecimiento.

Atentamente

A handwritten signature in dark ink, appearing to read 'Luis Madrid Herrera', with a stylized flourish at the end.

Luis Madrid Herrera

RESUMEN

ALGORITMO PARA DETERMINAR LA COMPLEJIDAD DE IMÁGENES APLICADO A SEGMENTACIÓN

M.C. Luis Madrid Herrera

Doctorado en Ciencias en Ingeniería Electrónica

División de Estudios de Posgrado e Investigación del

Tecnológico Nacional de México/I.T. Chihuahua

Chihuahua, Chih., 2022

Director de Tesis: Dr. Mario Ignacio Chacón Murguía

Esta tesis presenta un estudio para determinar la complejidad de imagen para su aplicación en algoritmos de segmentación de imágenes. Primero, se presenta la revisión de la literatura respecto a la complejidad de imagen para proponer una nueva definición conceptual y operacional de la complejidad de imagen, que traslada el concepto desde el espacio multidimensional de la percepción visual humana al espacio unidimensional de la respuesta a la complejidad de imagen. De tal manera que la respuesta a la complejidad de imagen es definida como una medida relativa y subjetiva de la cantidad de información semántica contenida en las imágenes, dentro de un conjunto determinado de imágenes. Luego, se proponen dos métodos computacionales para estimar la respuesta a la complejidad de imagen. Un método es basado en un sistema de inferencia difuso tipo Sugeno y el otro método se basa en una red neuronal artificial. Ambos métodos son alimentados mediante 11 características visuales de las imágenes procesadas por el algoritmo de análisis de componentes principales. Una vez obtenido el valor de respuesta a la complejidad de imagen, se proponen dos aplicaciones prácticas en los algoritmos de segmentación de imágenes. La primera aplicación trata sobre la determinación automática del número de grupos necesarios para segmentar una imagen mediante los algoritmos de agrupamiento difuso. La segunda aplicación propone un nuevo enfoque de evaluación difuso para los algoritmos de segmentación de imágenes.

De acuerdo con los resultados obtenidos, los métodos para estimar la respuesta a la complejidad de imagen computacionalmente tienen una alta correlación con el criterio humano, comparados contra otros métodos del estado del arte. La aplicación para determinar el número de grupos en los algoritmos de segmentación de imágenes por agrupamiento difuso demuestra ser de gran utilidad al mejorar los desempeños de algunos algoritmos del estado del arte. Por su

parte, el nuevo enfoque difuso de evaluación, mejora en diversos aspectos la evaluación de los algoritmos de segmentación de imágenes, ya que considera la subjetividad del criterio humano al realizar esta tarea tan importante en el área de visión por computadora.

CONTENIDO

LISTA DE FIGURAS	vii
LISTA DE TABLAS	x
GLOSARIO DE SÍMBOLOS	xii
INTRODUCCIÓN.....	1
CAPÍTULO 1.ANTECEDENTES	4
1.1 Respuesta a la complejidad de imagen	4
1.1.1 Enfoques computacionales.....	5
1.1.1.1 Información de bordes.....	5
1.1.1.2 Información del color y/o intensidad.....	6
1.1.1.3 Grado de compresión.....	8
1.1.1.4 Redes neuronales convolucionales.....	8
1.1.1.5 Combinado	9
1.1.2 Enfoques humanos	14
1.1.2.1 Criterio humano.....	14
1.1.2.2 Reacción humana	17
1.1.3 Bases de datos relacionadas a la complejidad de imágenes.....	20
1.2 Algoritmos de segmentación por agrupamiento difuso	24
1.2.1 Base de datos BSDS500.....	26
1.2.1.1 Documentación de la base de datos.....	26
1.2.1.2 Métricas de evaluación de la base de datos.....	27
CAPÍTULO 2.COMPLEJIDAD DE IMAGEN Y ANÁLISIS DEL CRITERIO HUMANO..	30
2.1 Definición de la respuesta a la complejidad de imagen propuesta	30
2.2 Definición operacional de la respuesta a la complejidad de imagen	31
2.2.1 Propiedades de la RCI.....	32
2.2.2 Resolución espacial y relación de aspecto	33
2.3 Análisis de la subjetividad del criterio humano.....	33
2.4 Coherencia del criterio humano en B_{RS1} y B_{RS2}	38
CAPÍTULO 3.DETERMINACIÓN DE LA COMPLEJIDAD DE IMAGEN	41

3.1 Características	41
3.1.1 Análisis de características	45
3.1.2 Preprocesamiento de características.....	50
3.2 Sistema de inferencia difuso para determinar la complejidad de imagen.....	51
3.2.1 Diseño del sistema difuso.....	52
3.2.2 Resultados del sistema difuso en B_{RS1} y B_{RS2}	54
3.3 Red neuronal artificial para determinar complejidad de imágenes.....	55
3.3.1 Diseño de la red neuronal artificial	56
3.3.2 Resultados y discusión del método basado en redes neuronales artificiales.....	59
3.3.2.1 Resultados en la base de datos B_{RS1} y B_{RS2}	59
3.3.2.2 Resultados en la base de datos $B_{SAVOIAS}$	61
3.4 Comparación de resultados con otros métodos.....	65
3.5 Conclusiones	66
CAPÍTULO 4.COMPLEJIDAD DE IMAGEN PARA DETERMINAR EL NÚMERO DE GRUPOS EN ALGORITMOS DE SEGMENTACIÓN DE IMÁGENES	68
4.1 Motivación para utilizar la complejidad de imagen.....	71
4.2 Problemas de utilizar un número de grupos fijo	71
4.3 Diseño de AENCIC	73
4.3.1 Relación entre complejidad de imagen y número de grupos	73
4.3.2 Determinación automática del número de grupos.....	76
4.4 Experimentos y resultados	79
4.4.1 Verificación de los códigos de los algoritmos de segmentación por agrupamiento y configuración de parámetros	79
4.4.1.1 FRFCM.....	80
4.4.1.2 SFFCM	80
4.4.1.3 Algoritmos de segmentación AFCE	80
4.4.1.4 FAS-SGC.....	83
4.4.2 Análisis de resultados.....	83
4.4.2.1 Resultados de FRFCM	85
4.4.2.2 Resultados de SFFCM.....	86

4.4.2.3 Resultados de las variantes de AFCF	86
4.4.3 Análisis de correlación entre los valores de C determinados	88
4.5 Conclusiones	89
CAPÍTULO 5. ENFOQUE DIFUSO PARA LA EVALUACIÓN DE LA SEGMENTACIÓN DE IMÁGENES BASADO EN LA COMPLEJIDAD DE IMAGEN	91
5.1 Métricas de evaluación y variación del criterio humano	92
5.1.1 Métricas de evaluación	93
5.1.2 La RCI y la variación en el criterio humano al segmentar imágenes	94
5.2 Metodología del enfoque difuso propuesto	97
5.2.1 Fusificación de métricas de acuerdo con la complejidad de imagen	98
5.2.2 Agregación de las métricas difusas uM_H	102
5.3 Resultados	103
5.3.1 Evaluación de algoritmos del estado del arte mediante las diferentes métricas y enfoque propuesto	103
5.3.2 Evaluación y análisis por medio de meta-métricas	108
5.3.2.1 Meta-métrica Peng	108
5.3.2.2 Meta-métrica SIHD	109
5.4 Conclusiones	111
CAPÍTULO 6. RESULTADOS Y CONCLUSIONES GENERALES	113
REFERENCIAS	116

LISTA DE FIGURAS

Figura 1.1 Clasificación general de los métodos para determinar la respuesta a la complejidad de imágenes.	4
Figura 1.2 Algunas imágenes de la base de datos RSIVL.....	21
Figura 1.3 Algunas imágenes de la base de datos SAVOIAS ordenadas de menor a mayor complejidad de izquierda a derecha para cada categoría.....	22
Figura 1.4 Ejemplo de la interfaz.	23
Figura 1.5 Ejemplos de segmentación de imágenes [83].	25
Figura 2.1 Imagen de césped.	32
Figura 2.2 Gráfica de la media y desviación estándar de los valores de Ψ^H de la base de datos B_{RS1} ordenados ascendentemente, y los valores de Ψ^H generados por el sujeto 1 and 13 para cada imagen.	35
Figura 2.3 Coeficientes de correlación de cada sujeto de B_{RS1} y B_{RS2} respecto al valor promedio de complejidad propuesta por todos los sujetos.	36
Figura 2.4 z_{ij} de cada evaluación de los sujetos para todas las imágenes de B_{RS1} y B_{RS2}	37
Figura 2.5 Valores de complejidad de imagen para B_{RS1} y B_{RS2} en orden ascendente.	38
Figura 2.6 Muestra de imágenes para B_{RS1} y B_{RS2} con valores similares de RCI. El número indica el valor de complejidad asignado por los sujetos.	39
Figura 2.7 Imágenes con menor y mayor complejidad de B_{RS1} y B_{RS2}	39
Figura 3.1 a) Imagen en escala de grises con una resolución de 8 niveles de gris b) matriz de coocurrencia en escala de grises c) matriz de coocurrencia en escala de grises normalizada.....	42
Figura 3.2 Generación de conjuntos de entrenamiento y validación. Las imágenes fueron ordenadas ascendentemente por su complejidad.	51
Figura 3.3 Esquema general del sistema difuso propuesto.....	52
Figura 3.4 Comparación entre los sujetos aceptados de las bases de datos B_{RS1} y B_{RS2} contra el desempeño del sistema difuso propuesto en términos de correlación con el promedio de Ψ^H	55
Figura 3.5 Arquitectura de la red neuronal artificial para la determinación de la RCI.	56

Figura 3.6 Correlación de la complejidad determinada por el humano y las 100 redes neuronales con 1 a 100 neuronas en la capa oculta.....	58
Figura 3.7 Gráficas de regresión para el entrenamiento y la validación de la red neuronal de 28 neuronas que fue entrenada con la base de datos B_{RS1} y validada con B_{RS2}	58
Figura 3.8 Imágenes de muestra de $B_{SAVOIAS}$	63
Figura 4.1 Diagrama general de un algoritmo de agrupamiento difuso para segmentación de imágenes utilizando C determinado por AENCIC.	70
Figura 4.2 Cantidad de imágenes en la que se obtuvo el máximo PRI de acuerdo con el número de grupos c utilizado para los algoritmos FRFCM [88] y SFFCM [89].	72
Figura 4.3 Mapa de calor que representa la relación entre la complejidad de imagen y número de grupos en términos de la métrica PRI normalizada (PRIN) del conjunto de entrenamiento de BSDS500.	74
Figura 4.4 Histograma de complejidad de imagen del conjunto de entrenamiento de la base de datos BSDS500.	75
Figura 4.5 Cantidad de imágenes en la que se obtuvo el máximo PRI de acuerdo con el número de grupos utilizado en el algoritmo FRFCM.	76
Figura 4.6 Mapa de calor que representa cual valor de C fue seleccionado para cada imagen mediante AENCIC de acuerdo con los valores de Ψ_{Lim} , en términos del PRI normalizado para las imágenes de entrenamiento de BSDS500.	78
Figura 4.7 Diagrama de flujo que ilustra el proceso de segmentación de una imagen mediante AENCIC.	79
Figura 4.8 Comparación de resultados del experimento de la segmentación de imágenes reportada por AFCF en [91] y la reproducción de resultados. a) Segmentación reportada en [91] por $AFCF^{SLIC}$, b) Segmentación obtenida con $AFCF^{SLIC}$, c) Segmentación publicada en [91] por $AFCF^{LSC}$, d) Segmentación generada con $AFCF^{LSC}$, e) Segmentación documentada en [91] por $AFCF^{MMGR}$, f) Segmentación obtenida por $AFCF^{MMGR}$	82
Figura 4.9 Resultados obtenidos en términos de PRI para cada algoritmo de segmentación de imágenes utilizando diferentes métodos para determinar C	85

Figura 5.1 a) Imagen de baja complejidad con sus respectivos GT, b) Imagen de alta complejidad con sus respectivos GT.	95
Figura 5.2 Variación humana de acuerdo con las métricas M en función de la complejidad de imagen Ψ	96
Figura 5.3 Esquema general del método de evaluación difuso propuesto.....	98
Figura 5.4 Ajuste lineal de V_{PRI} en función de Ψ	99
Figura 5.5 Representación de la función $uPRI_H$, donde se especifica la localización de los valores VF_{PRIi} y a_{PRI}	100
Figura 5.6 Representación gráfica de las funciones uM_H	102
Figura 5.7 Ejemplo de evaluación de la segmentación de una imagen. a) Imagen original b) Segmentación correcta c) Segmentación incorrecta.	107
Figura 5.8 Gráficas de distribución de pares comparados por medio de M para el cálculo de SIHD.	110
Figura 5.9 Gráfica de distribución pares comparados por medio de $u^{\omega}SEG_H$ para el cálculo de SIHD.	111

LISTA DE TABLAS

Tabla 3.1 Coeficiente de correlación entre la RCI subjetiva de cada base de datos B y cada característica f .	46
Tabla 3.2 Tiempo en segundos para calcular cada característica f .	50
Tabla 3.3 Desempeño del sistema difuso expresado en términos de correlación entre Ψ^C y Ψ^H obtenidos mediante validación cruzada.	54
Tabla 3.4 Desempeño en términos del coeficiente de correlación obtenido por los diferentes experimentos en la base de datos B_{RS} , cambiando el conjunto de muestras utilizado para entrenamiento y validación.	60
Tabla 3.5 Desempeño de la red neuronal artificial usando validación cruzada en la base de datos B_{RS} .	61
Tabla 3.6 Desempeño de la RNA en términos del coeficiente de correlación utilizando validación cruzada en la base de datos $B_{SAVOIAS}$.	62
Tabla 3.7 Desempeño en validación en términos del coeficiente de correlación obtenido mediante el entrenamiento de la RNA con imágenes de una categoría de $B_{SAVOIAS}$ y validada con imágenes de otra categoría de $B_{SAVOIAS}$.	64
Tabla 3.8 Desempeño de los métodos al estimar la RCI.	65
Tabla 4.1 Desempeño de los algoritmos FRFCM y SFFCM con la base de datos BSDS500 utilizando $2 \leq C \leq 10$ para la métrica PRI.	73
Tabla 4.2 Comparación del número de grupos C reportado en [91] y el obtenido reproduciendo el experimento.	83
Tabla 4.3 Resultados obtenidos en los algoritmos difusos variando el número de grupos C .	84
Tabla 4.4 Porcentaje de eficiencia alcanzada en cada algoritmo de segmentación de acuerdo con la métrica PRI mediante los diferentes métodos para determinar C .	87
Tabla 4.5 Coeficientes de correlación entre C obtenido por cada método adaptativo y C correspondiente al máximo PRI de cada algoritmo de segmentación.	89
Tabla 5.1 Algoritmos de segmentación de imágenes evaluados por M y ordenados del mejor al peor desempeño en cada métrica.	104

Tabla 5.2 Algoritmos de segmentación de imágenes evaluados por uM_H , ordenados del mejor al peor desempeño.	105
Tabla 5.3 Algoritmos de segmentación de imágenes evaluados por $u^\omega \text{SEG}_H$, utilizando $\omega_M = 1$, ordenados del mejor al peor desempeño.....	106
Tabla 5.4 Resultados obtenidos mediante la meta-métrica propuesta por Peng <i>et al.</i> [129]...	109

GLOSARIO DE SÍMBOLOS

Símbolo	Definición	Capítulo donde se define
B	Base de datos de imágenes	1
I	Imagen digital	
GT	<i>Ground Truth</i>	
PRI	Índice rand probabilístico	
VI	Variación de la información	
GCE	Error de consistencia global	
BDE	Error de desplazamiento de fronteras	
S	Imagen segmentada	
v	Número de píxeles	
v, η	Índice de un grupo	
N	Número total de píxeles en una imagen	
E	Entropía	
Γ	Información mutua	
p	Pixel de una imagen digital	
$\hat{i}$	Índice de un pixel	
d	Distancia	
$i, \bar{i}$	Índice de una imagen	2
NI	Número total de imágenes en una base de datos	
Ψ	Respuesta a la complejidad de imagen	
Ψ^H	Respuesta a la complejidad de imagen obtenida por el criterio humano	
Ψ^C	Respuesta a la complejidad de imagen obtenida computacionalmente	
Ξ	Cantidad de información semántica en una imagen	
Θ	Relación dependiente del criterio humano para comprar Ξ	
SI	Conjunto de información semántica contenida en una imagen	

si	Unidad de información semántica	2
j	Índice de un sujeto	
z	z-score	
σ	Desviación estándar	
L	Conjunto de evaluaciones de complejidad aceptadas para una imagen	
f	Característica de una imagen	3
φ	Índice de característica de una imagen	
G	Matriz de coocurrencia en escala de grises normalizada	
g	Elemento de la matriz G	
l	Índice de filas de la matriz G	
m	Índice de columnas de la matriz G	
μ	Media	
$SRAD$	Frecuencia correspondiente a la energía de una imagen	
F_N	Frecuencia de Nyquist	
N_{BW}	Cantidad de píxeles detectados como borde en una imagen digital	
w	Ancho de una imagen en píxeles	
q	Alto de una imagen en píxeles	
MS	Conjunto de regiones encontradas por el algoritmo MeanShift	
r	Región de una imagen segmentada	
d_{ab}	Distancia del centro de gravedad al eje neutral en el espacio ab	
γ	Imagen en el espacio de color RGB reducido a 4 bits	
p	Pixel de una imagen digital en el espacio de color RGB reducido a 4 bits	
CS	Conjunto de píxeles indexados	
H_C	Efecto cromático	
H_L	Efecto de luminancia	
H_H	Efecto de matiz	

x	Característica normalizada	
ρ	Componente principal	
τ	Índice de componente principal	
t	Objetivo	
$\ddot{u}$	Función de membresía	
δ	Valor difuso	
R	Regla de un sistema difuso	
n	Índice de una regla de un sistema difuso	
A	Valor difuso de una variable lingüística	
k	Número de grupos utilizado por el algoritmo <i>k-means</i>	
y	Salida de una regla difusa	
ς	Centroide de función de membresía	3
NR	Número de reglas en un sistema difuso	
ω	Peso o ponderación	
$\ddot{A}$	Número de neuronas	
$\ddot{a}$	Índice de una neurona	
ω_h	Pesos de la capa oculta	
ω_o	Pesos de la capa de salida	
b_h	Polarización de la capa oculta	
b_o	Polarización de la capa de salida	
$\hat{o}$	Salida de las neuronas en la capa oculta	
Φ	Función de activación sigmoideal	
α	Argumento de función de activación sigmoideal	
C	Número de grupos	
c_f	Número de grupos fijo	
C_{maxPRI}	Número de grupos donde se obtiene el máximo PRI para cada imagen	4
$\grave{i}$	Índice de imagen de entrenamiento	
Ra	Rango	

Pa	Paso	4
H	Histograma	
$\hat{a}$	Índice del histograma por límites	
rd	Radio del disco de reconstrucción morfológica	
M	Conjunto de métricas de evaluación de segmentación de imágenes	5
V_M	Variación humana de cierta métrica M	
K	Número de GT en una imagen	
VF	Ajuste a variación humana mediante la ecuación de la recta	
ϖ	Índice de GT	
$\hat{e}, \ddot{e}$	Índice que representa a una persona	
α	Pendiente de una recta	
β	Sesgo de una recta	
R	Rango de variación humana	
u	Función de membresía	
a	Parámetro utilizado en las funciones de membresía tipo S y Z	

INTRODUCCIÓN

La determinación de la Respuesta a la Complejidad de Imagen (RCI), Ψ , consiste en el mapeo de la percepción de la complejidad de imagen en un valor que puede ser utilizado en diferentes aplicaciones. Esta tarea se basa en la teoría de estímulo-respuesta [1]. La RCI puede ser determinada mediante un enfoque humano, Ψ^H o un enfoque computacional, Ψ^C . Ya que la percepción de la complejidad de imagen es afectada por muchos factores como la edad, experiencia, aspectos psicológicos y demográficos [2]–[5], la RCI es un término que actualmente no tiene una definición estándar. Una de las posibles definiciones de la RCI se basa en la definición de complejidad visual propuesta por Snodgrass y Vanderwart [6], quienes fueron los primeros en darla a conocer. Ellos relacionan la complejidad visual con la cantidad de detalles o que tan intrincado es el entorno que se percibe mediante el sentido de la vista. Sin embargo, el término RCI continúa siendo subjetivo, ya que no existe una manera concreta para medirla. Lo anterior, llevó a algunos investigadores a trabajar en el diseño de algoritmos para medir la complejidad de las imágenes y comúnmente es definida de acuerdo con alguno de los métodos que se utiliza para medirla. Chikhman *et al.* [7] mencionan diversas definiciones para la complejidad de imagen, una de ellas se refiere a una medida de aleatoriedad que tiene algo en común con la entropía de Shannon de la imagen. Otra definición menciona que se puede expresar cómo el número de frecuencias espaciales o número de Wavelets activas. Además, mencionan que la complejidad se puede estimar como el producto de la media cuadrática de la distribución espacial de frecuencias y el área de la imagen. Para Chacón *et al.* [8] el término complejidad de imagen está relacionado al nivel de atención que requiere una persona para reconocer y detectar objetos en una imagen. En [9] se relaciona la complejidad de la imagen con la atención visual y el trabajo de la memoria que requiere una persona para entenderla, por lo que consideran que existe una relación entre la cantidad de objetos contenidos en una imagen y su complejidad. En [6] se define la complejidad de imagen como la cantidad de detalles y que tan confusa es una imagen. Por su parte, Heaps *et al.* [10] definen la complejidad de imagen como el grado de dificultad de dar una descripción verbal de la imagen [11]. Sin embargo, la definición de Snodgrass y Vanderwart [6] ha sido la más aceptada en el estado del arte pero, no

es una definición aceptable que detalle de manera adecuada el concepto de complejidad de imagen.

La complejidad de imagen ha sido utilizada en aplicaciones como: detección de bordes [12], determinación del nivel de compresión de una imagen [7], análisis de la complejidad visual de los tableros de carros y como afecta al conducir [13] y para analizar cuantas capas son necesarias en las redes neuronales convolucionales para reconocimiento de objetos [14], entre otras aplicaciones [15]–[21]. Sin embargo, sólo son aplicaciones de usuario final, es decir, para que el humano tome acciones en función de su valor, por lo que la complejidad de imagen es un campo abierto para futuras aportaciones.

Aunque ya existen algunos métodos para estimar la complejidad de imagen como los reportados en [15], [22]–[25], la mayoría no tiene una buena correlación con el criterio humano, trabaja solamente con imágenes artificiales o no son factibles para aplicaciones computacionales posteriores.

Contemplando la falta de información sobre la complejidad de imagen, este trabajo de tesis tiene el objetivo de estudiar y determinar la complejidad de imagen para su uso en los algoritmos de segmentación de imágenes, específicamente en la segmentación de imágenes por agrupamiento difuso. Por lo tanto, primero se da una definición conceptual y operacional de la respuesta a la complejidad de imagen. Luego, se analiza el criterio de los humanos al determinarla. Después se proponen dos métodos computacionales para determinar la respuesta a la complejidad de imagen. Por último, se proponen dos aplicaciones de la complejidad en algoritmos de segmentación.

Con la finalidad de describir el desarrollo de este trabajo, esta tesis se organiza de la siguiente manera:

El Capítulo 1 presenta el estado del arte para la determinación de la respuesta a la complejidad de imagen desde sus dos enfoques principales. Los enfoques presentados son: el enfoque computacional y el enfoque humano. También, presentan las bases de datos de

complejidad de imagen utilizadas en la literatura. Además, describe los algoritmos de agrupamiento difuso para la segmentación de imágenes.

El Capítulo 2 propone una nueva definición de la respuesta a la complejidad de imagen de acuerdo con el análisis del estado del arte y las bases de datos más populares. Además, presenta un preprocesamiento de una base de datos popular en la que la respuesta a la complejidad fue determinada mediante escalas psicométricas.

El Capítulo 3 presenta dos algoritmos para la determinación de la respuesta a la complejidad de imagen. Un algoritmo trabaja mediante un sistema de inferencia difuso y el otro mediante una red neuronal artificial. Ambos algoritmos son alimentados con 11 características visuales de las imágenes.

El Capítulo 4 describe un método para determinar automáticamente el número de grupos necesarios para segmentar una imagen mediante algoritmos de agrupamiento difuso. El número de grupos se determina mediante una relación directa con la complejidad de cada imagen. Este método es utilizado principalmente en imágenes de entornos reales.

El Capítulo 5 especifica un nuevo enfoque difuso para la evaluación de la segmentación de imágenes. Este enfoque evalúa cada imagen de diferente manera de acuerdo con la complejidad de imagen. La idea es evaluar la segmentación de una manera exigente cuando se trata de imágenes simples y de una manera relajada las segmentaciones de imágenes complejas.

Por último, el Capítulo 6 menciona los resultados y conclusiones generales de este trabajo de tesis.

CAPÍTULO 1. ANTECEDENTES

Este capítulo presenta los antecedentes para el desarrollo de este trabajo de tesis. Primero, la Sección 1.1 muestra los antecedentes referentes a la determinación de la Respuesta a la Complejidad de Imagen, RCI. Luego, la Sección 1.2 reporta los antecedentes sobre la segmentación de imágenes mediante algoritmos de agrupamiento difuso.

1.1 Respuesta a la complejidad de imagen

Mediante el análisis del estado del arte se encontraron diversos métodos para determinar la RCI, los cuales son agrupados en dos tipos de enfoques: computacionales y humanos, tal como se muestra en el diagrama de la Figura 1.1. Los enfoques computacionales son divididos en cinco categorías: información de bordes, información del color y/o intensidad, grado de compresión, redes neuronales convolucionales y combinado. Por otro lado, los enfoques humanos se dividen en dos categorías: criterio humano y reacción humana.

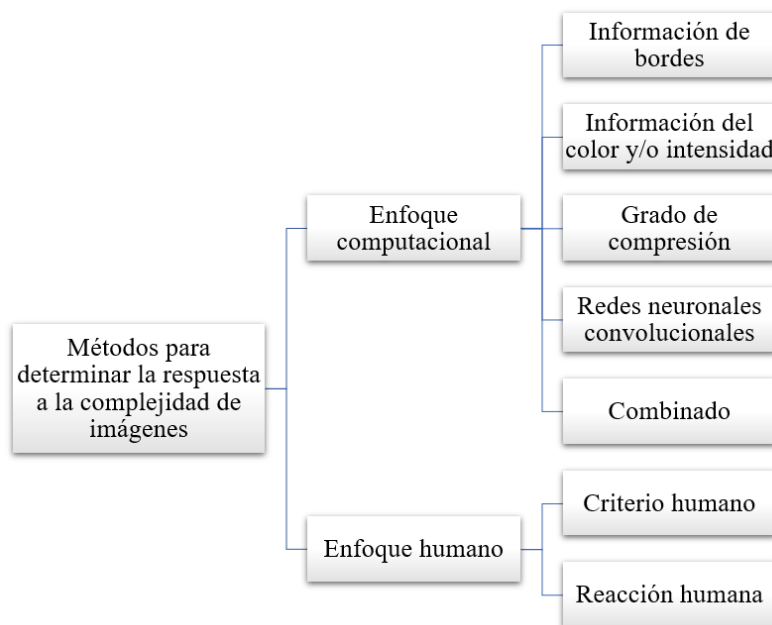


Figura 1.1 Clasificación general de los métodos para determinar la respuesta a la complejidad de imágenes.

De esta manera, en la Sección 1.1.1 se da un resumen de los métodos basados en enfoques computacionales encontrados y en la Sección 1.1.2 de los basados en enfoques humanos.

Seguidamente, en la Sección 1.1.3 se dan a conocer brevemente las bases de datos utilizadas para evaluar la RCI.

1.1.1 Enfoques computacionales

Los métodos para determinar la RCI por medio de enfoques computacionales son aquellos en los que el cálculo se realiza mediante algoritmos que utilizan recursos informáticos. A continuación, se describen los métodos encontrados basados en este tipo de técnica.

1.1.1.1 Información de bordes

Los bordes han sido utilizados para estimar la RCI debido a que se considera que estos están fuertemente relacionados, ya que una imagen con mucha información presente cuenta con una gran cantidad de objetos y detalles, y estos están relacionados con la cantidad de bordes. Diversos trabajos emplean los bordes como características, sin embargo, no es la única información que toman en cuenta, es por esto por lo que no todos quedan dentro de esta categoría. En total, se encontraron dos métodos donde se utilizan los bordes como característica principal para determinar la RCI, uno de ellos es propuesto por Chacón *et al.* [8], donde se basan en el análisis de los porcentajes de niveles de bordes contenidos en toda la imagen. La información que proveen los bordes es procesada por medio de lógica difusa y se calcula el grado de pertenencia de las imágenes a las clases: *Poco Compleja*, *Más o Menos Compleja* y *Muy Compleja*. Estas clases son determinadas mediante el algoritmo de agrupamiento difuso (*fuzzy C-means*) descrito en [26]. Los resultados de la medición de la RCI son consistentes y se correlacionan con el criterio humano. Este método para determinar la RCI fue probado en una base de datos con 150 imágenes. La ventaja principal es que no se involucra el criterio humano, lo que hace que sea muy consistente. Como aplicación del método, se encuentra un algoritmo detector de bordes, el cual utiliza la RCI para determinar cuáles niveles de bordes serán incluidos en la imagen de bordes. Machado *et al.* [27] miden la RCI mediante el uso de operaciones de detección de bordes, algunas métricas basadas en la compresión de imágenes y la ley Zipf. La información que proveen los bordes es extraída respecto a su cantidad y su distribución, ya que consideran que una imagen es más compleja si cuenta con una gran cantidad de bordes y su distribución es menos predecible. Para extraer los bordes se utilizan filtros de Canny y de Sobel.

Las métricas basadas en compresión son utilizadas ya que existe una relación entre la complejidad y la compresión. Debido a que una imagen poco compleja tiene información redundante y predecible. Por lo tanto, esta información puede ser representada con menos datos, es decir que puede ser comprimida; y una imagen con complejidad alta, la información no puede ser predicha tan fácilmente, por lo tanto, es más difícil comprimirla ya que no se quiere perder información importante. La ley de Zipf es utilizada para calcular el rango y tamaño de frecuencia. Para calcular el rango de frecuencia Zipf es necesario contar el número de veces que se presenta cierta intensidad en los píxeles y ordenarlos. El tamaño de frecuencia Zipf es la diferencia de intensidad que tiene cada píxel con sus vecinos. Es importante mencionar que este trabajo es clasificado en esta categoría de densidad de bordes ya que, aunque se utilizan otros métodos, con la densidad de bordes es con la que obtuvieron los mejores resultados.

1.1.1.2 Información del color y/o intensidad

Dentro de esta categoría se encuentran los métodos para determinar la RCI que utilizan características extraídas directamente de la información del color y/o intensidad de los píxeles. Uno de estos trabajos es el método propuesto por Ivanovici [28], el cual se basa solamente en los colores y la resolución para que pueda ser utilizado en aplicaciones de tiempo real. Para proponer el método que determina la RCI estudian el comportamiento de la entropía de Shannon y de esta manera tratan de que tenga un comportamiento parcialmente similar. Por lo que, el cálculo de la RCI es el número normalizado de colores distintivos en un soporte espacial definido. Además, se realizaron pruebas con algunas imágenes artificiales donde se varió el número de colores y la resolución, y se extendió su análisis a la base de datos VisTex [29]. Los resultados muestran que este sencillo enfoque para la medición de la RCI se correlaciona con algunas otras características como la entropía. Por otro lado, Perki y Hyvarinen [30] presentan un método para evaluar la RCI basado en el Análisis de Componentes Independientes (ICA). La finalidad de utilizar ICA es tomarlo como una medida de dispersión del color y la intensidad de los píxeles, es decir, se mide el grado de no semejanza a la distribución Gaussiana y se aproxima a una medición de la entropía. Este método muestra que la complejidad es poca si los componentes son escasos, o si los datos están concentrados en subespacios pequeños. Por medio del algoritmo *Support Vector Machine* (SVM) en [31] clasifican a las imágenes como *Poco*

Compleja, Medio Compleja y Muy compleja, un resultado similar al reportado en [8]. Las características globales que se utilizan como entrada del clasificador son: la medida de complejidad de color [32], la saturación y la simetría. Para encontrar las características locales primero se segmenta la imagen y se juntan las regiones con color similar y adyacencia [33]. Posteriormente, de cada región se extrae la saturación, el contraste y la intensidad del color, además, también se utiliza el número de segmentos creados, el área y la forma de los segmentos. Este método se utiliza para evaluar la complejidad visual de las pinturas artísticas y obtienen correlación con el criterio humano, como trabajo futuro se aplicará en diferentes tipos de imágenes. Otro método que se basa en las regiones de la imagen pero con información del histograma, es el propuesto en [34], en él, se relacionan la complejidad de la imagen como la entropía de la intensidad de su histograma. El proceso para calcular la RCI toma como entrada la imagen segmentada y el histograma dividido por regiones. De la imagen segmentada se obtiene el número de regiones que se formaron, tomando en cuenta que entre más regiones contenga la imagen más compleja será y el histograma provee información de intensidad de cada una de las regiones. Este método demuestra otra manera en la cual puede ser medida la RCI, siendo afectada directamente por la resolución de las imágenes, el contraste y la cantidad de detalles contenidos. Por otro lado, Cavalcante *et al.* [35] proponen un método para evaluar la RCI de los paisajes urbanos basándose en la estadística del contraste local y la frecuencia espacial. Las características que se extraen del contraste son la media y la desviación estándar. La frecuencia espacial es segmentada por medio de mapas de Kurtosis para obtener su valor y la oblicuidad. El objetivo de evaluar la RCI de estos paisajes es observar cómo afectan a los conductores, con el fin de evitar incidentes como el no ver a algún peatón. Los resultados demuestran que se tiene una alta correlación con la subjetividad del criterio humano y tiene buen desempeño también con paisajes urbanos nocturnos. Por su parte, Cardaci *et al.* [36] trabajan con un modelo matemático difuso de complejidad visual basado en la medida difusa de la entropía, para generar una medida objetiva de la complejidad de las imágenes. Para crear el modelo matemático de complejidad visual se utilizaron características locales y globales de las imágenes. Se eligieron dos características locales, una es la información de los bordes y la otra son las simetrías locales que son obtenidas por la transformada discreta de simetría. De esta manera, los píxeles tienen tres valores: la intensidad de los bordes, la intensidad de borde

umbralizada y binarizada de cero a uno, y la intensidad de simetría local. La medida de complejidad global es evaluada usando tres nuevas definiciones de entropía difusa las cuales son detalladas en [37]. Los experimentos que realizaron son comparados con el criterio humano, para ver si existe correlación entre la subjetividad del humano para percibir la complejidad de las imágenes y su algoritmo. El método obtuvo una buena correlación entre las dimensiones espaciales y temporales de la complejidad asemejándose al criterio humano, lo que puede llevar a la posibilidad de incluir el procesamiento de información humana en un estándar para medir la RCI. Por otro lado, Pham e Yan [38] presentan una estadística de regularidad como una medida de la RCI en términos de la entropía. El método que proponen consiste en aproximar la entropía de las imágenes en el marco de una cadena estacionaria de Markov, la cual es construida a partir de un gráfico ponderado que viene de la divergencia Kullback-Leibler de la imagen. Los resultados de este método demuestran que es un buen enfoque para cuantificar la RCI basado en la aproximación de la entropía y que puede ser utilizado para la evaluación de la calidad de imágenes.

1.1.1.3 Grado de compresión

Existen varios métodos para determinar la RCI que utilizan el grado de compresión, sin embargo, como no es la única característica que utilizan, no todos se encuentran dentro de esta categoría. Sólo se encuentra el trabajo de Forsythe *et al.* [39] ya que el grado de compresión es la característica principal de su método. Al igual que Guo *et al.* [31] este trabajo se enfoca en medir la RCI de pinturas artísticas pero para relacionarla con la estética y la belleza. Su método se basa en el grado de compresión GIF y el cálculo de la dimensión fractal. El grado de compresión GIF fue tomado en cuenta debido a que presentó una fuerte correlación con el criterio humano cuando se trata de pinturas artísticas. Los resultados demuestran que la combinación de la compresión GIF y la dimensión fractal dan una buena aproximación a la belleza de las pinturas en términos de la complejidad.

1.1.1.4 Redes neuronales convolucionales

Dentro de esta categoría se encuentran principalmente dos métodos que hacen uso de las redes neuronales convolucionales para determinar la RCI. El primer método es propuesto por

Saraee *et al.* en [15], quienes determinan la RCI mediante la exploración de la información extraída en las capas convolucionales de diferentes modelos profundos. Proponen una métrica de activación de energía que combina la activación de las capas convolucionales para cuantificar la complejidad de imagen. Los autores obtuvieron un buen desempeño correlacionándose con el criterio humano con un coeficiente de 0.71 utilizando las imágenes de la base de datos SAVOIAS [40]. Por su parte, Xiao *et al.* [22] también trabajaron en un método para determinar la RCI utilizando características extraídas mediante una red neuronal convolucional. Aunque su propósito no es determinar la RCI, muestran un poco de información sobre como lo hacen. Por lo que, para determinar la RCI, utilizan las características extraídas de la red neuronal en el método *Support Vector Ordinal Regression*. De esta manera, logran obtener la RCI de una manera correlacionada al criterio humano.

1.1.1.5 Combinado

Los métodos que se encuentran en esta categoría son aquellos que no se les pudo dar una clasificación específica ya que utilizan dos o más características principales. Los investigadores combinan las características ya que opinan que la complejidad de las imágenes depende de varios aspectos, pero se le da un peso diferente a cada una de ellas.

Por medio de una combinación lineal de características Corchs *et al.* [8] [27] crean un método para determinar la RCI basándose en características del dominio del espacio y la frecuencia, y las propiedades de color. Para la combinación lineal se utilizan varias características visuales simples y fueron seleccionadas de acuerdo con la literatura referente a complejidad de imágenes, estas características se listan a continuación:

- Contraste. El contraste de cada píxel con sus vecinos a lo largo de la imagen.
- Correlación. Cómo se correlaciona cada píxel con sus píxeles vecinos a lo largo de la imagen.
- Energía. Se refiere a la suma cuadrática de los elementos en GLCM (Matriz de Coocurrencia de Escala de Grises).

- Homogeneidad. Mide la cercanía de las distribuciones de los elementos en GLCM con respecto a la diagonal de GLCM.
- Factor de frecuencia. Es la tasa entre la frecuencia correspondiente al 99% de la energía de la imagen y la frecuencia de Nyquist.
- Densidad de los bordes. Es obtenida aplicando el detector de bordes de Canny en las imágenes a escala de grises.
- Tasa de compresión JPEG con un factor $Q = 100$ en comparación a la imagen sin comprimir.
- Número de regiones calculadas usando el algoritmo de segmentación descrito en [41].
- Coloración de la imagen. La coloración de la imagen consiste en una combinación lineal de la media y la desviación estándar de la nube de píxeles en el plano de color [42].
- Número de colores distintos que existen en la imagen.
- La armonía de los colores [43].
- Congestión de características. Se obtiene una sola métrica utilizando tres mapeos de desorden de las imágenes, representando la congestión de color, textura y orientación [44].
- Entropía de subbanda. Esta métrica está relacionada al número de bits requeridos por subbanda para codificar la imagen [44].

Otro esquema que se ha trabajado es el uso de redes neuronales artificiales, en [45] Chen *et al.* las utilizan para generar un algoritmo que describa la RCI cuantitativamente tomando en cuenta tres aspectos: la textura, información de los bordes y la región significativa. Las características de la textura son extraídas de la matriz de coocurrencia en escala de gris, las cuales corresponden a la energía, entropía, contraste, homogeneidad y correlación. De los bordes sólo se utiliza como característica la densidad de bordes que existen en la imagen, debido a que relacionan que entre mayor densidad de bordes tenga la imagen, más compleja será. Para la región significativa se utilizan los coeficientes extraídos mediante la Transformada Discreta del Coseno. La clasificación de las características se realiza con una red neuronal artificial tipo *feedforward* la cual provee un valor que indica la RCI de cada una de las imágenes, de esta manera cada imagen se clasifica dentro de tres categorías: *No Compleja*, *Complicada* y *Muy Compleja*. Una ventaja de este método es que puede calcular la RCI de una manera automática e independiente del humano. En esta categoría también se encuentra el método propuesto por

Chikhman *et al.* [7] donde desarrollaron un modelo para medir la RCI donde utilizan características espaciales, características espacio-frecuenciales, una combinación de las propiedades espaciales y la Transformada de Fourier, y el tamaño de la imagen codificada en formato JPEG. Se realiza un experimento psicofísico para comparar sus resultados donde algunas personas evalúan la complejidad de dos grupos de imágenes, un grupo son 15 jeroglíficos chinos con los que no están familiarizados y el otro grupo es de 24 imágenes con la silueta de objetos ya conocidos. La finalidad de escoger estos dos conjuntos de imágenes fue para analizar si se requiere medir de diferente manera la RCI dependiendo del tipo de imagen utilizada. Los resultados indican que existe correlación entre los métodos que proponen para medir la RCI y el criterio humano. Sin embargo, concluyen que su modelo de medición puede ser usado para estimar la RCI, pero para diferentes clases de imágenes se tienen que utilizar mediciones distintas. Por otro lado, Yaghmaee y Jamzad en [46] desarrollan un método basado en el concepto de región de interés (ROI). Las ROI son aquellas regiones de las imágenes que más atraen la atención humana. La idea principal de este método es encontrar las ROI en las imágenes dividiéndolas por bloques y posteriormente la complejidad de las imágenes se mide basándose en la intensidad, contraste, localización, bordes y textura. El objetivo principal es aplicar la RCI para estimar la capacidad de marca de agua en imágenes a escala de grises, donde sus resultados demuestran que el método por ROI es mejor que los métodos de Complejidad Compositiva de Imagen [34], Dimensión Fractal [47] y el método *Quad Tree* [48] para calcular la capacidad de marca de agua. En [49] proponen un método para determinar la RCI, no directamente en el análisis del contenido de las imágenes, sino a partir de la perspectiva de la complejidad a través de un filtro de atención. Utilizan mapas de atención, mapas sobresalientes y mapas de calor para el cálculo de la complejidad de las imágenes. También, muestran que los puntos de enfoque dinámicos generados por la atención no centralizada pueden generar una mejor estimación de la RCI. Como trabajo futuro proponen enfocarse en el estudio de la complejidad en secuencias de video.

El algoritmo de agrupamiento *k-means* combinado con el algoritmo matemático de transformación de características invariantes a la escala (SIFT por sus siglas en inglés) es utilizado por Huo [14] [35] para desarrollar un algoritmo que mide el desajuste de los objetos

que se desean observar de los objetos más sobresalientes de las imágenes. Este desajuste es utilizado para medir que tan compleja es una imagen, tratando de representar el grado de dificultad que tienen los humanos para poner atención a los objetos de interés cuando existen objetos de no interés sobresalientes. Otro método objetivo para determinar la RCI que combina características es el propuesto por Purchase [52], donde utilizan siete características: 1) la cantidad de colores distintos en la imagen, 2) la cantidad de colores después de un algoritmo que normaliza el color por regiones similares y adyacencia, 3) la cantidad de colores cuando un algoritmo de posterización es aplicado, 4) la cantidad de regiones después de segmentar la imagen, 5) el área de los bordes al utilizar el algoritmo de detección de Canny [53], 6) la desviación estándar de la intensidad de los píxeles con la imagen en escala de grises y 7) el grado de compresión JPEG, PNG y GIF. Los resultados de este método demostraron que es complicado asimilarse con el criterio humano mediante el uso de características simples, ya que la regresión no dio los resultados esperados.

En el trabajo de Guo *et al.* [38] [39] determinan cinco características que afectan la complejidad de la percepción visual de las texturas, estas características son: la regularidad, rugosidad, direccionalidad, densidad y comprensibilidad. Proponen un conjunto de métodos para medir cada una de estas características, la regularidad es estimada por una función de autocorrelación, la rugosidad es medida basándose en los cambios locales, la direccionalidad es indicada de acuerdo con las direcciones de los bordes, la densidad se calcula mediante la densidad de los bordes y la comprensibilidad es definida mediante el criterio humano. Por lo que, la definición de complejidad de las imágenes está en función de todas estas características basadas en un análisis de la textura. De esta manera, como trabajo futuro proponen crear un método para determinar la RCI que involucre estas características. Además, se realizaron experimentos para verificar si cada una de las características propuestas se correlaciona con el criterio humano. Reinecke *et al.* en [56] predicen la primera impresión de la estética de las páginas web por medio de la cuantificación de la complejidad visual. Para medir la RCI utilizan las siguientes características: el promedio del valor de los píxeles en el espacio de color HSV para matiz y saturación, la suma del valor promedio de saturación y su desviación estándar, la suma ponderada de la longitud trigonométrica de la desviación estándar, el número de hojas

obtenidas por el método *Quadtree*, el número de áreas de imágenes, número de grupos de texto, relación de área con texto y sin texto, la simetría a lo largo del eje horizontal y vertical, y el equilibrio. Los resultados demuestran que la complejidad visual tiene gran importancia para la impresión de los usuarios en páginas web, sin embargo, concluyen que las características utilizadas no son universales para evaluar complejidad visual, es decir, los usuarios de las páginas web perciben la complejidad de diferente manera de acuerdo con sus antecedentes demográficos. Otro método para determinar la RCI que utiliza diversas características es propuesto en [57], el cual es utilizado para analizar la relación de la complejidad con la estética y de esta manera poder predecir la belleza en las fotos. Las características utilizadas para medir la RCI son extraídas de la composición, la estadística y la distribución de los píxeles. La composición es calculada utilizando el método momentos ortogonales variantes descrito en [58] y las características que se extraen son los segmentos, el color, la nitidez y la información de color relativo. Como información estadística se utilizan los bordes para contar el número de objetos, la curvatura, y textura calculan la entropía y áreas. La distribución es representada como la combinación del histograma y sus diferencias con algunas plantillas de histograma predefinidas. De esta manera muestran que la RCI que calculan por las características propuestas tiene una relación similar con la expectativa de belleza de los seres humanos en miles de fotos.

De acuerdo con el estado del arte de la determinación de la RCI por medio de métodos computacionales, es posible darse cuenta de que la complejidad se ha tratado de determinar mediante la relación de las características visuales con la percepción visual del cerebro humano. Sin embargo, algunos autores concluyen que no puede ser posible determinar la RCI con una sola característica visual, por lo que se han utilizado métodos para determinarla mediante la combinación de características. Debido a que la complejidad es un término subjetivo, no es esperado que los algoritmos que la determinen se correlacionen perfectamente con el criterio humano, ya que incluso entre humanos existe una variación considerable a la percepción de la complejidad, es por esto que a continuación se muestran algunos de los enfoques humanos que se han desarrollado para determinar la RCI de una manera subjetiva.

1.1.2 Enfoques humanos

Los métodos para determinar la RCI por enfoques humanos son aquellos en que la medición de la complejidad es especificada mediante el criterio humano o mediante un estudio de la reacción humana al presentar imágenes con diferentes grados de complejidad a los sujetos de pruebas. De esta manera, en esta sección se resumen los métodos encontrados que se basan en este tipo de técnica.

1.1.2.1 Criterio humano

Dentro de los artículos revisados acerca de la complejidad de imágenes, se encontraron algunos experimentos para poder evaluar el criterio humano y tener una medida subjetiva contra la cual comparar los métodos que proponen o para estudios psicológicos. La mayoría de los métodos consiste en presentar imágenes con diferentes grados de complejidad a personas, para que de esta manera se puedan realizar diferentes análisis y evaluaciones.

Un método para determinar la RCI es propuesto por Da Silva *et al.* [49], donde consideran que la complejidad de imagen está relacionada a la dificultad que se tiene para describir una imagen. Miden el tiempo de observación promedio de un conjunto de 148 imágenes de dos bases de datos que contienen información de movimiento ocular para imágenes con diferentes grados de complejidad [43], [44]. Se les pidió a 12 sujetos que observaran cada imagen de la base de datos el tiempo suficiente para poder describirlas con una cantidad media de detalles, posteriormente el tiempo fue normalizado, promediado y ordenado para poder obtener la puntuación promedio de tiempo de observación. Después, se obtiene la diferencia de tiempo de cada sujeto hacia el tiempo promedio de los sujetos por imagen, y con esta información se construye una curva ROC (Característica Operativa del Receptor por sus siglas en inglés) para las 148 imágenes, considerando un clasificador binario que estima la RCI correctamente si el tiempo que tardó cada sujeto tiene una distancia menor a un umbral del tiempo promedio. De esta manera, la ROC representa el número total de predicciones correctas a cierto umbral, por lo tanto, la RCI es estimada por medio del área bajo la curva ROC, donde toman en cuenta la variación del tiempo de cada participante con la media. Un método diferente es propuesto por Cardaci *et al.* [36], donde participan dos grupos, uno de 50 sujetos y otro de 15 sujetos, para

evaluar la complejidad de seis pinturas artísticas en tres categorías: baja, media y alta. Los dos grupos pertenecen a diferentes universidades con diferentes carreras, esto se realizó con la finalidad de evitar algún sesgo cultural. Cada imagen se mostró durante 90 segundos, después de este tiempo una persona les realizaba las siguientes preguntas a los participantes:

- ¿Definirías la imagen cómo agradable o molesta?
- ¿Definirías la imagen cómo silenciosa o llena de actividad?
- ¿Cuántas personas había en la imagen?
- ¿Cuántas mujeres y hombres estaban en la imagen?
- ¿Definirías la imagen cómo violenta o alegre?
- ¿Hubo algún niño?
- ¿Qué estación del año era?
- ¿La escena es por la mañana o por la tarde?
- ¿Había animales en la escena?
- ¿Cuántos tonos de color distintos puedes recordar?
- ¿Ha percibido líneas curvas o rectas en la imagen?
- ¿Cuántas formas distintas había en la imagen?

Posteriormente, se evalúan las respuestas de cada uno de los sujetos para cada imagen, y con esto se mide la RCI basándose también en el tiempo que tardan para analizar cada una. Otro método se presenta en [25], donde realizan 4 experimentos para evaluar la RCI de las imágenes en una escala de 0 a 100, las imágenes son presentadas por un tiempo indefinido hasta que el sujeto envía su respuesta, la RCI es determinada por la media de la puntuación que se le da a cada imagen. En el experimento 1 participaron 26 sujetos, 15 mujeres y 11 hombres con edad promedio de 37 años, en un rango de 18 a 53 años, utilizaron 49 imágenes de la base de datos RSIVL [61]. Para el experimento 2 participaron 16 mujeres y 20 hombres con edad promedio de 33 años, en un rango de 18 a 53, para este experimento utilizaron 49 imágenes diferentes a las del experimento 1, las cuales se encuentran en la base de datos LIVE [62]–[64] y IVL [49] [50]. En el experimento 3 participaron 17 sujetos 5 mujeres y 12 hombres con edad promedio de 35 años en un rango de 23 a 51, en este experimento se utilizaron 54 imágenes de textura, el

conjunto TXT1 de la base de datos VisTex [67]. Por último, para el experimento 4 participaron 10 mujeres y 13 hombres con edad promedio de 28 años en un rango de 19 a 68, y para este experimento utilizaron 68 imágenes de la base de datos RawFoot [68], [69].

Por otro lado, Cavalcante *et al.* [35] proponen una puntuación subjetiva donde involucran a 40 sujetos de dos nacionalidades: 27 japoneses y 13 argelinos. En el experimento, los sujetos son colocados a una distancia aproximada de 80 cm del monitor donde se muestran las imágenes a evaluar. Los sujetos debían agrupar según su criterio las imágenes de los paisajes urbanos en tres grupos: *simples*, *normales* y *complejos*. Después, se les pidió que ordenaran las imágenes de menor a mayor complejidad para cada grupo. De esta manera, la RCI fue estimada por la media de la puntuación que dieron cada uno de los sujetos participantes. En [57] también proponen una puntuación subjetiva donde utilizan 80 fotografías (de animales, arquitectura, paisajes urbanos, flores, comida, retratos y naturaleza muerta) para que cinco sujetos, dos mujeres y tres hombres con edad entre 23 y 28 años evalúen su complejidad en una escala del uno al cinco, donde uno es una fotografía muy sencilla y cinco es una muy compleja. Para cada imagen se calcula la media y la desviación estándar de la complejidad prevista por los sujetos. Otro método basado en el criterio humano es el propuesto por Ramanarayanan [70], las pruebas piloto para entender como es la RCI consistieron en que 21 sujetos ordenaran por grado de complejidad 40 imágenes obtenidas de la web. Cabe mencionar que no se les dio una explicación de lo que es complejidad, sino que cada sujeto realizó el experimento según su definición de complejidad. Después de este experimento se eligieron solamente 21 imágenes para las pruebas, donde los sujetos comparaban las imágenes por pares, diciendo cual es más compleja. Finalmente, los resultados obtenidos fueron comparados contra algunos métodos para determinar la complejidad de imágenes por enfoques computacionales, como el conteo de polígonos de escena y el grado de compresión de imágenes JPEG, donde se dieron cuenta de que no existía gran correlación. Por otro lado, Palumbo *et al.* [71] estiman la RCI mediante dos experimentos subjetivos 1A y 2A. El experimento 1A consistió en evaluar imágenes de patrones de bloques a blanco y negro. Participaron 10 sujetos con edad entre 19 y 45 años, cinco hombres y cinco mujeres. Los patrones de bloques fueron presentados a los sujetos aleatoriamente y tuvieron que evaluar su RCI en una escala de 1 a 6, donde 1 es muy simple y 6 muy compleja,

en total se les presentaron 180 imágenes por lo que la evaluación duró aproximadamente 20 min. Los resultados de este experimento demostraron una correlación positiva con una medida de RCI computacional basada en el grado de compresión GIF [72]. El experimento 2A participaron 10 sujetos con edad entre 17 y 76 años, seis mujeres y seis hombres. Este experimento es similar al 1A con la diferencia en que se utilizaron bloques de distintos tamaños. Para el experimento 2A se obtuvo una alta correlación positiva con el grado de compresión GIF.

Además de proponer un método computacional, Purchase *et al.* [52] también proponen una medida subjetiva de la RCI donde recolectan la información por medio de un sitio web, esto con la finalidad de que lo puedan hacer desde el lugar donde se encuentren. La primera etapa del experimento consiste en pedirles a los sujetos que ordenen cuatro imágenes basándose en que tan complejas son para ellos, cada imagen es mostrada dos veces durante esta etapa, y cada sujeto realiza esta acción 30 veces, por lo tanto, se tiene un total de 60 imágenes distintas. En la segunda etapa se les muestran las 60 imágenes a los sujetos y se les pide que indiquen la RCI de cada una de acuerdo con la escala Likert de cinco puntos, la cual es una escala utilizada comúnmente en encuestas para evaluar opiniones y actitudes de una persona, en este caso es adaptada para que se pueda elegir la complejidad de la imagen del 1 al 5, donde 1 indica que es una imagen poco compleja y 5 una imagen muy compleja. Este experimento duró en promedio 13 minutos y lo realizaron 54 sujetos. Los resultados de este experimento demostraron que la evaluación subjetiva de la RCI de los humanos es consistente tanto individual como grupal.

1.1.2.2 Reacción humana

La RCI no sólo ha sido estudiada para lograr definirla o cuantificarla, algunos trabajos de investigación se enfocan en estudiar cómo es la reacción de los seres humanos al observar imágenes con diferentes grados de complejidad. Las reacciones estudiadas pueden ser desde el movimiento ocular hasta las señales eléctricas que se generan en el cerebro.

El análisis del movimiento ocular al observar imágenes con diferentes grados de complejidad ha sido estudiado por Bonev *et al.* [73], y su hipótesis es que la complejidad de las imágenes y la cantidad de tareas que se tienen que realizar para entender la imagen interactúan

entre sí. Para el experimento utilizan a ocho personas, a las cuales se le muestran imágenes a una resolución fija de 1400 x 931 pixeles mientras que se utiliza un dispositivo para registrar el movimiento ocular (SR Research Eyelink II) y otro para registrar el movimiento de la cabeza (VICON MX). La distancia de observación y las imágenes utilizadas son normalizadas para que la percepción de todas las imágenes siempre sea del mismo tamaño, ya que, aunque se trate de la misma imagen, pero con diferente tamaño, el movimiento ocular será menor o mayor. Debido a esto, se puede concluir que el movimiento ocular para encontrar objetos en una imagen está relacionado a la complejidad siempre y cuando se tenga una calibración en el experimento tanto en el tamaño de la imagen como en la distancia a la que se encuentra el observador. En [74] también analizan el movimiento ocular, pero con la finalidad de observar el comportamiento de los ojos al leer oraciones japonesas y los efectos en palabras largas o con complejidad visual alta. Para los experimentos participan diez estudiantes de la universidad de Teikyo en Japón, cuyo idioma natal es el japonés. El movimiento ocular fue monitoreado utilizando el dispositivo seguidor de ojos Dual Purkinje generación 5.5. Los participantes leen un total de 142 oraciones, de las cuales se les pide que las lean hasta que sean comprendidas totalmente. Los resultados demuestran que la complejidad de las oraciones influye significativamente en el tiempo de lectura. Por otro lado, Roth *et al.* [75] analizan el movimiento ocular para medir como es la experiencia de los usuarios en una página web ficticia con la finalidad de ver cómo afecta la localización de los objetos en tiendas en línea, páginas de noticias o de compañías. 40 sujetos participan en este experimento y el movimiento ocular es registrado mediante el dispositivo Tobii 1750 eye-tracker. Los resultados muestran que el colocar los objetos en posiciones esperadas de los usuarios y diseñar su apariencia de acuerdo con la orientación ayuda a mejorar la primera impresión de las páginas web y también la navegación en éstas. De esta manera, el movimiento ocular ha sido estudiado para saber cómo se relaciona con la complejidad de las imágenes, sin embargo, al ser la complejidad un concepto subjetivo, la relación del movimiento ocular y la complejidad puede ser afectada por diversos factores, tales como el tamaño de las imágenes y la familiaridad con ellas.

Huo en [9] además de proponer un método para evaluar la RCI de las imágenes, también realizó un experimento para medir si existe correlación entre su método y la actividad que tiene

el cerebro humano mediante señales electroencefalográficas (EEG) cuando se observan imágenes con diferentes grados de complejidad. El experimento se realiza con cuatro hombres y cuatro mujeres con una edad promedio de 23.8 años donde se le pidió recordar cuantos objetos observan en una imagen que les es presentada en un monitor, las cuales van variando de simples a complejas. El número de objetos a recordar es de 3 a 50 en cada imagen. Los resultados obtenidos demuestran que existe correlación entre el método para medir la RCI que utilizan [14] [35] y el resultado del experimento con señales electroencefalográficas. Cho *et al.* En [76] también realizaron un estudio electroencefalográfico, además de uno magnetoencefalográfico (MEG) para medir la actividad cerebral de las personas cuando se les presentan imágenes 3D estereoscópicas, y de esta manera poder analizar su respuesta en imágenes con profundidad excesiva, diferentes tipos de complejidad en las texturas, objetos en movimiento, entre otras. La hipótesis de este trabajo es que la actividad cerebral puede aumentar a medida que aumenta la complejidad de forma de la imagen estereoscópica. Para probar su hipótesis proponen un experimento que usa tres imágenes distintas, consiste en recolectar las señales EEG y MEG de 10 sujetos de pruebas sanos (seis hombres y cuatro mujeres con edad promedio de 24.4 años) y ver si existe correlación de la actividad cerebral con la complejidad de las imágenes. Después del análisis de los experimentos se presentaron varios resultados donde se comprueba que existe correlación entre la actividad cerebral del área occipital y el tipo de imágenes que se les fueron mostradas. Basándose también en un estudio magnetoencefalográfico, Hsu *et al.* [2] estudian los efectos de la complejidad visual con un experimento que consiste en analizar las respuestas M100 y M170 en las zonas occipital y temporal de personas inglesas y chinas al observar caracteres chinos. Los resultados demostraron que para los ingleses y los chinos la M100 tiende a incrementar cuando se presentan caracteres con alto número de trazos, pero la M170 en el hemisferio izquierdo tiende a incrementar sólo con los chinos cuando están leyendo los caracteres con alto número de trazos, y la M170 del hemisferio derecho incrementa cuando leen los caracteres menos complejos. De esta manera, es posible darse cuenta de que la complejidad de imágenes sigue siendo subjetiva aún y cuando se analiza la actividad cerebral, lo cual es causado por una gran cantidad de aspectos psicológicos y demográficos. Lo que sucede con este experimento puede ser un fenómeno similar al encontrado en [77] donde se menciona que la señal P300 obtenida por medio de electroencefalografía se genera mejor cuando los estímulos

visuales del deletreador de Donchin [78] se dan en las columnas que en las filas, es decir que el cerebro humano reacciona de diferentes maneras de acuerdo con las costumbres y características demográficas.

1.1.3 Bases de datos relacionadas a la complejidad de imágenes

Las bases de datos más populares en el estado del arte para estudios de la complejidad de imagen son RSIVL (*Real Scenes, Imaging and Vision Laboratory*) [61] y SAVOIAS (*Scenes, Advertisements, Visualization and infographics, Objects, Interior design, Art, and Suprematism*) [40]. La base de datos RSIVL fue generada por Corchs *et al.* en [61] específicamente para el análisis de complejidad de imágenes. RSIVL contiene dos conjuntos de imágenes de entornos reales conocidos como RS1 y RS2, los cuales son definidos como:

$$\begin{aligned} B_{RS1} &= \{I_{RS1} \mid I_{RS1} \text{ es una imagen del conjunto RS1}\} \\ B_{RS2} &= \{I_{RS2} \mid I_{RS2} \text{ es una imagen del conjunto RS2}\} \\ B_{RS} &= \{I_{RS} \mid I_{RS} \text{ es una imagen de } B_{RS1} \cup B_{RS2}\} \end{aligned} \tag{1.1}$$

La cardinalidad de B_{RS1} y B_{RS2} es de 49 imágenes, y $B_{RS1} \cap B_{RS2} = \emptyset$. Para determinar el grado de complejidad realizaron un experimento que consistió en que varias personas evaluaran cada una de las imágenes de los conjuntos de datos. En este experimento se indicó la RCI en una escala psicométrica de 0 a 100 para cada imagen, logrando de esta manera representar la complejidad mediante el promedio de la puntuación de las personas tras haber filtrado los datos. La RCI de B_{RS1} fue calculada tomando en cuenta 26 personas, 15 mujeres y 11 hombres con edad promedio de 37 años en un rango de 18 a 53 años. En la evaluación de B_{RS2} participaron 36 personas diferentes a los que participaron en B_{RS1} , para esta base de datos fueron 16 mujeres y 20 hombres con edad promedio de 33 años en el mismo rango de edad que B_{RS1} . En la Figura 1.2 se muestran algunas imágenes de los conjuntos de datos de B_{RS} .



Figura 1.2 Algunas imágenes de la base de datos RSIVL.

Por su parte, la base de datos SAVOIAS fue creada por Saraee *et al.* [40] con la finalidad de contar con una base de datos abierta al público, que sirva como *Ground Truth* para realizar análisis de complejidad de imágenes en diferentes categorías de imágenes. Esta base de datos es definida como:

$$\begin{aligned}
 B_{Sc} &= \{I_{Sc} \mid I_{Sc} \text{ es una imagen del conjunto Escenas}\} \\
 B_{Ad} &= \{I_{Ad} \mid I_{Ad} \text{ es una imagen del conjunto Anuncios}\} \\
 B_{Vi} &= \{I_{Vi} \mid I_{Vi} \text{ es una imagen del conjunto Visualizaciones}\} \\
 B_{Ob} &= \{I_{Ob} \mid I_{Ob} \text{ es una imagen del conjunto Objetos}\} \\
 B_{In} &= \{I_{In} \mid I_{In} \text{ es una imagen del conjunto Interiores}\} \\
 B_{Art} &= \{I_{Art} \mid I_{Art} \text{ es una imagen del conjunto Arte}\} \\
 B_{Sup} &= \{I_{Sup} \mid I_{Sup} \text{ es una imagen del conjunto Suprematismo}\} \\
 B_{SAVOIAS} &= \{I_{SAVOIAS} \mid I_{SAVOIAS} \text{ es una imagen de } B_{Sc} \cup B_{Ad} \cup B_{Vi} \cup B_{Ob} \cup B_{In} \cup B_{Art} \cup B_{Sup}\}
 \end{aligned} \tag{1.2}$$

La cardinalidad de B_{Sc} , B_{Ad} , B_{Vi} y B_{Ob} es de 200 imágenes, para B_{In} y B_{Sup} es de 100 imágenes y para B_{Art} es 420 imágenes. $B_{Sc} \cap B_{Ad} \cap B_{Vi} \cap B_{Ob} \cap B_{In} \cap B_{Sup} \cap B_{Art} = \emptyset$. Cada conjunto de la base de datos contiene imágenes de siete categorías diferentes (Escenas, Anuncios, Visualizaciones, Objetos, Interiores, Arte y Suprematismo) y la RCI determinada mediante el criterio humano para cada imagen. Algunos ejemplos de las imágenes de la base de datos SAVOIAS son mostradas en la Figura 1.3.

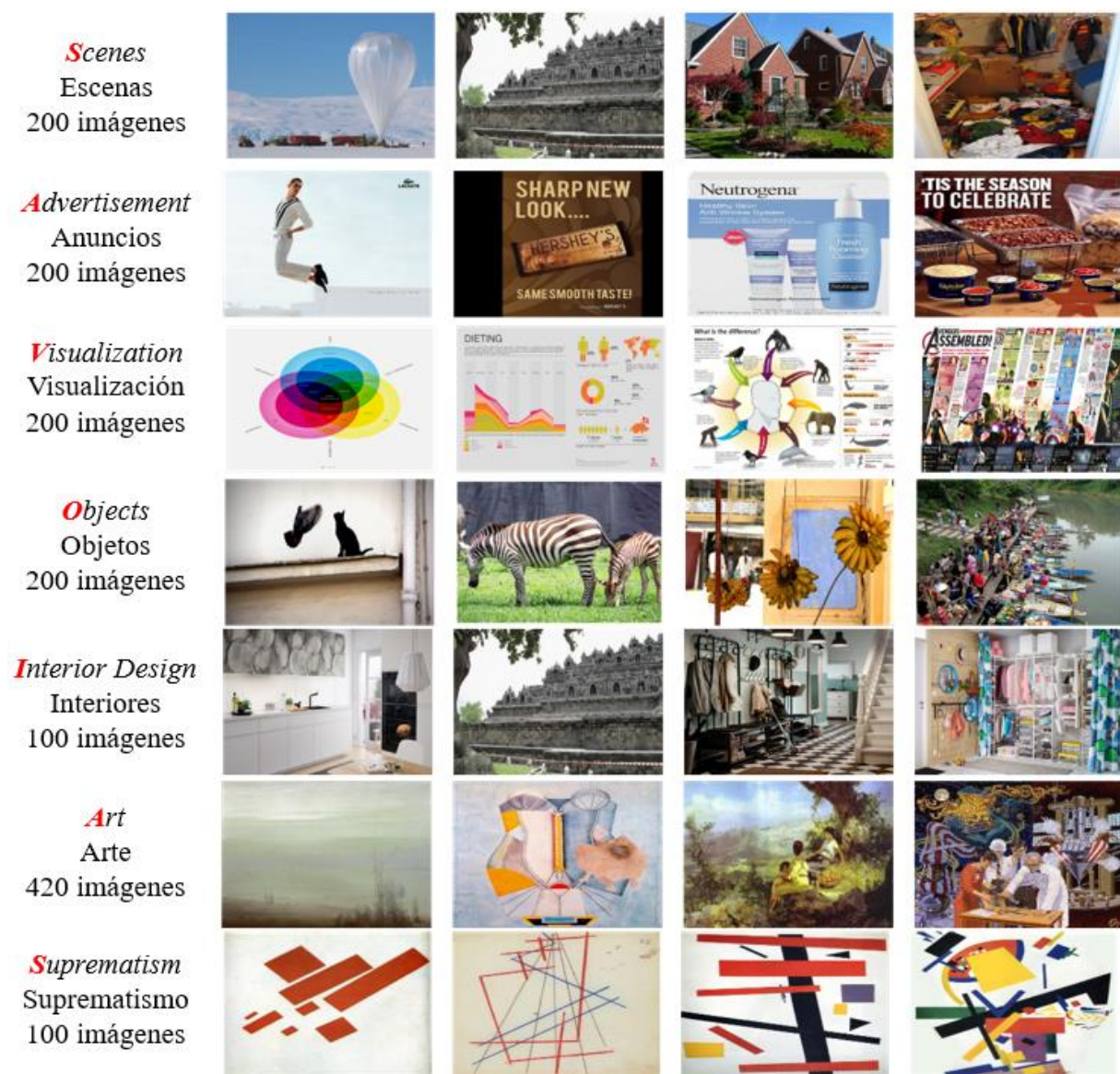


Figura 1.3 Algunas imágenes de la base de datos SAVOIAS ordenadas de menor a mayor complejidad de izquierda a derecha para cada categoría.

Las imágenes contenidas en cada categoría fueron elegidas con una distribución uniforme de complejidad de acuerdo con el criterio de los autores del artículo, bajo la definición de Snodgrass y Vanderwart [6] tomando la RCI como el nivel de detalles de las imágenes y que tan intrincadas son. Además de la base de datos con las imágenes, proponen una puntuación subjetiva para cada una de las imágenes. Dicha puntuación fue calculada mediante un

experimento donde participaron 1687 personas. Con la finalidad de eliminar uno de los problemas de la subjetividad del criterio humano, decidieron evitar el uso de escalas para evaluar cada imagen y el experimento se realizó por medio de una comparación por pares de imágenes. En la Figura 1.4 se muestra un ejemplo de la interfaz utilizada por las personas que colaboraron con la base de datos, en esta interfaz se tenía que elegir la respuesta que la persona considerara correcta, ya sea que la imagen A es visualmente más compleja que B o que B es visualmente más compleja que A.



Figura 1.4 Ejemplo de la interfaz.

No se compararon todos los pares de imágenes ya que esto repercutiría en que fuera una cantidad muy extensa de comparaciones y se tendría que tomar mucho tiempo para realizar la evaluación. Cabe mencionar que hubo un estímulo económico para cada una de las personas que contribuyó en la evaluación de las imágenes. Para obtener la puntuación de la RCI subjetiva de cada imagen basada en las comparaciones, primero, se calcula la tasa para cada comparación. Esta tasa es la relación que existe entre el número de veces que una imagen fue elegida como más compleja que otra imagen entre la cantidad de comparaciones que se realizaron. Posteriormente se utilizaron dos métodos, el método Bradley-Terry y el de completación de matriz. El método de Bradley-Terry describe la probabilidad de elegir una imagen como más compleja que otra imagen basándose en la comparación de éstas. La puntuación de complejidad para cada imagen puede ser estimada resolviendo el problema de Máximo a Posteriori (MAP) donde la distribución a priori se toma como uniforme. Por otro lado, el método de completación

de matriz tiene como base la inferencia de los pares que no se compararon basándose en los que sí. Es decir, se asume que, si la imagen 1 tiene mayor complejidad que la imagen 2, y la imagen 2 es más compleja que la imagen 3, entonces la imagen 1 es más compleja que la imagen 3. De esta manera, una vez que la matriz es completada, la puntuación de la RCI de cada imagen está dada por el promedio de la tasa de comparaciones para dicha imagen con todas las demás. Los resultados que dieron estos dos métodos se correlacionaron con un coeficiente de 0.98, es decir que se obtienen resultados fuertemente correlacionados, sin embargo, el método utilizado para mostrar los resultados fue el de Bradley-Terry.

1.2 Algoritmos de segmentación por agrupamiento difuso

Dentro de las aplicaciones de estimar la RCI se encontró que: es utilizada para la definición de los bordes de interés en una imagen [79]. Para obtener el grado de compresión de las imágenes [4] [10]. Como un ajuste automático de la tasa de transferencia de bits en transmisiones de videos [49]. Para definir el grado de marca de agua sin distorsionar las imágenes [46]. Para evaluar cómo afecta la complejidad de los paisajes urbanos al manejar [35]. Para medir el grado de dificultad al fijar la atención en los objetos de interés dependiendo el tipo de fondo utilizado en el diseño de páginas web [51]. Verificar como afecta la complejidad en la evaluación de la calidad de las imágenes [80]. Para el análisis de la complejidad visual de los tableros de automóviles y como afecta al conducir [13]. Para determinar el grado de compresión de redes neuronales convolucionales [20]. Para evaluar algoritmos de detección de fuego [21], entre otras aplicaciones.

Como se mencionó anteriormente este trabajo de tesis tiene la intención de determinar la RCI y darle una aplicación en los algoritmos de segmentación, específicamente en los algoritmos de segmentación por agrupamiento difuso. Por lo tanto, en esta sección, se define que es la segmentación de imágenes, se mencionan algunos algoritmos populares que son utilizados para la aplicación de la RCI y se describe la base de datos con la que se trabaja en esta tesis.

El objetivo de la segmentación de imágenes es dividir las imágenes en regiones, particiones o segmentos (grupos de píxeles) para localizar las partes u objetos de interés en una imagen digital [81]. Los segmentos creados están compuestos por píxeles adyacentes que cumplen con ciertos criterios de similitud, o cuando los píxeles adyacentes no se contemplan dentro de algún criterio de similitud se crean bordes que marcan las fronteras de cada segmento. El objetivo de la segmentación es identificar los segmentos que tienen algún significado y que pueden ser utilizados para describir el contenido de una imagen, tal como lo es identificar el fondo, los objetos o ciertas partes de los objetos [82]. Algunos ejemplos de segmentación de imágenes son mostrados en la Figura 1.5, donde se muestra la imagen original y la imagen segmentada.



Figura 1.5 Ejemplos de segmentación de imágenes [83].

La segmentación de imágenes por agrupamiento difuso ha contribuido en diversas áreas de investigación y ha sido utilizada en aplicaciones tales como: la conducción automática de automóviles, realidad aumentada, navegación en interiores, para sistemas de videovigilancia, algunas pruebas médicas como localización de tumores, sistemas de reconocimiento de patrones como el reconocimiento de huellas digitales o iris, motores de búsqueda en imágenes y análisis de imágenes satelitales [84]–[86]. Uno de los métodos más populares para llevar a cabo este tipo de segmentación es mediante el algoritmo *Fuzzy C-means* [87]. Este tipo de algoritmos sigue teniendo gran impacto en la comunidad científica debido a su eficiencia y buenos resultados. Algunos de los algoritmos recientes y con buenos resultados para este tipo de segmentación son el algoritmo FRFCM (*Fast and Robust Fuzzy C-Means*) [88], SFFCM (*Superpixel-Based Fast Fuzzy C-Means*) [89], RSSFCA (*Robust Self-Sparse Cuzzy Clustering Algorithm*) [90] y MMGR (*Multiscale Morphological Gradient Reconstruction*) [91]. Debido a

que esta tesis no se enfoca en proponer un nuevo algoritmo de segmentación de imágenes, estos no son descritos a detalle en este capítulo, por lo que serán descritos en la Sección 4.4.1.

1.2.1 Base de datos BSDS500

Para evaluar estos algoritmos de segmentación por agrupamiento difuso es comúnmente utilizada la base de datos “*Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500)*” [92]. BSDS500 cuenta con un total de 500 imágenes de entornos naturales divididas en 200 imágenes de entrenamiento, 100 imágenes de validación y 200 imágenes de prueba. Se eligió usar la presente base de datos ya que es una de las más populares en el estado del arte y además cuenta con un *benchmark* para realizar comparaciones.

Los autores de la base de datos proporcionan lo siguiente:

- Las 500 imágenes de entornos naturales que han sido segmentadas manualmente en promedio por cinco diferentes sujetos, las segmentaciones sirven como *Ground Truth* (GT) y como punto de referencia para realizar las evaluaciones en los algoritmos de segmentación y detección de fronteras.
- Programa para evaluar los resultados de los algoritmos con diferentes métricas.
- El algoritmo desarrollado por los autores de la base de datos para la segmentación y detección de contornos.
- Artículo con la documentación de la base de datos [92].

1.2.1.1 Documentación de la base de datos

En el artículo [92] que viene como documentación al descargar de la base de datos BSDS500 se menciona el lanzamiento de un nuevo conjunto de imágenes, haciendo referencia a la base de datos BSDS500, sin embargo, sólo se menciona que se agregaron 200 imágenes a la base de datos BSDS300 y no se encuentra información alguna de cómo fueron elegidas las imágenes ni cómo se realizaron los GT.

Por otro lado, cuando en el artículo [92] hacen referencia a la base de datos BSDS300 citan al artículo [93] donde si se encuentra información de cómo fue realizada la base de datos y las

consideraciones que se tomaron en cuenta. Cabe mencionar que todas las imágenes de la base de datos BSDS300 se encuentran en la base de datos BSDS500.

Para la creación de la base de datos BSDS300 en [93] se menciona que no hay una referencia absoluta para la segmentación de imágenes estáticas ya que depende del criterio de cada persona que observe la imagen y para hacer más confiable los resultados obtenidos deben ser comparadas contra múltiples segmentaciones humanas, cada una de una persona diferente.

Para el desarrollo de los GT a las personas se les dieron las siguientes instrucciones:

- Dividir cada imagen en segmentos, donde cada segmento representa un objeto distintivo en la imagen.
- Todos los objetos de la imagen tienen aproximadamente la misma importancia.
- El número de segmentos de cada imagen depende de la persona, sin embargo, se les dijo que la mayoría de las veces entre 2 y 20 segmentos son razonables.

Mencionan que las instrucciones no fueron muy estrictas con la finalidad de no restringir o sesgar criterio de las personas y poder tener un resultado que se acerque más al criterio humano, ya que dos personas pueden tener exactamente la misma percepción organizacional de una imagen, sin embargo, pueden diferir en la cantidad de segmentos y detalles que se tomarán en cuenta para la segmentación.

Los autores de [93] desarrollaron su propio software para la segmentación de las imágenes, el cual se escribió en lenguaje Java para que cualquier persona con internet pudiera utilizarlo. El software contenía herramientas que facilitaba la segmentación de las imágenes y se encargaba de que cada persona no segmentara la misma imagen dos veces.

1.2.1.2 Métricas de evaluación de la base de datos

En [92], en la sección 2.3 se menciona que las métricas utilizadas para evaluar la segmentación de imágenes son basadas en regiones, refiriéndose al *Probabilistic Rand Index* (PRI) [94], a la *Variation of Information* (VI) [95], el *Global Consistency Error* (GCE) [93] y el *Boundary Displacement Error* (BDE) [96].

El PRI [94] fue desarrollado con el fin de evaluar la precisión de los algoritmos de agrupamiento. Trabaja por medio de la comparación de los elementos pares en los grupos. El PRI entre la segmentación de una imagen S y el GT está dado mediante la siguiente ecuación:

$$\text{PRI}(S, \text{GT}) = 1 - \frac{2 \left(\sum_v \left(\sum_\eta v_{v\eta} \right)^2 / 2 + \sum_\eta \left(\sum_v v_{v\eta} \right)^2 / 2 - \sum_v \sum_\eta v_{v\eta}^2 \right)}{N(N-1)} \quad (1.3)$$

donde $v_{v\eta}$ es el número de pixeles en el v -ésimo grupo de S y el η -ésimo grupo del GT y N es el número total de pixeles en la imagen.

La métrica VI [95] tiene el propósito de comparar los agrupamientos generados por los algoritmos. Mide la distancia entre dos segmentaciones en términos de su promedio de entropía condicional y está dada por:

$$\text{VI}(S, \text{GT}) = E(S) + E(\text{GT}) - 2\Gamma(S, \text{GT}) \quad (1.4)$$

donde E representa la entropía y Γ representa la información mutua de la imagen segmentada S y el GT.

El GCE [93] mide el grado en que una segmentación puede verse como un refinamiento de la otra, el cual es representado por:

$$\text{GCE}(S, \text{GT}) = \min \left(1 - \frac{\sum_\eta \left(\sum_v v_{v\eta}^2 / \sum_\eta v_{v\eta} \right)}{N}, 1 - \frac{\sum_v \left(\sum_\eta v_{v\eta}^2 / \sum_v v_{v\eta} \right)}{N} \right) \quad (1.5)$$

El BDE [96] indica el desplazamiento promedio de los pixeles que son considerados como fronteras en dos segmentaciones y se mide mediante:

$$\text{BDE}(S, \text{GT}) = \left(\sum_i^{N_1} d(p_i, S) \right) / N_1 + \left(\sum_i^{N_2} d(p_i, \text{GT}) \right) / N_2 \Big/ 2 \quad (1.6)$$

donde N_1 y N_2 denotan el total de número de píxeles en las fronteras de la imagen S y el GT respectivamente. d es la distancia entre un píxel p_i en el GT y su frontera más cercana de un píxel p en S .

En este capítulo se abordaron los conceptos asociados a la respuesta a la complejidad de imagen, RCI y a los algoritmos de segmentación de imágenes por agrupamiento difuso. La RCI fue dividida en enfoques computacionales y enfoques humanos. Además, se presentaron las bases de datos relacionadas con la RCI. Respecto a la segmentación de imágenes, se mostraron algunos algoritmos populares, la base de datos más utilizada en el estado del arte y las métricas de evaluación utilizadas. De esta manera, en los próximos capítulos se presentará la contribución de este trabajo de tesis.

CAPÍTULO 2. COMPLEJIDAD DE IMAGEN Y ANÁLISIS DEL CRITERIO HUMANO

En este capítulo se presenta una definición conceptual y operacional de la respuesta a la complejidad de imagen, RCI, basada en el estudio del criterio humano al determinar la RCI en las dos bases de datos más populares del estado del arte: RSIVL [25] y SAVOIAS [40]. Además, presenta un análisis del criterio humano al determinar la RCI por medio de escalas en la base de datos RSIVL. Este análisis es utilizado para procesar la información de la base de datos y obtener una medición coherente de la RCI. Para esto, este capítulo se organiza de la siguiente manera: la Sección 2.1 presenta una definición conceptual de la RCI y la Sección 2.2 presenta la definición operacional de la RCI. Por su parte, la Sección 2.3 muestra un análisis de la subjetividad del criterio humano al determinar la RCI y la Sección 2.4 estudia la coherencia del criterio humano cuando determina la RCI.

2.1 Definición de la respuesta a la complejidad de imagen propuesta

Esta nueva definición es basada en el estudio del estado del arte sobre complejidad de imagen, el método utilizado por Saraee *et al.* para generar la base de datos SAVOIAS [40] y como la percepción de la complejidad de imagen depende de muchos factores [2]. Esta definición traslada el concepto de la percepción de la complejidad de imagen desde un espacio multidimensional al espacio unidimensional de la respuesta a la complejidad de imagen [1]. En esta tesis, la percepción de la complejidad de imagen está relacionada a los factores que una persona puede percibir en una imagen para juzgar su complejidad. Por su parte, la respuesta a la complejidad de imagen, RCI, es un valor que es obtenido mediante la comparación de esos factores con los factores de otras imágenes.

De esta manera, se sugiere relacionar a los factores involucrados en el juicio de la complejidad de imagen a la cantidad de información con significado (información semántica) que un observador puede percibir. Por lo tanto, la RCI se define como una medida relativa y subjetiva de la cantidad de información semántica que puede ser percibida mediante el sentido de la vista en una imagen, considerando un conjunto determinado de imágenes. El sustento de esta definición es presentado a continuación mediante la definición operacional.

2.2 Definición operacional de la respuesta a la complejidad de imagen

Sea B un conjunto de imágenes digitales I_i definido como:

$$B = \{I_i \mid I \text{ es una imagen digital y } 1 \leq i \leq NI\} \quad (2.1)$$

donde NI es el número de imágenes bajo análisis, la RCI determinada por un humano es definida como:

$$\Psi^H_i = \Theta\left(\Xi(I_i), \Xi\left(\{I_{\bar{i}}\}_{\bar{i}=1}^{NI}\right)\right) \quad \bar{i} \neq i \quad (2.2)$$

donde Ξ representa la cantidad de información semántica percibida mediante el sentido de la vista en una imagen, Θ es una relación dependiente del criterio humano para comparar Ξ de las imágenes. La cantidad de información semántica Ξ de una imagen es definida como:

$$\Xi(I) = |SI| \quad (2.3)$$

donde SI representa el conjunto de información semántica contenida en una imagen y $|\cdot|$ denota la cardinalidad del conjunto, por lo tanto:

$$SI = \{si \mid si \text{ es una unidad de información semántica}\} \quad (2.4)$$

donde:

$$si = \{p \mid p \text{ es un pixel de una región de una imagen digital que tiene significado para el humano}\} \quad (2.5)$$

Una de las definiciones de la RCI más utilizadas es la propuesta por Snodgrass y Vanderwart [6]. Ellos relacionan la RCI con la cantidad de detalles de las imágenes y que tan intrincado es el entorno que se percibe mediante el sentido de la vista. Sin embargo, esta definición no considera el significado de las regiones, ni la subjetividad que se presenta en el conjunto B . Por ejemplo, en una imagen de césped como se muestra en la Figura 2.1 se puede apreciar una gran cantidad de detalles por las hojas que se traslapan en todas direcciones. De acuerdo con la

definición de Snodgrass y Vanderwart, esta imagen sería considerada compleja sin importar el conjunto de imágenes B que se esté analizando. Para los humanos evaluadores de la RCI, la cantidad de información semántica cambiaría de acuerdo con el conjunto de imágenes que se desee evaluar. Por ejemplo, si B contiene imágenes de parques, considerarían que la imagen del césped es una imagen poco compleja ya que representa sólo una pequeña parte de la información semántica que puede contener una imagen de B . Por otro lado, si B contiene imágenes de texturas, posiblemente los humanos considerarían la imagen de césped como compleja, debido a la gran cantidad de hojas sin un patrón a seguir que presenta la imagen y cada hoja puede ser considerada como una unidad semántica. De acuerdo con lo anterior, esta tesis propone una definición de RCI más adecuada, donde se relaciona a la cantidad de información semántica contenida en la imagen, debido a que el ser humano tiende a determinar la RCI mediante el significado que poseen las regiones de las imágenes bajo cierto contexto.



Figura 2.1 Imagen de césped.

2.2.1 Propiedades de la RCI

- **Relatividad.** La RCI es relativa ya que compara la cantidad de información semántica de cada imagen. Esta cantidad de información depende del universo del discurso que se esté analizando. Por lo tanto, la relación Θ debe realizar el siguiente mapeo:

$$\Theta: \Psi^H_i \mid I \in B \rightarrow [0,1] \quad (2.6)$$

donde cada valor de Ψ^H_i es definido de acuerdo con Ξ de las imágenes, donde 0 representa la RCI de la imagen menos compleja del universo del discurso y 1 la RCI de la imagen más compleja.

- Subjetividad. La RCI es una medida subjetiva que depende de la percepción de cada persona al percibir la cantidad de información semántica de las imágenes. Sin embargo, esta subjetividad es reducida cuando se define correctamente el criterio a emplear para determinar la complejidad de las imágenes de cierto conjunto.

2.2.2 Resolución espacial y relación de aspecto

- Resolución espacial. La resolución espacial de las imágenes tiene un papel muy importante en la medición de la RCI y se deben tener en cuenta 2 consideraciones:
 1. La resolución espacial de toda $I \in B$ tiene que ser aproximadamente igual en cuestión del número de píxeles, preservando la relación de aspecto de la imagen original.
 2. La resolución espacial debe ser adecuada de tal manera que el humano sea capaz de distinguir visualmente toda la información semántica contenida en la imagen. Ya que grandes resoluciones en las imágenes que no puedan ser mostradas adecuadamente en los monitores, pueden causar una estimación errónea de la RCI por parte de los métodos computacionales, Ψ^C . Esto es debido a que los métodos computacionales pueden considerar información de las imágenes que los humanos no serían capaces de visualizar sin zum.
- Relación de aspecto. La relación de aspecto tiene un rol muy importante al determinar la RCI. Por lo que, no debe ser modificada de la imagen original. La variación en la relación de aspecto puede generar conflictos en el criterio humano al determinar la complejidad de imagen.

2.3 Análisis de la subjetividad del criterio humano

Una vez definida formalmente la RCI, es posible analizar la subjetividad del criterio humano. Este análisis es debido a que los datos utilizados en este trabajo de tesis provienen del criterio humano y uno de los pasos principales en el diseño de algoritmos o sistemas basados en

datos es asegurar que los datos utilizados son confiables, especialmente cuando no existe un reporte detallado de la manera en la que se obtuvieron. Por lo tanto, antes de proceder a diseñar un método para determinar la RCI, Ψ^C , es necesario verificar la confiabilidad de los datos. Este paso es importante especialmente cuando la determinación de la RCI mediante el criterio humano es muy subjetiva y se realiza por medio de escalas psicométricas, como es el caso de la base de datos B_{RS} . Aunque en esta tesis también se trabaja con la base de datos $B_{SAVOIAS}$, el proceso descrito en este capítulo no aplica a esta base de datos, ya que $B_{SAVOIAS}$ fue generada con un método más confiable que está relacionado a la definición de complejidad de imagen presentada en la Sección 2.2.

La percepción de la complejidad de imagen para una persona puede ser muy diferente a la de otra persona, sin embargo, debe de existir coherencia en un grupo de personas al determinar la RCI. Ya que la complejidad de imagen es aún un término subjetivo, se cree que no todas las personas que participaron en el desarrollo de B_{RS1} y B_{RS2} consideraron el mismo criterio que el presentado en la Sección 2.1 para determinar el valor de complejidad de cada imagen. Esto es porque la complejidad especificada por los humanos Ψ^H de esta base de datos fue especificada una escala psicométrica de 0 a 100 para cada imagen y las escalas pueden sesgar las evaluaciones cuando el criterio humano está involucrado [40]. De hecho, una persona puede especificar un valor de RCI diferente para una misma imagen, bajo las mismas circunstancias. Por lo tanto, es importante validar la información de esta base de datos antes de utilizar su información.

En el proceso de desarrollo de las B_{RS1} y B_{RS2} , los sujetos realizaron una prueba para ver si podían ser contemplados en el experimento. La prueba consistió en descartar sujetos que no reportaran correctamente 6 cartas Ishihara [97]. La prueba Ishihara es utilizada para asegurar que los sujetos no presenten alguna deficiencia para percibir los colores, sin embargo, asegurar que los sujetos no presenten un problema con los colores, no es suficiente para garantizar que ellos tienen el mismo concepto de RCI o que ellos entendieron correctamente el proceso para evaluar la RCI. Consecuentemente, el análisis propuesto en este capítulo para la base de datos B_{RS} ayuda a detectar y descartar evaluaciones fuera de lo común realizadas por los sujetos. El análisis considera los valores de Ψ^H de cada sujeto con respecto a los valores medios de Ψ^H del

grupo de personas en ambos conjuntos de datos. La Figura 2.2 muestra un gráfico del valor medio de Ψ^H ordenado de menor a mayor y la desviación estándar de B_{RS1} . La Figura 2.2 también presenta los valores de complejidad Ψ^H generados por los sujetos 1 y 13. El coeficiente de correlación de Pearson con el valor medio de Ψ^H con el sujeto 1 es de 0.91, mientras que para el sujeto 13 es de -0.09. Por lo tanto, el sujeto 1 es aceptado para el estudio y el sujeto 13 es rechazado. Realizando un análisis del valor de complejidad de todos los sujetos, un umbral de correlación de 0.43 fue determinado para aceptar a los sujetos. En este análisis, se pudo observar como el sujeto con un coeficiente de correlación de 0.43 realizó una evaluación aceptable, pero los sujetos con un menor coeficiente de correlación tuvieron muchos valores de complejidad inconsistentes con los valores de complejidad de la mayoría de los sujetos. Por lo tanto, 0.43 fue determinado como el umbral para aceptar o descartar las evaluaciones de los sujetos.

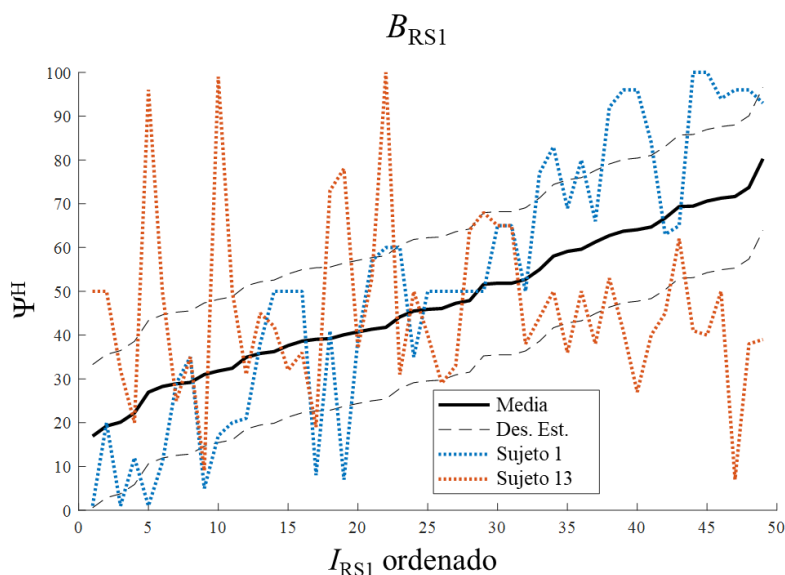


Figura 2.2 Gráfica de la media y desviación estándar de los valores de Ψ^H de la base de datos B_{RS1} ordenados ascendentemente, y los valores de Ψ^H generados por el sujeto 1 and 13 para cada imagen.

A continuación, la Figura 2.3 muestra el coeficiente de correlación de cada uno de los sujetos de las bases de datos B_{RS1} y B_{RS2} con respecto a la media de todos los sujetos. Se puede observar que 3 sujetos de B_{RS1} y 11 de B_{RS2} fueron descartados debido a que no superaron el umbral de

correlación de 0.43. Una vez realizado este análisis, el número de sujetos considerados para el estudio fue de 23 y 25 para B_{RS1} y B_{RS2} respectivamente.

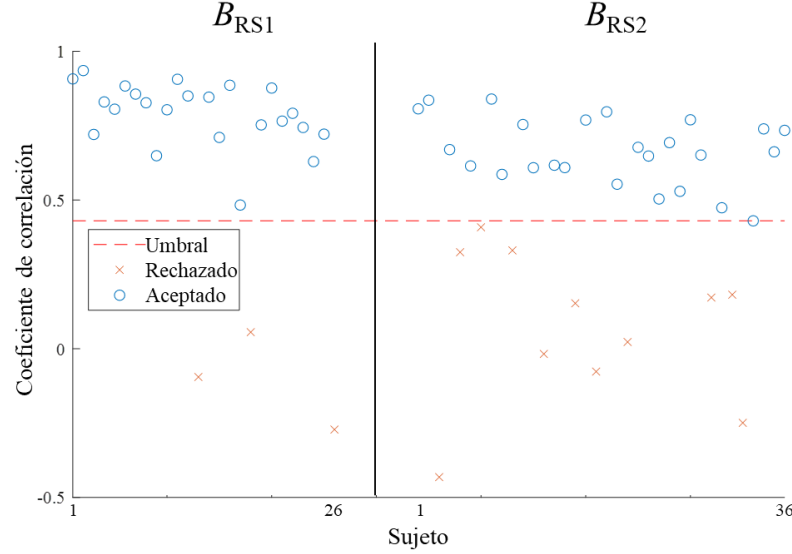


Figura 2.3 Coeficientes de correlación de cada sujeto de B_{RS1} y B_{RS2} respecto al valor promedio de complejidad propuesta por todos los sujetos.

Además del análisis de correlación por sujeto, que permitió descartar algunos sujetos de la base de datos, también se realizó un análisis de los valores de RCI, Ψ^H , asignados por cada sujeto aceptado. Para este análisis, algunos valores de Ψ^H fueron descartados porque fueron considerados incorrectos, debido a algún error por parte de los sujetos. El criterio para eliminar aquellos valores de complejidad es análogo al método descrito en [25]. Sin embargo, el cálculo de z-score fue por imagen, no por sujeto, como en [25]. Se utilizó este enfoque porque es esperado que un sujeto tenga variaciones cuando evalúa todas las imágenes, sin embargo, no se esperan muchas variaciones en las evaluaciones de una imagen por todos los sujetos. El z-score es obtenido para cada sujeto j en la imagen i de acuerdo con:

$$z_{ij} = \frac{\Psi_{ij}^H - \overline{\Psi_i^H}}{\sigma_i} \quad (2.7)$$

donde Ψ_{ij}^H es el valor de RCI subjetivo asignado por el sujeto j en la imagen i , $\overline{\Psi}_i^H$ y σ_i son la media y la desviación estándar de todos los valores subjetivos Ψ^H asignados por los sujetos para cada imagen i . Una vez que z_{ij} es calculado para cada valor de RCI, es utilizado para descartar aquellas evaluaciones identificadas como atípicas de acuerdo con el siguiente criterio. La Figura 2.4 se muestra una gráfica para B_{RS1} y B_{RS2} indicando todos los valores z_{ij} , los valores aceptados y rechazados también son indicados. La Figura 2.4 muestra que los valores rechazados son aquellos que se desvían más de dos desviaciones estándar. El umbral fue colocado en dos debido a que se puede observar cómo este umbral agrupa densamente las evaluaciones consistentes de las imágenes. Por lo tanto, las evaluaciones que desvían más de dos desviaciones estándar son rechazadas, ya que exhiben un comportamiento dudoso y pueden ser consideradas como atípicas.

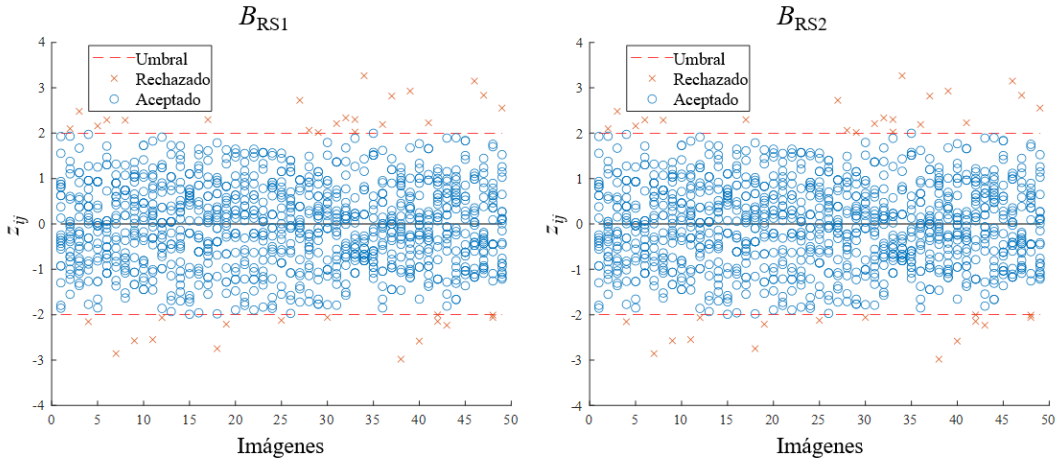


Figura 2.4 z_{ij} de cada evaluación de los sujetos para todas las imágenes de B_{RS1} y B_{RS2} .

Por último, el valor de RCI subjetivo de cada imagen Ψ_i^H es representado por la media de los valores subjetivos aceptados de dicha imagen, el cual se calcula mediante:

$$\Psi_i^H = \frac{1}{|L|} \sum_{j \in L} \Psi_{ij}^H \quad (2.8)$$

donde L es el conjunto de evaluaciones aceptadas de cada imagen y $|\cdot|$ indica cardinalidad.

2.4 Coherencia del criterio humano en B_{RS1} y B_{RS2}

Ya que dos grupos diferentes de personas evaluaron la RCI de B_{RS1} y B_{RS2} , se consideró necesario verificar si los dos grupos evaluaron ambos conjuntos de imágenes de una forma coherente. Los dos conjuntos contienen imágenes diferentes, pero las imágenes son del mismo tipo. Por lo tanto, el propósito de este experimento es verificar si estos dos conjuntos pueden ser unidos, es decir, si las evaluaciones de los dos son coherentes con la finalidad de tener más imágenes para trabajar en los algoritmos que determinan la respuesta a la complejidad de imagen, Ψ^C . El rango de los valores subjetivos Ψ^H determinado por los sujetos fue principalmente comparado entre los dos grupos. La Figura 2.5 tiene la intención de mostrar la posible duda en la coherencia de la evaluación de los conjuntos de imágenes B_{RS1} y B_{RS2} . Por lo que, la Figura 2.5 muestra los valores de la RCI de ambos conjuntos de imágenes de una manera ascendente. El rango de RCI de B_{RS1} va desde 10.72 hasta 83.95 y el de B_{RS2} de 24 hasta 72.

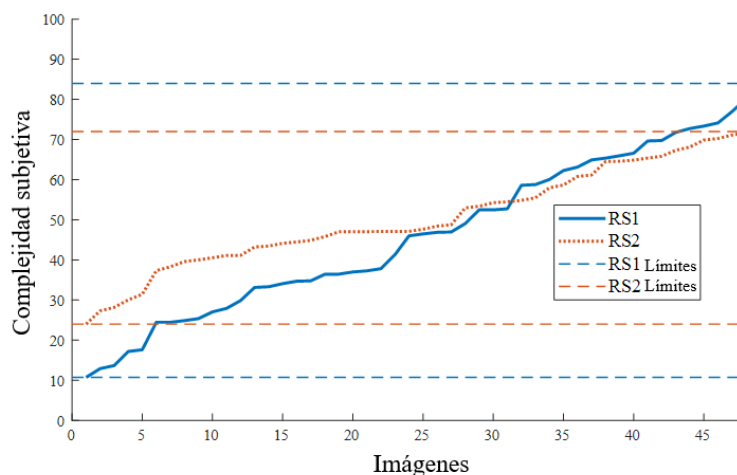


Figura 2.5 Valores de complejidad de imagen para B_{RS1} y B_{RS2} en orden ascendente.

De la gráfica de la Figura 2.5 se puede inferir lo siguiente: B_{RS2} contiene imágenes con un rango menor de complejidad que B_{RS1} o que el criterio de los humanos que participaron para crear B_{RS2} es menos sensible que los sujetos de B_{RS1} . Por lo tanto, se decidió comparar algunas imágenes de las dos bases de datos que tuvieran complejidades similares, tal como se muestra en la Figura 2.6.



Figura 2.6 Muestra de imágenes para B_{RS1} y B_{RS2} con valores similares de RCI. El número indica el valor de complejidad asignado por los sujetos.

Como fue mencionado anteriormente, la RCI es subjetiva y depende del criterio de cada persona, sin embargo, analizando las imágenes de la Figura 2.6 es posible notar que las imágenes en ambas bases de datos que tienen un valor similar de complejidad son percibidas aparentemente con un nivel de complejidad similar. También, para comparar el rango de complejidad, la Figura 2.7 muestra las imágenes con el menor y el mayor valor de complejidad de B_{RS1} y B_{RS2} . Se puede notar de la Figura 2.7 que en efecto la imagen con el menor valor de complejidad de B_{RS1} puede ser considerada menos compleja que la imagen con menor valor de complejidad de B_{RS2} y la imagen con el mayor valor de complejidad de B_{RS1} es más compleja que la imagen con el mayor valor de complejidad en B_{RS2} .

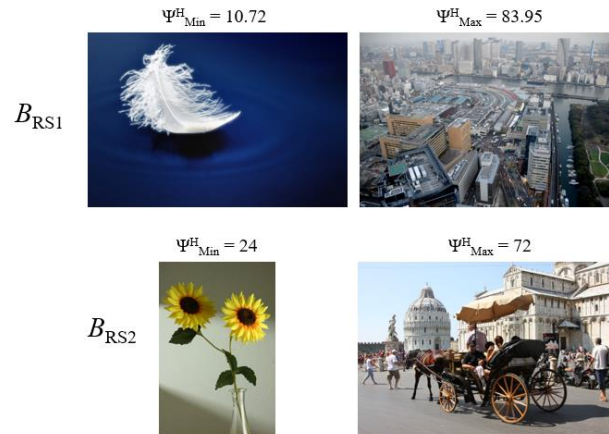


Figura 2.7 Imágenes con menor y mayor complejidad de B_{RS1} y B_{RS2} .

El análisis de las bases de datos B_{RS1} y B_{RS2} presentado en este capítulo consistió en descartar sujetos que realizaron evaluaciones atípicas de la RCI considerando la definición presentada en la Sección 2.1. Este análisis ayudó a obtener una mejor correlación de cada sujeto con la media de la RCI de acuerdo con el criterio de un grupo de personas. Además, se previene utilizar evaluaciones incorrectas que pueden generarse por errores durante las evaluaciones humanas. También, el análisis contribuyó en determinar que el criterio de los grupos de personas que realizaron los experimentos de B_{RS} es coherente una vez que las correcciones en los datos fueron realizadas.

El análisis y preprocesamiento de B_{RS} fue realizado debido a los problemas que se presentan cuando se utilizan escalas psicométricas para determinar la RCI, ya que estas escalas afectan a las evaluaciones. Ya que $B_{SAVOIAS}$ fue generada de una manera más robusta mediante la comparación de pares de imágenes, como fue mencionado en la Sección 1.1.3, esta etapa de preprocesamiento no fue necesaria.

CAPÍTULO 3. DETERMINACIÓN DE LA COMPLEJIDAD DE IMAGEN

La determinación de la respuesta a la complejidad de imagen, RCI, involucra diferentes procesos subjetivos. Por lo tanto, para determinarla computacionalmente es necesario utilizar métodos que simulen la manera en la que el cerebro humano procesa la información. De esta manera, este capítulo presenta dos métodos para determinar la RCI. El primer método es un sistema de inferencia difuso, ya que es eficaz para manejar problemas de vaguedad e incertidumbre [8], puede determinar la RCI de una manera muy similar a como lo hace un humano. El segundo método presentado está basado en una red neuronal artificial, ya que este tipo de algoritmos simulan de una manera eficiente la toma de decisiones del cerebro humano.

El presente capítulo se organiza de la siguiente manera. La Sección 3.1 presenta las características utilizadas por los dos métodos propuestos. En la Sección 3.2 se muestra el desarrollo del método basado en un sistema difuso. La Sección 3.3 contiene el desarrollo del método basado en una red neuronal artificial. La Sección 3.4 analiza y compara los resultados obtenidos contra otros métodos del estado del arte. Por último, la Sección 3.5 presenta las conclusiones obtenidas en este capítulo.

3.1 Características

Las características f_φ donde $1 \leq \varphi \leq 11$ utilizadas para determinar la RCI fueron elegidas de acuerdo con el estado del arte de la complejidad de imágenes y se mencionan a continuación:

Las características f_1 a f_4 son obtenidas de la Matriz de Coocurrencia en Escala de Grises Normalizada (GLCMN) la cual es definida como la matriz G que contiene elementos $g(l,m)$, donde l corresponde a las filas y m a las columnas [25]. Esta matriz es utilizada como un método estadístico para examinar la textura de las imágenes, considerando la relación espacial de los píxeles. La GLCMN caracteriza la textura de una imagen mediante el cálculo de la frecuencia con la que aparecen pares de píxeles con valores específicos y en una relación espacial específica de una imagen [98]. G es creada como se muestra en la Figura 3.1, donde en a) se muestra una imagen con una resolución de 8 niveles de gris. Luego, en la Figura 3.1 b) se muestra la matriz GLCM, que contiene el conteo de que tan frecuente un pixel con cierto nivel de gris l aparece

horizontalmente adyacente a un pixel con valor de gris m , donde cada elemento especifica el número de veces que un pixel con valor l ocurre horizontalmente adyacente a un pixel con valor m . Por último, en la Figura 3.1 c) se normaliza la GLCM de b) para obtener la GLCMN.

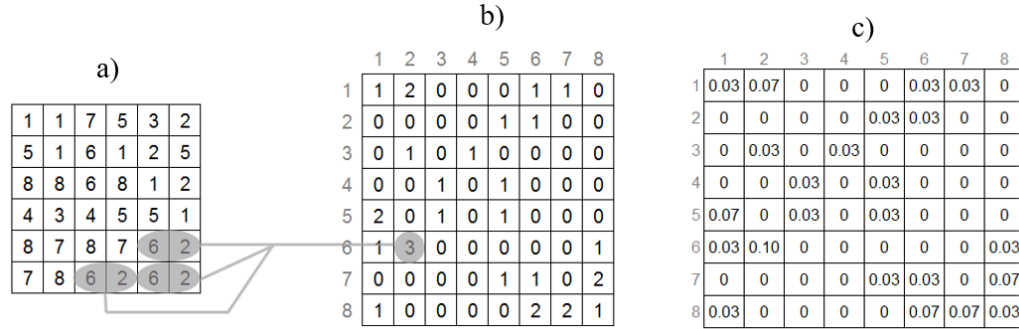


Figura 3.1 a) Imagen en escala de grises con una resolución de 8 niveles de gris b) matriz de coocurrencia en escala de grises c) matriz de coocurrencia en escala de grises normalizada.

- Contraste (f_1): es la medida de la intensidad del contraste entre cada elemento de G con su vecino en toda la matriz.

$$f_1 = \sum_{l,m} (l - m)^2 g(l, m) \quad (3.1)$$

- Correlación (f_2): mide que tan correlacionado está cada elemento de G con su vecino en toda la matriz.

$$f_2 = \sum_{l,m} \frac{(l - \mu_l)(m - \mu_m) g(l, m)}{\sigma_l \sigma_m} \quad (3.2)$$

donde μ_l y μ_m representa la media de la fila l y columna m y σ_l y σ_m la desviación estándar respectivamente.

- Energía (f_3): se refiere a la suma del cuadrado de los elementos de la matriz G .

$$f_3 = \sum_{l,m} g(l, m)^2 \quad (3.3)$$

- Homogeneidad (f_4): mide la proximidad de la distribución de elementos de G a su diagonal.

$$f_4 = \sum_{l,m} \frac{g(l,m)}{1 + \text{abs}(l - m)} \quad (3.4)$$

Las siguientes características f_5 a f_{11} son obtenidas de la información espacial contenida en las imágenes:

- Factor de frecuencia (f_5): esta métrica mide la relación entre las frecuencias correspondientes al 99% de la energía de la imagen y la frecuencia de Nyquist. Esta métrica es definida por la siguiente ecuación:

$$f_5 = \frac{SRAD_{0.99}}{F_N} \quad (3.5)$$

donde $SRAD_{0.99}$ es la frecuencia correspondiente al 99% de la energía y F_N la frecuencia de Nyquist [99].

- Densidad de bordes (f_6): mide el porcentaje de pixeles que son considerados como bordes en la imagen. Se obtiene como una relación entre la cantidad de pixeles detectados como bordes en la imagen en escala de grises aplicando el detector de bordes de Canny con un umbral bajo de 0.11 y umbral alto de 0.27 y la cantidad total de pixeles de la imagen [44]. Esta relación se define como:

$$f_6 = \frac{N_{BW}}{w \cdot q} \quad (3.6)$$

donde N_{BW} es la cantidad de pixeles detectados como borde, w es el ancho de la imagen y q la altura de la imagen en pixeles.

- Grado de compresión (f_7): caracteriza la cantidad de detalles de las imágenes mediante la optimización de los elementos redundantes y se representa como la relación entre la

imagen comprimida en formato JPEG y la imagen en tamaño original [65], el cual se define como:

$$f_7 = \frac{w_{JPEG} \cdot q_{JPEG}}{w \cdot q} \quad (3.7)$$

donde w_{JPEG} y q_{JPEG} son el ancho y alto de la imagen comprimida en formato JPEG respectivamente con un factor de calidad de 100.

- Número de regiones (f_8): es el número de regiones encontradas al utilizar el algoritmo de segmentación MeanShift descrito en [41]. El número de regiones es definido de la siguiente manera:

$$f_8 = |MS| \quad (3.8)$$

donde MS es un conjunto de regiones tal que $MS = \{r | r \text{ es una región encontrada por el algoritmo MeanShift}\}$.

- Coloración (f_9): esta métrica aproxima como los colores percibidos parecen ser más o menos cromáticos. Consiste en una combinación lineal de la media y la desviación estándar de la nube de pixeles en el espacio de color CIELab [42] y es definida como:

$$f_9 = \sqrt{\sigma_a^2 + \sigma_b^2} + 0.37d_{ab} \quad (3.9)$$

donde $\sqrt{\sigma_a^2 + \sigma_b^2}$ es la distancia euclidiana de la desviación estándar en el espacio ab y d_{ab} es la distancia del centro de gravedad al eje neutral en el espacio ab .

- Número de colores (f_{10}): es el número de colores contenidos en la imagen con el algoritmo descrito en [100]. Esta métrica se define de la siguiente manera: sea la imagen

$$I = \{p_{RGB} \mid p_{RGB} \text{ es un pixel en el espacio de color RGB con una resolución de 8 bits}\} \quad (3.10)$$

a cada componente del espacio RGB se le eliminan los 4 bits menos significativos definido por:

$$b_{\text{RGB}} = p_{\text{RGB}} \gg 4 \quad (3.11)$$

donde $\gg$ denota corrimiento a la derecha de bits. De esta manera se genera una nueva imagen

$$\gamma = \{b_{\text{RGB}} \mid b_{\text{RGB}} \text{ es un pixel en el espacio de color RGB reducido a 4 bits}\} \quad (3.12)$$

y los valores de los pixeles de la nueva imagen γ_{RS} obtenida son indexados $CS = \{\gamma\}_{b \in \{1, \dots, 512\}}$. De esta manera, el número de colores es definido como:

$$f_{10} = |CS| \quad (3.13)$$

- Armonía de colores (f_{11}): se refiere a la propiedad que tienen las combinaciones de colores de ser estéticamente agradables. Es definida como la sumatoria del efecto cromático (H_C), efecto de luminancia (H_L) y el efecto de matiz (H_H) [43].

$$f_{11} = H_C + H_L + H_H \quad (3.14)$$

3.1.1 Análisis de características

Las características f_1 a f_{11} son las más utilizadas en el estado del arte para determinar la RCI. Sin embargo, no todas ellas tienen el mismo nivel de correlación con el criterio humano al determinar la RCI. Por lo tanto, esta sección presenta un análisis de estas características. El análisis tiene el objetivo de determinar cuáles características son las más relevantes para determinar la RCI. Algunas de estas características son utilizadas por separado y directamente en algunos artículos del estado del arte, como una medida directa de la complejidad. La métrica comúnmente empleada para establecer una relación entre la RCI predicha por un humano y la RCI determinada computacionalmente es el coeficiente de correlación de Pearson entre cada característica f y el valor subjetivo de complejidad. De esta manera, la Tabla 3.1 muestra los coeficientes de correlación de Pearson obtenidos entre cada característica f y el criterio humano al determinar la RCI de cada una de las bases de datos B , también se muestra la media del valor

absoluto de los coeficientes de correlación. El valor absoluto es utilizado sólo para fines del análisis, ya que no importa si las características se correlación directa o indirectamente.

Tabla 3.1 Coeficiente de correlación entre la RCI subjetiva de cada base de datos B y cada característica f .

Dataset	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}
B_{RS}	0.607	-0.330	-0.612	-0.621	0.460	0.562	0.639	0.695	-0.160	0.518	-0.389
B_{Sc}	0.190	0.083	-0.349	-0.152	-0.062	0.142	0.281	0.575	0.182	0.636	-0.479
B_{Ad}	0.483	-0.169	-0.519	-0.571	0.128	0.617	0.681	0.663	0.310	0.420	-0.454
B_{Vi}	0.402	-0.015	-0.405	-0.487	0.240	0.571	0.572	0.377	0.231	0.462	-0.181
B_{Ob}	0.160	0.014	-0.270	-0.134	0.139	0.223	0.139	0.357	-0.016	0.411	-0.360
B_{In}	0.471	-0.408	-0.462	-0.500	0.384	0.639	0.705	0.691	0.412	0.706	-0.300
B_{Art}	0.231	-0.215	-0.521	-0.426	0.170	0.466	0.536	0.658	0.087	0.370	-0.359
B_{Sup}	0.127	-0.125	-0.606	-0.344	0.019	0.085	0.573	0.823	0.453	0.796	-0.337
Media del valor absoluto	0.334	0.170	0.468	0.404	0.200	0.413	0.516	0.605	0.231	0.540	0.357

A continuación, se da una breve explicación de cómo se correlaciona cada característica individualmente con la RCI.

Contraste f_1 . Ya que esta característica obtenida de la GLCMN mide la intensidad del contraste entre un pixel y sus vecinos en toda la imagen, también es conocida como una característica de varianza e inercia [98]. Para una imagen constante se tiene $f_1 = 0$. Por lo tanto, se correlaciona directamente con la RCI, como es mostrado en la Tabla 3.1, ya que conforme las imágenes dejan de ser constates, el valor de f_1 incrementa.

Correlación f_2 . La correlación medida en la GLCMN expresa una relación lineal y predecible entre los dos píxeles vecinos dentro de una ventana, expresada por la ecuación de regresión [98]. La correlación es 1 para una imagen perfectamente correlacionada positivamente o -1 para una imagen perfectamente correlacionada negativamente. Por lo tanto, a veces se correlaciona directa o inversamente con la RCI dependiendo de la categoría de imagen, así como es mostrado en la Tabla 3.1. Lo cual significa que el 0 representa una imagen de alta complejidad y 1 o -1 una imagen de baja complejidad.

Energía f_3 . La energía de la GLCMN también es conocida como una medida de uniformidad donde para una imagen constante $f_3 = 1$ [98]. Esto es debido a la normalización, ya que entre más dispersos sean los valores de la GLCMN menor será el valor de f_3 , porque se trata de una

sumatoria de todos los elementos elevados al cuadrado. Y cuando se trata de una imagen uniforme, la GLCMN sólo tendrá en una posición el valor de 1 mientras que en los demás valores será 0. Por lo tanto, f_3 se correlaciona inversamente con la RCI, así como es mostrado en la Tabla 3.1.

Homogeneidad f_4 . Ya que la homogeneidad mide la proximidad de la distribución de los elementos en la GLCMN a la diagonal, esta se correlaciona inversamente con la RCI. Ya que una imagen cuyos valores de GLCMN en la diagonal obtiene un $f_4 = 1$, lo cual significa que es una imagen poco compleja, debido a su uniformidad en los patrones de pixeles. Mientras tanto, en una imagen compleja los valores se encuentran dispersos en la GLCMN o lejanos a su diagonal.

Factor de frecuencia f_5 . El factor de frecuencia es una medida espectral basada en la transformada de Fourier. Comúnmente, es utilizada para describir la direccionalidad de los patrones periódicos o casi periódicos en las imágenes. Esta característica se correlaciona directamente con la RCI, ya que, entre más alto sea su valor hace referencia a una imagen de mayor complejidad. Esto es debido a que las imágenes que presentan un alto valor de f_5 es porque tienen muchos detalles que hacen parecer a las imágenes más complejas.

Densidad de bordes f_6 . En el caso de la densidad de bordes, se puede observar en la Tabla 3.1 cómo se correlaciona directamente con la RCI. Esto es debido a que normalmente entre más bordes tengan las imágenes mayor cantidad de información semántica contienen, por lo tanto, mayor complejidad. Por su parte, para una imagen constante, es decir, una imagen simple se tiene $f_6 = 0$.

Grado de compresión f_7 . El grado de compresión también es una métrica que se correlaciona directamente con la RCI, es decir, entre más alto sea el valor de esta métrica hace referencia a una imagen de mayor complejidad. Esto es debido a que entre menos complejas sean las imágenes más se pueden comprimir sin perder detalles importantes, por lo tanto, se espera un valor bajo de f_7 . En el caso de las imágenes muy complejas, no pueden ser tan comprimidas ya que se perdería información relevante de las imágenes, esperando un valor alto de f_7 .

Número de regiones f_8 . El número de regiones es la métrica con mayor grado de correlación a la RCI de una manera directa. Para este caso, entre más tenga la imagen se hace referencia a una imagen de alta complejidad. Mientras tanto, una imagen con pocas regiones es considerada poco compleja. Esto es debido a que la cantidad de regiones tiene una alta relación con la cantidad de información semántica contenida en las imágenes, ya que los algoritmos de segmentación son diseñados para segmentar las imágenes en regiones con significado para el humano.

Coloración f_9 . La coloración es una característica que comúnmente tiene una correlación directa con la RCI, de acuerdo con el tipo de imágenes en cuestión. Esto es debido a que normalmente se consideran más complejas las imágenes entre más coloridas son, mientras que si una imagen carece de coloración es posible que se considere una imagen poco compleja. Pero puede haber una excepción a esto dependiendo del tipo de imágenes, como lo es en el caso de los conjuntos de imágenes de B_{RS} y B_{Ob} donde se tiene una correlación inversa muy cercana a cero.

Número de colores f_{10} . Esta característica se correlación directamente con la RCI. Esto es debido a que el criterio humano normalmente considera más compleja una imagen que contiene un mayor número de colores a una imagen con una cantidad baja de colores.

Armonía de colores f_{11} . La armonía de colores se correlaciona inversamente con la RCI. Es decir, el criterio humano considera menos compleja una imagen cuyos colores están en mayor armonía y conforme los colores empiezan a tener menos armonía la complejidad aumenta. Esta relación es respaldada por el trabajo realizado en [57].

Por su parte, los valores remarcados en negritas de la Tabla 3.1 indican el valor máximo de correlación para cada conjunto de datos B mediante cierta característica f . Es posible darse cuenta de que $f_{7,8,10}$ se correlacionan muy bien con el criterio humano. Sin embargo, estas características no siempre presentan valores altos de correlación en todos los conjuntos de datos, por ejemplo, f_7 es bien correlacionada con B_{Ad} y B_{In} , pero no con B_{Ob} . Mientras que f_2 presenta un valor alto de correlación con B_{In} incluso cuando se correlaciona indirectamente.

Se ha mencionado que la definición propuesta de RCI implica una relación basada en la cantidad de información semántica percibida en las imágenes. Esta definición se relaciona con la segmentación de imágenes, cuyo objetivo es generar regiones en las imágenes con significado para el humano. Por lo tanto, la característica número de regiones f_8 tiene una alta correlación con la RCI. f_8 obtiene su mejor resultado en B_{Sup} , ya que son imágenes que no tienen muchas interpretaciones, debido a que están compuestas normalmente por figuras geométricas. Sin embargo, el desempeño con B_{Ob} es menor, esto es porque este tipo de imágenes tiende a tener múltiples interpretaciones semánticas a diferentes granularidades. Otras características como f_{10} contribuyen más en determinar la respuesta a la complejidad en el tipo de imágenes de B_{Ob} .

El número de regiones f_8 tiene una relación directa con la RCI. Sin embargo, existen muchas situaciones donde otras características son necesitadas para estimar la RCI de una mejor manera. Por ejemplo, reconsiderando el ejemplo de la imagen de césped en la Figura 2.1, es posible explicar como el uso de diversas características ayudan a determinar la RCI. Suponiendo que B contiene imágenes de parques, las características de densidad de bordes f_6 , y el número de colores f_{10} pueden trabajar juntas, donde la complejidad de la imagen de césped debe de ser baja porque sólo contiene una pequeña porción de lo que es una imagen de parques. La imagen del césped tiene alta densidad de bordes, pero tiene un número bajo en la característica número de colores, lo cual debería impactar en un valor de baja complejidad.

Ya que, en aplicaciones reales, el tiempo de procesamiento es importante, un análisis del tiempo necesario para extraer cada característica es presentado a continuación. La Tabla 3.2 muestra el tiempo de cómputo necesario para extraer cada característica. De acuerdo con esta tabla, es posible darse cuenta de que todas las características, excepto f_8 , tienen un tiempo de procesamiento relativamente bajo. f_8 tiene el tiempo computacional más alto y podría ser descartada para poder determinar la RCI más rápidamente. Sin embargo, de acuerdo con la Tabla 3.1 es la característica que más provee información para determinar adecuadamente la RCI y debido a que se desea priorizar la correlación con el criterio humano en vez del tiempo de procesamiento, f_8 no es descartada. Es importante mencionar que los tiempos mostrados en la Tabla 3.2 fueron obtenidos con una computadora con procesador Ryzen 5 3600 a 3.7 GHz, 16 GB de RAM a 2666 MHz y una GPU Nvidia GTX 1070ti.

Tabla 3.2 Tiempo en segundos para calcular cada característica f .

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}
Tiempo (s)	0.0265	0.0276	0.0273	0.0275	0.0175	0.0152	0.0216	1.7559	0.0076	0.0253	0.0183

De acuerdo con este análisis, se puede afirmar que no hay un resultado concluyente sobre una característica que pueda ser relaciona directamente con la RCI. Por lo tanto, se decidió utilizar el análisis de componentes principales para determinar la mejor información de estas características en el diseño de los algoritmos mostrados a continuación.

3.1.2 Preprocesamiento de características

La primera parte del preprocesamiento consiste en la normalización de las características para que su escala se encuentre en el mismo rango y así facilite el diseño de los algoritmos. La normalización es en la escala de 0 a 1, porque es el mismo rango utilizado por la RCI y se mantienen los datos en valores positivos con la misma tendencia, ya que la mayoría de las características f son positivas excepto f_{11} . La normalización es llevada a cabo mediante

$$x_{\varphi i} = \frac{f_{\varphi i} - \min(f_{\varphi i})}{\max(f_{\varphi i}) - \min(f_{\varphi i})} \quad (3.15)$$

donde $1 \leq \varphi \leq 11$ es el índice de cada característica y i indica el índice de cada imagen del conjunto de entrenamiento.

Una vez que las características son normalizadas, el método de análisis de componentes principales (PCA) es utilizado para representar las características en un nuevo espacio dimensional. Por lo que se encuentran características no correlacionadas que ayudan a reducir la dimensión del vector de entrada de los algoritmos. El uso de PCA permite que los algoritmos funcionen de una manera más eficiente [101]. Las primeras cuatro componentes principales para cada imagen son representadas por $\rho_{\tau i}$ para $1 \leq \tau \leq 4$. La misma matriz de transformación de PCA utilizada en este proceso es aplicada a las muestras de validación.

Finalmente, el valor de complejidad subjetivo Ψ^H_i obtenido de las bases de datos es utilizado como los objetivos t_i para el diseño de los algoritmos, pero en la escala de 0 a 1:

$$t_i = \frac{\Psi_i^H}{100} \quad (3.16)$$

3.2 Sistema de inferencia difuso para determinar la complejidad de imagen

Para el diseño y evaluación del algoritmo de inferencia difuso se utilizó la base de datos B_{RS} . Primero, se definieron los conjuntos de entrenamiento y validación. La manera en la que fueron elegidos fue pensada en hacer validación cruzada, la cual es útil para garantizar que las imágenes utilizadas no están sesgadas a dar cierto resultado y obtener un desempeño confiable. Debido a la cantidad de imágenes contenidas en B_{RS} , los conjuntos se forman de la siguiente manera para que las 98 imágenes se utilicen en la etapa de entrenamiento y validación al menos una vez:

1. Las imágenes son ordenadas de menor a mayor complejidad Ψ_i^H y se seleccionan de tal manera que se tomen imágenes de todos los rangos de complejidad, tanto para entrenamiento como validación.
2. Se forman 7 conjuntos de entrenamiento con 84 imágenes, dejando una imagen cada 7 a partir de la primera para validación, tal como se muestra en la Figura 3.2.

# Entrenamiento	# Validación																																																																																																																																																																																																					
<table><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr><tr><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td></tr><tr><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td></tr><tr><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td></tr><tr><td>29</td><td>30</td><td>31</td><td>32</td><td>33</td><td>34</td><td>35</td></tr><tr><td>36</td><td>37</td><td>38</td><td>39</td><td>40</td><td>41</td><td>42</td></tr><tr><td>43</td><td>44</td><td>45</td><td>46</td><td>47</td><td>48</td><td>49</td></tr><tr><td>50</td><td>51</td><td>52</td><td>53</td><td>54</td><td>55</td><td>56</td></tr><tr><td>57</td><td>58</td><td>59</td><td>60</td><td>61</td><td>62</td><td>63</td></tr><tr><td>64</td><td>65</td><td>66</td><td>67</td><td>68</td><td>69</td><td>70</td></tr><tr><td>71</td><td>72</td><td>73</td><td>74</td><td>75</td><td>76</td><td>77</td></tr><tr><td>78</td><td>79</td><td>80</td><td>81</td><td>82</td><td>83</td><td>84</td></tr><tr><td>85</td><td>86</td><td>87</td><td>88</td><td>89</td><td>90</td><td>91</td></tr><tr><td>92</td><td>93</td><td>94</td><td>95</td><td>96</td><td>97</td><td>98</td></tr></table> Conjunto 1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	<table><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr><tr><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td></tr><tr><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td></tr><tr><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td></tr><tr><td>29</td><td>30</td><td>31</td><td>32</td><td>33</td><td>34</td><td>35</td></tr><tr><td>36</td><td>37</td><td>38</td><td>39</td><td>40</td><td>41</td><td>42</td></tr><tr><td>43</td><td>44</td><td>45</td><td>46</td><td>47</td><td>48</td><td>49</td></tr><tr><td>50</td><td>51</td><td>52</td><td>53</td><td>54</td><td>55</td><td>56</td></tr><tr><td>57</td><td>58</td><td>59</td><td>60</td><td>61</td><td>62</td><td>63</td></tr><tr><td>64</td><td>65</td><td>66</td><td>67</td><td>68</td><td>69</td><td>70</td></tr><tr><td>71</td><td>72</td><td>73</td><td>74</td><td>75</td><td>76</td><td>77</td></tr><tr><td>78</td><td>79</td><td>80</td><td>81</td><td>82</td><td>83</td><td>84</td></tr><tr><td>85</td><td>86</td><td>87</td><td>88</td><td>89</td><td>90</td><td>91</td></tr><tr><td>92</td><td>93</td><td>94</td><td>95</td><td>96</td><td>97</td><td>98</td></tr></table> Conjunto 2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	...
1	2	3	4	5	6	7																																																																																																																																																																																																
8	9	10	11	12	13	14																																																																																																																																																																																																
15	16	17	18	19	20	21																																																																																																																																																																																																
22	23	24	25	26	27	28																																																																																																																																																																																																
29	30	31	32	33	34	35																																																																																																																																																																																																
36	37	38	39	40	41	42																																																																																																																																																																																																
43	44	45	46	47	48	49																																																																																																																																																																																																
50	51	52	53	54	55	56																																																																																																																																																																																																
57	58	59	60	61	62	63																																																																																																																																																																																																
64	65	66	67	68	69	70																																																																																																																																																																																																
71	72	73	74	75	76	77																																																																																																																																																																																																
78	79	80	81	82	83	84																																																																																																																																																																																																
85	86	87	88	89	90	91																																																																																																																																																																																																
92	93	94	95	96	97	98																																																																																																																																																																																																
1	2	3	4	5	6	7																																																																																																																																																																																																
8	9	10	11	12	13	14																																																																																																																																																																																																
15	16	17	18	19	20	21																																																																																																																																																																																																
22	23	24	25	26	27	28																																																																																																																																																																																																
29	30	31	32	33	34	35																																																																																																																																																																																																
36	37	38	39	40	41	42																																																																																																																																																																																																
43	44	45	46	47	48	49																																																																																																																																																																																																
50	51	52	53	54	55	56																																																																																																																																																																																																
57	58	59	60	61	62	63																																																																																																																																																																																																
64	65	66	67	68	69	70																																																																																																																																																																																																
71	72	73	74	75	76	77																																																																																																																																																																																																
78	79	80	81	82	83	84																																																																																																																																																																																																
85	86	87	88	89	90	91																																																																																																																																																																																																
92	93	94	95	96	97	98																																																																																																																																																																																																
	<table><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr><tr><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td></tr><tr><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td></tr><tr><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td></tr><tr><td>29</td><td>30</td><td>31</td><td>32</td><td>33</td><td>34</td><td>35</td></tr><tr><td>36</td><td>37</td><td>38</td><td>39</td><td>40</td><td>41</td><td>42</td></tr><tr><td>43</td><td>44</td><td>45</td><td>46</td><td>47</td><td>48</td><td>49</td></tr><tr><td>50</td><td>51</td><td>52</td><td>53</td><td>54</td><td>55</td><td>56</td></tr><tr><td>57</td><td>58</td><td>59</td><td>60</td><td>61</td><td>62</td><td>63</td></tr><tr><td>64</td><td>65</td><td>66</td><td>67</td><td>68</td><td>69</td><td>70</td></tr><tr><td>71</td><td>72</td><td>73</td><td>74</td><td>75</td><td>76</td><td>77</td></tr><tr><td>78</td><td>79</td><td>80</td><td>81</td><td>82</td><td>83</td><td>84</td></tr><tr><td>85</td><td>86</td><td>87</td><td>88</td><td>89</td><td>90</td><td>91</td></tr><tr><td>92</td><td>93</td><td>94</td><td>95</td><td>96</td><td>97</td><td>98</td></tr></table> Conjunto 7	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98																																																																																																			
1	2	3	4	5	6	7																																																																																																																																																																																																
8	9	10	11	12	13	14																																																																																																																																																																																																
15	16	17	18	19	20	21																																																																																																																																																																																																
22	23	24	25	26	27	28																																																																																																																																																																																																
29	30	31	32	33	34	35																																																																																																																																																																																																
36	37	38	39	40	41	42																																																																																																																																																																																																
43	44	45	46	47	48	49																																																																																																																																																																																																
50	51	52	53	54	55	56																																																																																																																																																																																																
57	58	59	60	61	62	63																																																																																																																																																																																																
64	65	66	67	68	69	70																																																																																																																																																																																																
71	72	73	74	75	76	77																																																																																																																																																																																																
78	79	80	81	82	83	84																																																																																																																																																																																																
85	86	87	88	89	90	91																																																																																																																																																																																																
92	93	94	95	96	97	98																																																																																																																																																																																																

Figura 3.2 Generación de conjuntos de entrenamiento y validación. Las imágenes fueron ordenadas ascendentemente por su complejidad.

3.2.1 Diseño del sistema difuso

El sistema difuso propuesto es un Sugeno de orden cero. Cuenta con cuatro variables lingüísticas de entrada las cuales se refieren a las cuatro primeras componentes principales $\rho_{\tau i}$. Estas variables pueden tomar los siguientes valores difusos: ρ_{1i} y ρ_{2i} toman los valores de *Level 1*, *Level 2*, *Level 3*, *Level 4*, *Level 5* y ρ_{3i} y ρ_{4i} de *Level 1*, *Level 2*, *Level 3*, *Level 4*. El propósito de esta diferencia en el número de valores difusos es darle más resolución a las componentes que son más discriminativas para determinar la complejidad y evitar caer en un exceso de valores difusos que incrementen las reglas innecesariamente. La Figura 3.3 muestra un esquema general del sistema difuso propuesto para determinar la RCI.

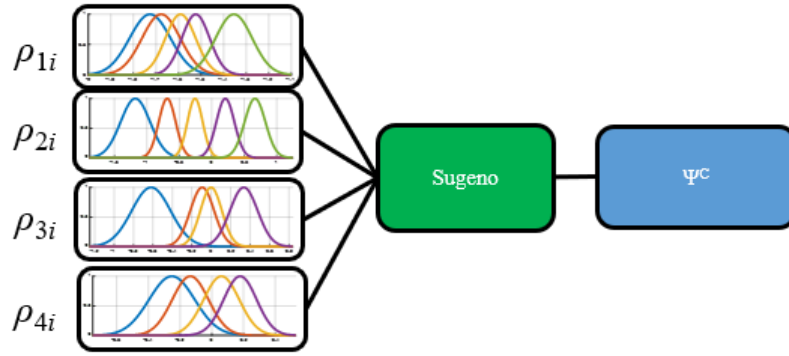


Figura 3.3 Esquema general del sistema difuso propuesto.

Las funciones de membresía fueron determinadas aplicando el algoritmo de agrupamiento *k-means* en el conjunto de entrenamiento. El algoritmo de agrupamiento es utilizado para encontrar 5 grupos para ρ_{1i} y ρ_{2i} y 4 para ρ_{3i} y ρ_{4i} . Las funciones de membresía para cada valor difuso son asignadas considerando los centroides. El centroide con el valor difuso más pequeño es asignado a *Level 1*, mientras que el centroide con el valor difuso más grande es asignado a *Level 5* o *Level 4* respectivamente para $\rho_{\tau i}$. Las funciones de membresía son denominadas $\ddot{u}_{\tau\delta}$ donde τ corresponde a la variable lingüística y δ corresponde a los valores difusos.

Las funciones de membresía son Gaussianas. Los centroides $\zeta_{\tau\delta}$ corresponden a la media de la función Gaussiana y $\sigma_{\tau\delta}$ es la desviación estándar de Ψ^H_i para las imágenes i del grupo correspondiente. Por lo tanto

$$\ddot{u}_{\tau\delta} = e^{\frac{-(p_{\tau i} - \zeta_{\tau\delta})^2}{2\sigma_{\tau\delta}^2}} \quad (3.17)$$

El siguiente paso es determinar las reglas difusas R_n del tipo:

SI ρ_{1i} es A_1 , **Y** ρ_{2i} es A_2 , **Y** ..., **ENTONCES** la salida es y_n .

donde A_r representa los valores difusos de las variables lingüísticas $p_{\tau i}$, y_n la salida de la regla y n el índice de cada regla. Sólo se eligen aquellas reglas en donde al menos uno de los antecedentes las cumple, es decir, una regla es tomada en cuenta cuando la siguiente expresión es verdadera:

$$\text{Seleccionar } R_n \text{ si } \exists p_{\tau i} : p_{\tau i} \in (A_1, A_2, A_3, A_4) \quad (3.18)$$

Por lo tanto, el sistema diseñado para cada conjunto de entrenamiento y validación tiene NR reglas. NR es variable y depende en cuantas reglas son generadas considerando la ecuación (3.18). La y_n de cada regla R_n es también determinada mediante *k-means*. El grupo k que mejor representa a Ψ^H_i del conjunto de entrenamiento es seleccionado. En este trabajo, número de grupos utilizados fue de 10. Así, y_n de cada regla toma el valor de uno de los centroides ζ_t de los 10 grupos generados. El valor promedio de Ψ^H_i de las imágenes que activan una regla R_n , $\bar{\Psi}_n^H$, es calculado para definir qué valor tomará y_n

$$\bar{\Psi}_n^H = \frac{1}{|I_n|} \sum_{i \in I_n} t_i \quad (3.19)$$

donde I_n es el conjunto de imágenes i cuyos valores $\rho_{\tau i}$ activan la regla n . Entonces, y_n toma como valor el centroide más cercano al promedio $\bar{\Psi}_n^H$, el cual es definido por:

$$y_n = \zeta_t : \zeta_t \left(\min \left(\bar{\Psi}_n^H - \zeta_t \right) \right) \quad (3.20)$$

Por último, una vez que las reglas son evaluadas, Ψ^C_i está definida por promedio ponderado de todas las reglas del sistema, como se muestra en la siguiente ecuación:

$$\Psi_i^C = \frac{\sum_{n=1}^{NR} \omega_{ni} y_n}{\sum_{n=1}^{NR} \omega_{ni}} \quad (3.21)$$

donde ω_{ni} es el peso generado por cada antecedente de cada regla.

3.2.2 Resultados del sistema difuso en B_{RS1} y B_{RS2}

Los resultados del sistema difuso para cada uno de los conjuntos de entrenamiento y validación son mostrados en la Tabla 3.3. Los resultados se encuentran en términos del coeficiente de correlación entre la complejidad determinada por el sistema difuso Ψ^C y la complejidad promedio Ψ^H especificada por el criterio humano. Además, también se muestra el número de reglas generado para cada conjunto de entrenamiento y validación.

Tabla 3.3 Desempeño del sistema difuso expresado en términos de correlación entre Ψ^C y Ψ^H obtenidos mediante validación cruzada.

Conjunto	Correlación		
	Entrenamiento	Validación	NR
1	0.8910	0.6786	48
2	0.8172	0.9394	50
3	0.8112	0.9289	47
4	0.8350	0.9132	51
5	0.8623	0.8099	46
6	0.8585	0.9165	41
7	0.8551	0.8097	51
Media	0.8472	0.8566	
Desviación estándar	0.0279	0.0957	

La media en el desempeño para entrenamiento y validación es de 0.8472 y 0.8566 respectivamente, con baja desviación estándar. Este desempeño es considerado aceptable por que el sistema es capaz de determinar la RCI con un nivel de correlación cercano al de los sujetos que participaron en la evaluación de las bases de datos B_{RS1} y B_{RS2} . El número de reglas difusas necesarias para los sistemas generados por cada conjunto de imágenes se encuentra en el rango de 41 a 51, las cuales no representan un gran costo computacional.

La Figura 3.4 muestra en términos de correlación como cada sujeto de las bases de datos se correlaciona con la media de Ψ^H . Además, se muestra como el sistema difuso propuesto se

correlaciona con Ψ^H . Este resultado del sistema difuso fue calculado utilizando el promedio de todos los conjuntos de la validación cruzada. Se puede observar como el sistema difuso propuesto tiene una mejor correlación que la mayoría de los sujetos considerados en este trabajo, es decir, estima adecuadamente el valor promedio de la RCI de acuerdo con los sujetos que participaron en el experimento.

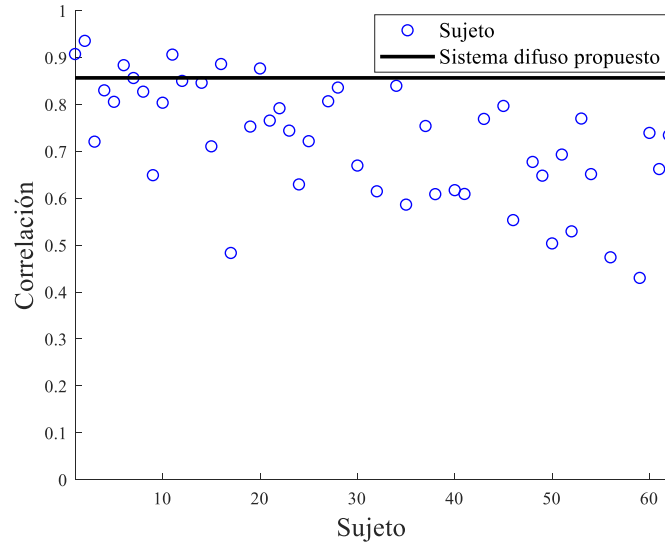


Figura 3.4 Comparación entre los sujetos aceptados de las bases de datos B_{RS1} y B_{RS2} contra el desempeño del sistema difuso propuesto en términos de correlación con el promedio de Ψ^H .

Los resultados obtenidos por el sistema difuso serán comparados en la Sección 3.4. Cabe mencionar que este sistema no se probó en la base de datos $B_{SAVOIAS}$ ya que se profundizó más en el método basado en una red neuronal artificial debido a la mejora en los resultados, el cual es presentado en la siguiente sección.

3.3 Red neuronal artificial para determinar complejidad de imágenes

El modelo de red neuronal artificial (RNA) incorpora una etapa de preprocesamiento de las características y una de estimación de la RCI. La base de datos B_{RS} es utilizada para determinar la arquitectura de la RNA y entrenarla para obtener una comparación directa con el modelo de Corchs [25]. La base de datos $B_{SAVOIAS}$ es utilizada para probar la arquitectura de la RNA. Para determinar el número óptimo de neuronas en la capa oculta de la RNA, se realiza el experimento mostrado en la Sección 3.3.1.

3.3.1 Diseño de la red neuronal artificial

La primera parte involucrada en el diseño de la red neuronal artificial es el preprocesamiento. El preprocesamiento consiste en la normalización de las características f para que los pesos de la red se ajusten dentro del mismo rango y se pueda mejorar la convergencia. El método de normalización utilizado es el mismo que el utilizado en la Sección 3.1.2. Además, con la finalidad de reducir la cantidad de entradas de la red neuronal artificial, también se utiliza el método de análisis de componentes principales, en este caso la red es alimentada con las primeras 5 componentes principales $\rho_{\tau i}$ para $1 \leq \tau \leq 5$, obtenidas de las 11 características descritas en la Sección 3.1.

La arquitectura de la red neuronal tiene 5 entradas y 1 salida. El número de neuronas en la capa oculta fue determinado mediante el siguiente experimento. Se diseñó una red neuronal de tipo *feedforward* con $\ddot{A}$ neuronas en la capa oculta donde el índice de cada neurona es representado por $\ddot{a}$ donde $1 \leq \ddot{a} \leq \ddot{A}$, la arquitectura es mostrada en la Figura 3.5.

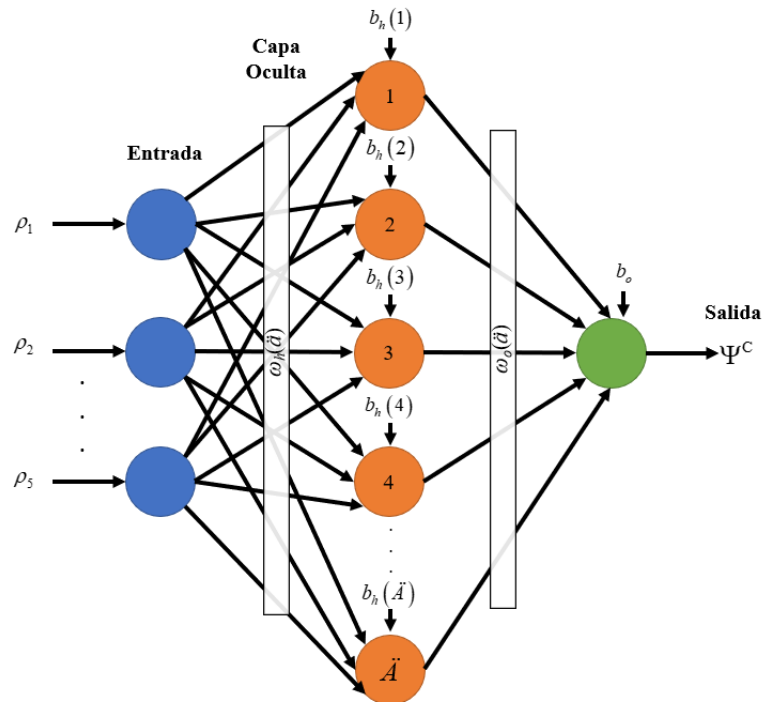


Figura 3.5 Arquitectura de la red neuronal artificial para la determinación de la RCI.

La salida de las neuronas en la capa oculta es definida como:

$$\hat{o}(\ddot{a}) = \rho_{\tau} \cdot \omega_h(\ddot{a}) + b_h(\ddot{a}) \quad (3.22)$$

donde $\omega_h(\ddot{a})$ representa los pesos y $b_h(\ddot{a})$ la polarización para cada neurona $\ddot{a}$ en la capa oculta.

Cada una de las neuronas en la capa oculta y la capa de salida utilizan una función de activación sigmoideal $\Phi(\alpha)$ la cual está representada como:

$$\Phi(\alpha) = \frac{1}{1 + e^{-\alpha}} \quad (3.23)$$

de esta manera, la salida de la red neuronal es:

$$\Psi^C = \Phi(\omega_o(\ddot{a}) \cdot \Phi(\hat{o}(\ddot{a})) + b_o) \quad (3.24)$$

donde Ψ^C es el valor de RCI estimado de la imagen de entrada, $\omega_o(\ddot{a})$ son los pesos de la neurona en la capa de salida y b_o la polarización.

El entrenamiento utilizado para esta red neuronal es el backpropagation de gradiente conjugado escalado [102] con los siguientes valores: épocas de entrenamiento = 1000, factor de aprendizaje = 5×10^{-5} , tratando de llegar a un rendimiento de 0 en el cambio del valor del gradiente, y si la red no disminuye el error durante 50 iteraciones el entrenamiento termina. Al igual que en sistema difuso, el objetivo de cada imagen en el entrenamiento es representado por t_i de la Ecuación (3.16).

El número de neuronas en la capa oculta fue determinado mediante el entrenamiento de 100 redes. Cada red tenía un número diferente de neuronas en la capa oculta, de 1 hasta 100. Se utilizaron a las imágenes de B_{RS1} como muestras de entrenamiento y las imágenes de B_{RS2} como muestras de validación. En este experimento se observó la correlación de cada red con los valores de RCI que habían sido determinados por el criterio humano para estas bases de datos. La Figura 3.6 muestra los valores de correlación de cada red con el criterio humano en el entrenamiento y validación. Se puede observar que la red con 98 neuronas en la capa oculta

obtiene el mejor desempeño durante el entrenamiento con un coeficiente de correlación de 0.904 y en la etapa de validación sólo obtiene 0.627, lo cual indica sobreentrenamiento. Tratando de encontrar el mejor balance entre el desempeño de entrenamiento y validación, se decidió utilizar 28 neuronas en la capa oculta, ya que esta red obtiene un desempeño de 0.8456 y 0.8409 en el entrenamiento y validación respectivamente.

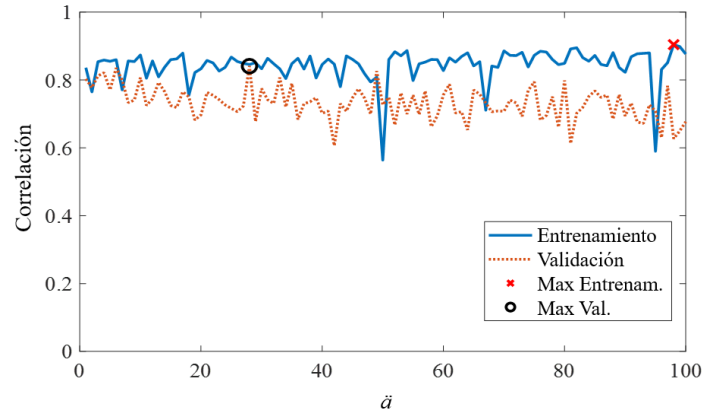


Figura 3.6 Correlación de la complejidad determinada por el humano y las 100 redes neuronales con 1 a 100 neuronas en la capa oculta.

La Figura 3.7 ilustra la regresión lineal entre la salida de la red con 28 neuronas en la capa oculta y el criterio humano, considerando el coeficiente de correlación, para el entrenamiento y validación.

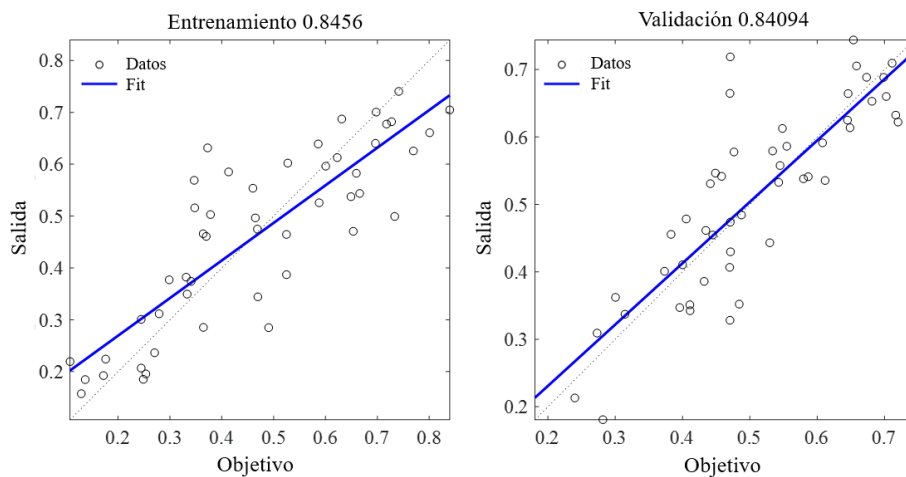


Figura 3.7 Gráficas de regresión para el entrenamiento y la validación de la red neuronal de 28 neuronas que fue entrenada con la base de datos B_{RS1} y validada con B_{RS2} .

3.3.2 Resultados y discusión del método basado en redes neuronales artificiales

El diseño de la RNA para la determinación de la RCI de imagen fue probado con diferentes esquemas experimentales y también con el método de validación cruzada, con la finalidad de obtener el desempeño en las bases de datos B_{RS} y $B_{SAVOIAS}$.

3.3.2.1 Resultados en la base de datos B_{RS1} y B_{RS2}

La base de datos B_{RS} fue diseñada con un método donde los humanos especificaban la RCI de cada imagen utilizando escalas psicométricas. Por lo tanto, se realizaron diferentes experimentos, no sólo para verificar el desempeño de la RNA basada en las características f , sino también para verificar la coherencia del criterio humano. El conjunto de muestras utilizado para entrenamiento y validación fueron diferentes en cada experimento. También, para obtener un desempeño confiable de la RNA se evaluó mediante el método de validación cruzada. Estos experimentos son descritos a continuación.

Primero, se presentan cinco experimentos, A-E, con la RNA propuesta utilizando las imágenes de B_{RS} . Cada experimento contiene diferentes configuraciones de imágenes utilizadas por la ANN en los conjuntos de entrenamiento y validación. El experimento A utiliza las imágenes de B_{RS1} para entrenamiento y las de B_{RS2} para validación. El experimento B usa B_{RS2} para entrenamiento y B_{RS1} para validación. El experimento C aleatoriamente combina las imágenes de B_{RS1} y B_{RS2} y usa 49 imágenes para entrenamiento y 49 imágenes para validación. El experimento D y E usan 30 imágenes para entrenamiento y 19 para validación utilizando sólo un conjunto de imágenes. Sin embargo, el experimento D sólo utiliza las imágenes de B_{RS1} y el experimento E las imágenes de B_{RS2} . El desempeño de cada experimento en términos del coeficiente de correlación de Pearson es mostrado en la Tabla 3.4. Estos experimentos tienen el objetivo de respaldar el argumento hecho en la Sección 2.4 sobre la coherencia del criterio humano en ambos conjuntos de datos una vez que B_{RS} ha sido procesado por el método descrito en la Sección 2.3.

Tabla 3.4 Desempeño en términos del coeficiente de correlación obtenido por los diferentes experimentos en la base de datos B_{RS} , cambiando el conjunto de muestras utilizado para entrenamiento y validación.

Experimento	Entrenamiento	Validación	Desempeño en entrenamiento	Desempeño en validación
A	B_{RS1}	B_{RS2}	0.8456	0.8409
B	B_{RS2}	B_{RS1}	0.8660	0.8315
C	B_{RS1} y B_{RS2} mezcladas, 49 muestras cada caso		0.8640	0.8635
D	30 muestras de B_{RS1}	19 muestras de B_{RS1}	0.8870	0.8838
E	30 muestras de B_{RS2}	19 muestras de B_{RS2}	0.8708	0.8651

El experimento A fue el realizado en la etapa de diseño de la RNA, las muestras en B_{RS1} son las de entrenamiento y las muestras en B_{RS2} de validación. El desempeño obtenido fue un coeficiente de correlación con el criterio humano 0.8456 y 0.8409 en entrenamiento y validación respectivamente. Este primer experimento equivale al experimento utilizado para evaluar el método reportado en [25], donde el desempeño obtenido en el entrenamiento es de 0.86 y en validación de 0.81. Cabe resaltar que con este experimento es posible confirmar la coherencia del criterio humano entre los dos grupos de personas que generaron B_{RS} , ya que se usó B_{RS1} para entrenamiento y B_{RS2} para validación. Recordando que cada base de datos fue realizada por un grupo de personas diferente. El experimento B tiene el objetivo de saber si la red tiene un buen desempeño para determinar correctamente la RCI si es entrenada con una base de datos (B_{RS2}) con un rango de complejidad menor al rango de la base de datos con la que se valida (B_{RS1}). Además, este experimento ayuda a verificar la coherencia del criterio humano en ambos conjuntos de datos ya que el proceso de entrenamiento y validación es realizado con imágenes que fueron evaluadas por diferentes grupos de personas. El desempeño obtenido en este experimento alcanza una correlación de 0.8660 para entrenamiento y 0.8315 en validación.

El experimento C es realizado con la finalidad de observar el comportamiento de la red neuronal cuando es expuesta al mismo número de muestras de entrenamiento y validación, pero seleccionando aleatoriamente las muestras de ambos conjuntos de datos. En este experimento, la RNA es entrenada con el criterio de los dos grupos de personas. Por lo tanto, es esperado que el desempeño en validación mejore. El resultado de este experimento fue de 0.8640 para entrenamiento y 0.8635 en validación. Se puede notar que los resultados mejoran en comparación con los de los experimentos A y B.

En el experimento D, sólo las muestras de B_{RS1} fueron utilizadas. La red fue entrenada con 30 muestras y validada con 19, ambos conjuntos de B_{RS1} . El propósito de este experimento fue observar cómo es el desempeño de la RNA cuando las muestras son tomadas del mismo grupo de personas. El resultado fue un mejor desempeño en comparación a los otros experimentos, obteniendo un desempeño de 0.8870 y 0.8838 para entrenamiento y validación respectivamente. Esto es debido a que la red aprende de un criterio más consistente, ya que sólo se trata del grupo

El experimento E es similar al experimento D, pero en E solamente se utiliza B_{RS2} . Sin embargo, los resultados no fueron tan buenos como los obtenidos en D. El desempeño en E fue de 0.8708 para entrenamiento y 0.8651 para validación. Un posible caso por el cual se bajó el desempeño puede ser por que el rango de complejidad de imágenes de B_{RS2} es más pequeño.

Para complementar estos experimentos, se utilizó el método de validación cruzada (experimento F) igual al mencionado en la Sección 3.2.2 del sistema difuso, por el cual se obtuvieron los resultados de la Tabla 3.5 en términos de correlación.

Tabla 3.5 Desempeño de la red neuronal artificial usando validación cruzada en la base de datos B_{RS} .

Conjunto de validación cruzada	Entrenamiento	Validación
1	0.8379	0.8405
2	0.8660	0.9774
3	0.8559	0.8860
4	0.9107	0.9056
5	0.8075	0.8040
6	0.8806	0.8637
7	0.8286	0.8200
Media	0.8553	0.8710
Desviación estándar	0.0345	0.0589

De acuerdo con los valores mostrados en la Tabla 3.5, se puede notar que el desempeño promedio de la red durante validación es 0.8710. Este resultado es más alto que el obtenido en los experimentos A, B, y C, y es el desempeño de la red esperado en una aplicación real.

3.3.2.2 Resultados en la base de datos $B_{SAVOIAS}$

Como es mencionado en la Sección 1.1.3, $B_{SAVOIAS}$ fue generada de una manera más confiable y robusta que B_{RS} . Por lo tanto, $B_{SAVOIAS}$ fue seleccionada para probar la RNA que estima la respuesta a la RCI. La RNA es probada mediante el método de validación cruzada,

donde es entrenada y validada con cada conjunto de la base de datos $B_{SAVOIAS}$. Esto es porque el criterio humano es subjetivo y como se comentó en la Sección 2.1, la percepción y la RCI cambia de acuerdo con el tipo de imágenes bajo análisis. Para respaldar este argumento hecho en la Sección 2.1, se realiza también un pequeño experimento, el cual es descrito al final de esta sección y consiste en entrenar la RNA con imágenes de una categoría de $B_{SAVOIAS}$ y validarla con imágenes de otra categoría de $B_{SAVOIAS}$.

Los conjuntos de entrenamiento y validación para la validación cruzada fueron generados de la misma forma que el experimento en la base de datos B_{RS} . Pero debido a que $B_{SAVOIAS}$ tiene más imágenes y son múltiplos de 10, se generaron 10 conjuntos de entrenamiento y validación para cada categoría de imágenes de la base de datos. Para estos conjuntos de imágenes, el 90% de las imágenes fueron utilizadas para el entrenamiento y el 10% para validación. Todas las imágenes son utilizadas una vez para validación y cada conjunto de validación tiene imágenes de todo el rango de RCI. La Tabla 3.6 muestra los resultados en términos de correlación para la RNA en cada conjunto de imágenes mediante validación cruzada de la base de datos $B_{SAVOIAS}$, durante el entrenamiento Ent y validación Val.

Tabla 3.6 Desempeño de la RNA en términos del coeficiente de correlación utilizando validación cruzada en la base de datos $B_{SAVOIAS}$.

Conjunto de validación	B_{Sc}		B_{Ad}		B_{Vi}		B_{Ob}		B_{In}		B_{Art}		B_{Sup}	
	Ent	Val	Ent	Val	Ent	Val	Ent	Val	Ent	Val	Ent	Val	Ent	Val
1	0.862	0.733	0.852	0.844	0.897	0.760	0.856	0.668	0.893	0.911	0.831	0.849	0.930	0.961
2	0.899	0.799	0.850	0.855	0.911	0.863	0.880	0.589	0.953	0.972	0.805	0.822	0.941	0.951
3	0.833	0.845	0.832	0.761	0.895	0.847	0.899	0.734	0.866	0.973	0.809	0.842	0.888	0.987
4	0.866	0.828	0.855	0.871	0.864	0.712	0.802	0.790	0.872	0.950	0.779	0.854	0.960	0.923
5	0.853	0.744	0.864	0.788	0.848	0.802	0.771	0.658	0.974	0.900	0.848	0.772	0.951	0.981
6	0.839	0.787	0.855	0.883	0.887	0.625	0.832	0.823	0.954	0.880	0.840	0.834	0.868	0.945
7	0.868	0.907	0.873	0.758	0.837	0.850	0.758	0.725	0.905	0.929	0.863	0.746	0.996	0.978
8	0.835	0.791	0.882	0.855	0.802	0.818	0.796	0.581	0.947	0.944	0.866	0.642	0.945	0.948
9	0.863	0.817	0.879	0.908	0.840	0.789	0.830	0.868	0.935	0.957	0.838	0.736	0.963	0.966
10	0.819	0.770	0.888	0.822	0.869	0.833	0.842	0.713	0.955	0.955	0.847	0.845	0.951	0.948
Media	0.854	0.802	0.863	0.835	0.865	0.790	0.827	0.715	0.925	0.937	0.833	0.794	0.939	0.959
Desviación estándar	0.023	0.051	0.017	0.051	0.034	0.074	0.045	0.094	0.038	0.031	0.027	0.069	0.037	0.020

Los resultados en la Tabla 3.6 cambian de acuerdo con el tipo de imagen utilizada para estimar su RCI. El mejor desempeño fue obtenido para las imágenes de B_{Sup} con un promedio de correlación de 0.939 en entrenamiento y 0.959 en validación. El desempeño más bajo fue con las imágenes de B_{Ob} con 0.827 en entrenamiento y 0.715 en validación. Estas variaciones en los

desempeños son por el tipo de imágenes. Por ejemplo, la Figura 3.8 muestra imágenes de cada categoría de $B_{SAVOIAS}$. En el caso de las imágenes de B_{Sup} , la mayoría son compuestas de figuras geométricas. Por lo tanto, basándose en las características utilizadas como el número de regiones, así como el número de regiones incrementa, la RCI estimada también incrementa, similar a lo que sucede con el criterio humano. El mismo caso ocurre con las imágenes de B_{In} . El valor de RCI esta correlacionado con el número de objetos en las imágenes. Para otras categorías, existe una gran subjetividad que depende de diversos factores como el nivel de conocimiento de los objetos presentes en las imágenes. B_{Ob} es considerada una de las categorías más subjetivas y por eso es donde se obtienen los más bajos resultados.



Figura 3.8 Imágenes de muestra de $B_{SAVOIAS}$.

Para mostrar que la percepción de la complejidad de imagen varía de acuerdo con el tipo de imágenes bajo cuestión, como fue comentado en la Sección 2.2. La Tabla 3.7 muestra el desempeño en validación obtenido mediante el entrenamiento de la RNA con imágenes de una categoría de $B_{SAVOIAS}$ y validada con imágenes de otra categoría de $B_{SAVOIAS}$. Para obtener estos resultados, la RNA fue entrenada y validada 30 veces. Luego, el desempeño fue calculado mediante la media de los valores obtenidos en validación para generalizar mejor el comportamiento de la ANN bajo diferentes condiciones iniciales. Estas condiciones iniciales se refieren al cambio del orden en que las imágenes alimentaban la RNA y la inicialización aleatoria de los pesos de la red. El desempeño obtenido en la etapa de entrenamiento no es mostrado ya que estos valores se encuentran documentados en la Tabla 3.6.

En la Tabla 3.7 se puede observar cómo los desempeños, de acuerdo con el coeficiente de correlación en validación son más bajos cuando la RNA es validada con imágenes de diferente categoría respecto a las imágenes utilizadas en el entrenamiento. Estos resultados pueden ser comparados con los resultados obtenidos en la Tabla 3.6. Por ejemplo, en la Tabla 3.6 se muestra que la RNA obtiene 0.827 en entrenamiento y 0.715 en validación para la categoría B_{Ob} . Si otra categoría es utilizada en el entrenamiento para validar las imágenes de B_{Ob} , el máximo desempeño es 0.5427, como se muestra en la Tabla 3.7 cuando la RNA es entrenada con las imágenes de B_{Sc} . Si la RNA es entrenada con B_{Sup} y validada con B_{Ob} el desempeño es 0.2541, el cual es el desempeño más bajo de este experimento. Por su parte, el desempeño más alto es obtenido por este experimento fue 0.7929. Este desempeño fue alcanzado cuando la RNA fue entrenada con B_{Sc} y validada con B_{In} . Es importante notar que los valores máximos de la Tabla 3.7 se encuentran en la diagonal principal y corresponden a los casos cuando los conjuntos de entrenamiento y validación fueron los mismos. Por lo tanto, el desempeño de 0.7929 podría ser aceptable, siendo mejor que el desempeño obtenido por B_{Ob} en la Tabla 3.6. Sin embargo, comparado con la misma categoría de imágenes, B_{In} , cuando el mismo tipo de imágenes son utilizadas en el entrenamiento, el desempeño incrementa a 0.937 como lo fue mostrado en la Tabla 3.6. En conclusión, el argumento hecho en la Sección 2.1 de que la percepción de la complejidad varía de acuerdo con el tipo de imágenes bajo análisis es respaldado mediante este experimento. Ya que, es probado que los resultados en validación son bajos cuando el modelo es validado con diferentes imágenes a las utilizadas en el entrenamiento.

Tabla 3.7 Desempeño en validación en términos del coeficiente de correlación obtenido mediante el entrenamiento de la RNA con imágenes de una categoría de $B_{SAVOIAS}$ y validada con imágenes de otra categoría de $B_{SAVOIAS}$.

		Conjunto de validación						
		B_{Sc}	B_{Ad}	B_{Vi}	B_{Ob}	B_{In}	B_{Art}	B_{Sup}
Conjunto de entrenamiento	B_{Sc}	0.9838	0.4920	0.3309	0.5427	0.7929	0.6301	0.7160
	B_{Ad}	0.4892	0.9776	0.6009	0.3742	0.6890	0.5894	0.6626
	B_{Vi}	0.4390	0.6298	0.9803	0.3846	0.7831	0.4309	0.3971
	B_{Ob}	0.6663	0.5409	0.4751	0.9750	0.7111	0.6224	0.5292
	B_{In}	0.5164	0.4995	0.4838	0.3156	1	0.5145	0.5841
	B_{Art}	0.6351	0.5374	0.4011	0.3596	0.6276	0.9204	0.7942
	B_{Sup}	0.4426	0.6264	0.4409	0.2541	0.6769	0.4508	1

3.4 Comparación de resultados con otros métodos

Existen muchos métodos para estimar la RCI. Desafortunadamente, no hay una base de datos estándar para comparar el desempeño de esos métodos. Con la finalidad de proveer algo de información sobre los desempeños, se presenta la Tabla 3.8, sin embargo, es importante saber que no todos los resultados se encuentran en términos de correlación. Los resultados mostrados en la Tabla 3.8 se encuentra de acuerdo con sus respectivas publicaciones.

Tabla 3.8 Desempeño de los métodos al estimar la RCI.

Método	Desempeño reportado contra el criterio humano	Número de imágenes	Base de datos	Base de datos pública
Descomposición por bloques [103]	0.53 Correlación	150	-Base de datos con patrones a blanco y negro	No
Características simples [40]	0.60 Correlación promedio del mejor resultado contra cada conjunto	1420	$-B_{SAVOIAS}$	Si
Red neuronal convolucional [15]	0.71 Correlación	1420	$-B_{SAVOIAS}$	Si
Sistema difuso [36]	0.73 Correlación	---	-Base de datos con imágenes artísticas a color	No
<i>Random forest</i> [104]	0.78 Correlación	500	-Base de datos con imágenes artísticas	No
Optimización de enjambre de partículas [25]	0.83 Correlación	98	$-B_{RS}$	Si
Red neuronal artificial [27]	0.83 Correlación	800	-Base de datos compuesta de imágenes en las categorías de arte abstracto, abstracto no artístico, arte representacional, representacional no artístico, escenas naturales y escenas hechas por humanos	No
Método propuesto basado en una RNA	0.84 Correlación	1518	$-B_{RS} -B_{SAVOIAS}$	Si
Método propuesto basado en un sistema difuso [105]	0.857 Correlación	98	$-B_{RS}$	Si
<i>Random forest</i> [106]	0.91 Correlación	1164	-Base de datos con patrones a blanco y negro	No
Grado de compresión GIF [71]	0.95 Correlación	180	-Base de datos con patrones a blanco y negro	No
Variedad y distribución de color [107]	“No consistente con la percepción humana”	---	-Base de datos con escenas reales	No
Sistema de lógica difusa [12]	“Algo correlacionado”	150	-Base de datos con escenas reales	No
Red neuronal artificial [45]	“Consistente”	600	---	No

Se puede observar en la Tabla 3.8 que los métodos propuestos (remarcados en negritas) tienen una alta correlación con el criterio humano, el método basado en una RNA obtiene una correlación promedio de 0.84 en las bases de datos B_{RS} y $B_{SAVOIAS}$. Por su parte, el sistema difuso obtiene una correlación de 0.857 en la base de datos B_{RS} . Como fue mencionado antes,

solamente el sistema basado en una RNA fue probado en $B_{SAVOIAS}$ ya que fue el que obtuvo mejores desempeños y robustez en B_{RS} con un coeficiente de correlación de 0.8710 en validación y fue con el que se continuó el trabajo de investigación. Estos desempeños quedan únicamente por debajo de los métodos en [106] y [71], sin embargo, estos métodos sólo trabajan con imágenes artificiales de patrones en blanco y negro. Lo cual facilita la estimación de la RCI y consistencia con el criterio humano, tal y como pasa con las imágenes de B_{Sup} . Por su parte, los métodos propuestos superan el desempeño de los métodos basados en redes neuronales artificiales, incluyendo redes neuronales convolucionales, además, el método propuesto basado en RNA es probado con una mayor cantidad de imágenes. Como se puede observar en la Tabla 3.8, no es posible realizar una comparación directa con todos los métodos porque algunos proveen sus desempeños en variables lingüísticas o subjetivas como en [12], [45] y [107].

3.5 Conclusiones

Este capítulo presentó dos métodos para determinar la RCI. Uno de los métodos determina la RCI mediante un sistema de inferencia difuso y el otro mediante una red neuronal artificial. El sistema de inferencia difuso es alimentado mediante 4 componentes principales obtenidas de 11 características visuales simples, la red neuronal artificial es alimentada con 5 componentes principales de estas 11 características. Ambos métodos obtuvieron una alta correlación con el criterio humano respecto a los algoritmos del estado del arte incluyendo redes neuronales profundas.

De acuerdo con los resultados obtenidos mediante el método basado en redes neuronales, se verificó la coherencia de los dos grupos de personas que desarrollaron la base de datos B_{RS} . Esta conclusión es respaldada mediante los resultados de los experimentos A, B, C, y F de la Sección 3.3.3.

En el caso de la red neuronal artificial, se utilizó un método para optimizar el número de neuronas en la capa oculta. De acuerdo con este método, 28 fue el número de neuronas que obtiene el mejor resultado para determinar la RCI, para las diferentes evaluaciones que se realizaron en este trabajo.

Basado en los comentarios anteriores, se puede decir que los métodos propuestos en este capítulo para estimar la RCI obtienen un alto desempeño respecto al criterio humano. Por lo tanto, estos métodos pueden ser utilizados en diferentes algoritmos computacionales para una toma de decisiones similar a la de los humanos. Por lo que, en los capítulos posteriores se presentaran dos aplicaciones prácticas de la RCI. Una de las aplicaciones se centra en determinar el número de grupos necesarios para segmentar una imagen y la otra aplicación tiene el objetivo de ayudar a realiza una mejor evaluación de la segmentación de imágenes.

CAPÍTULO 4. COMPLEJIDAD DE IMAGEN PARA DETERMINAR EL NÚMERO DE GRUPOS EN ALGORITMOS DE SEGMENTACIÓN DE IMÁGENES

La segmentación de imágenes es una de las tareas más importantes en el área de visión por computadora, su objetivo es etiquetar cada uno de los píxeles de una imagen para representarla en segmentos y que pueda ser analizada por humanos o máquinas [89]. Algunas de las aplicaciones de la segmentación de imágenes son el análisis de imágenes médicas, la navegación automática, realidad aumentada, sistemas de videovigilancia, reconocimiento de huellas dactilares o iris, motores de búsqueda en imágenes, análisis de imágenes satelitales, análisis de microorganismos, entre otras [84], [85], [108], [109].

A través de los años, muchos algoritmos de segmentación de imágenes han sido desarrollados. Dentro de estos algoritmos se encuentran los basados en agrupamiento difuso, debido a sus excelentes resultados y bajo costo computacional. Algunos de estos algoritmos están reportados en [88]–[90], [110]–[112]. Sin embargo, a pesar de sus buenos resultados, estos métodos comúnmente necesitan de la intervención humana para ajustar sus parámetros de acuerdo con cada imagen procesada. Por ejemplo, un parámetro que comúnmente necesita ser ajustado manualmente es el número de grupos, C . En algunos trabajos, el valor de C es constante, incluso cuando las imágenes tienen una composición o número de objetos diferente [91], [110]. En consecuencia, los métodos de segmentación basados en agrupamiento reducen su desempeño. Para prevenir esta problemática, el valor de C es manualmente modificado para cada imagen, con la finalidad de obtener el mejor resultado [88], [89]. Sin embargo, este enfoque no es factible para aplicaciones donde se tiene que procesar una gran cantidad de imágenes o donde no es posible que una persona esté supervisando el proceso.

Hay muchos métodos para determinar C que son utilizados en algoritmos de agrupamiento. Por ejemplo, los métodos *Elbow* [113], *Fuzzy Silhouette* [114], *Gap Statistic* [115], RL-FCM [116], entre otros. Sin embargo, estos métodos no están especializados para la segmentación de imágenes. Estos métodos no consideran la información espacial de las imágenes o sólo proveen información que debe ser analizada por humanos y algunos otros requieren de un alto costo computacional cuando trabajan con imágenes. Por lo tanto, la aplicación de estos métodos en la

segmentación de imágenes no es factible [91]. Por su parte, existen métodos actuales para segmentación de imágenes por agrupamiento que determinan automáticamente el valor de C . Estos métodos son AFCF propuesto por Lei *et al.* [91] y FAS-SGC propuesto por Jia *et al.* [117]. AFCF primero procesa las imágenes utilizando superpíxeles [118]–[120] y aplica el algoritmo *Density Peak*, DP [121]. Luego, AFCF transforma la información obtenida por DP a un nuevo espacio dimensional para facilitar la identificación de los superpíxeles que pueden ser considerados como centroides para determinar C . Finalmente, para segmentar las imágenes, AFCF utiliza el algoritmo *Prior Entropy-Based Fuzzy Clustering*, PEFC. FAS-SGC también trabaja con superpíxeles, utilizando una reconstrucción morfológica adaptativa basada en la eliminación mínima regional, AMR-RMR, y determina automáticamente C utilizando el método *Eigenvalue Gradient Clustering*, EGC [117]. AFCF y FAS-SGC determinan automáticamente el valor de C basándose en la información espacial de las imágenes. Sin embargo, el valor de C determinado sólo beneficia a sus algoritmos de agrupamiento. Ese valor de C no puede ser utilizado de una manera general en otros algoritmos de agrupamiento para segmentación de imágenes. Esto es demostrado mediante los resultados descritos en este capítulo, donde, se muestra que utilizar el valor de C determinado por AFCF y FAS-SGC afectan negativamente el desempeño de los algoritmos comparados con utilizar un valor fijo de C , c_f .

De acuerdo con la literatura, hay muchos algoritmos de agrupamiento difuso para segmentación de imágenes con buenos resultados, pero requieren de la intervención humana para determinar C y obtener esos buenos resultados. Por lo tanto, es necesario realizar más investigación con el objetivo de solucionar la determinación automática de C , para que estos algoritmos de segmentación puedan alcanzar su mejor desempeño. Por lo tanto, en este capítulo se presenta el método *Automatic Estimation of Number of Clusters by Image Complexity* (AENCIC). La contribución principal de AENCIC se centra en utilizar la respuesta a la complejidad de imagen, Ψ , para determinar automáticamente el mejor valor de C para cada imagen en los algoritmos de agrupamiento difuso del estado del arte para segmentación de imágenes, de acuerdo con las métricas de la base de datos BSDS500 [92]. La idea general de cómo trabaja AENCIC con los algoritmos de agrupamiento difuso para segmentación de imágenes es mostrada en la Figura 4.1, donde I representa una imagen digital y S una

segmentación. Es importante mencionar que para este caso $\Psi = \Psi^C$, ya que la respuesta a la complejidad utilizada en este capítulo y el Capítulo 5 es obtenida computacionalmente. AENCIC fue diseñado para lidiar con la segmentación de imágenes de entornos reales, ya que este tipo de segmentación es un problema mal definido donde existe gran variación del valor de C necesario para obtener un buen desempeño en las imágenes [122]. Esta variación de C para cada imagen afecta considerablemente el desempeño de los algoritmos de segmentación si c_f es utilizado. Otra contribución de AENCIC es que mejora el desempeño de los algoritmos de agrupamiento y permite que trabajen automáticamente.

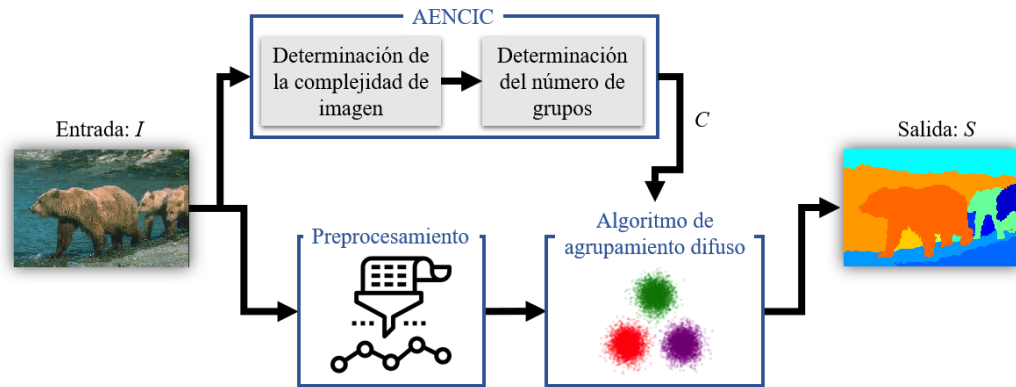


Figura 4.1 Diagrama general de un algoritmo de agrupamiento difuso para segmentación de imágenes utilizando C determinado por AENCIC.

Una restricción en el diseño de AENCIC es, como en otros trabajos, se asume que las imágenes son adquiridas con una buena relación señal ruido. Por lo tanto, el problema del ruido como en otros métodos del estado del arte no es considerado.

Este capítulo se organiza de la siguiente manera. La Sección 4.1 presenta la relación entre la RCI y el número de grupos necesario para segmentación una imagen correctamente. La Sección 4.2 muestra el problema de utilizar un número de grupos fijo. La Sección 4.3 describe el diseño de AENCIC. La Sección 4.4 presenta los experimentos realizados y los resultados. Finalmente, en la Sección 4.5 están las conclusiones de este capítulo.

4.1 Motivación para utilizar la complejidad de imagen

Cuando se trabaja con imágenes de entornos reales, existe una gran variación de la cantidad de información semántica Ξ contenidas en ellas. Desde imágenes con gran Ξ , como el interior de un supermercado hasta imágenes con poca Ξ , como una hoja sobre el agua. Comúnmente, para imágenes con gran Ξ , es conveniente utilizar un valor alto de número de grupos, C , para realizar la segmentación, mientras que para imágenes con poca Ξ es conveniente utilizar un valor bajo de C [123].

Algunos algoritmos de agrupamiento para segmentación de imágenes utilizan un número de grupos fijo c_f , sin lograr buenos resultados. Por lo que es mejor determinar C para cada imagen bajo cierto criterio. Por lo tanto, en este capítulo se propone el método AENCIC, el cual utiliza la RCI Ψ como criterio para determinar automáticamente el valor de C para segmentar una imagen. AENCIC considera que hay una relación directa entre el nivel de Ψ y el mejor C para cada imagen. Ψ es relacionada a la segmentación de imágenes porque Ψ se centra en una comparación de la cantidad de información con significado en las imágenes para el humano, mientras que la segmentación consiste en dividir las imágenes en regiones de pixeles con un significado. La estimación de C bajo la suposición anterior es probada mediante el incremento del desempeño de los algoritmos de agrupamiento difuso en la sección de resultados.

4.2 Problemas de utilizar un número de grupos fijo

Si un algoritmo de agrupamiento para segmentación de imágenes no tiene un método para estimar C , es común utilizar un valor fijo c_f . Sin embargo, este valor fijo no funcionaría en imágenes de entornos naturales debido a la gran variabilidad de la composición espacial que presentan. En consecuencia, no es recomendable utilizar el mismo C para todas las imágenes, ya que esto afectará directamente el desempeño de los algoritmos de agrupamiento. Una alternativa es utilizar un valor de C variable para cada imagen. Sin embargo, es un tema de investigación aún en desarrollo. La Figura 4.2 ilustra por que utilizar un valor de C variable es aún un tema de investigación en desarrollo.

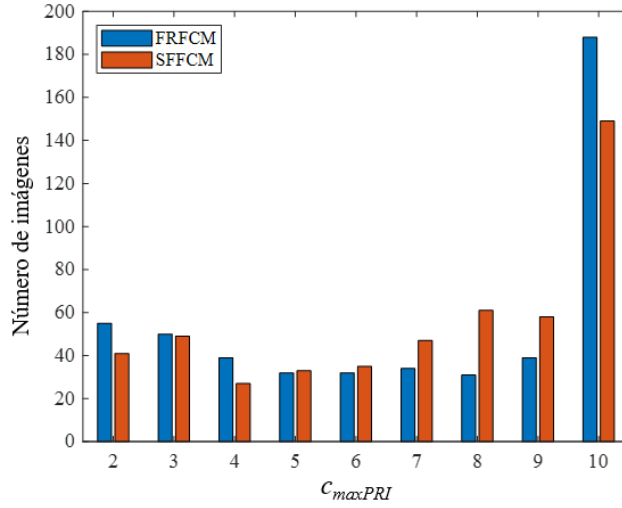


Figura 4.2 Cantidad de imágenes en la que se obtuvo el máximo PRI de acuerdo con el número de grupos c utilizado para los algoritmos FRFCM [88] y SFFCM [89].

La Figura 4.2 muestra el número de imágenes donde el valor máximo de PRI fue obtenido considerando un valor de C , C_{maxPRI} , en los algoritmos FRFCM (*Fast and Robust Fuzzy C-Means*) [88] y SFFCM (*Supapixel-Based Fast Fuzzy C-Means*) [89]. La métrica PRI fue seleccionada como la métrica principal para evaluar AENCIC ya que comparado con las otras métricas del *benchmark* BSDS500 [92], es la que tiene mejor correlación con el criterio humano en la tarea de segmentación de imágenes [124]. Para este experimento se utilizó $2 \leq C \leq 10$. Se puede observar que la distribución del número de imágenes de acuerdo con C_{maxPRI} es casi uniforme, excepto para $C_{maxPRI} = 10$. Para esta gráfica, parece ser que lo mejor sería utilizar un número alto de grupos ya que $C_{maxPRI} = 10$ produce muchas segmentaciones con valores altos de PRI. Sin embargo, esto no es totalmente cierto, debido a la información que presenta la Tabla 4.1. La información de la Tabla 4.1 fue obtenida con la base de datos BSDS500 utilizando $2 \leq c_f \leq 10$ para los métodos FRFCM y SFFCM. Esta información muestra que para $6 \leq c_f \leq 10$ el valor de PRI no cambia de 0.72 para FRFCM y 0.75 para SFFCM, pero el valor de la desviación estándar incrementa conforme c_f incrementa. Por lo que, aunque $c_f = 10$ obtiene muchos valores altos de PRI, también puede generar muchos valores bajos de PRI para otras imágenes. Otra desventaja es que utilizar valores de C altos, es que en la mayoría de los algoritmos de agrupamiento el costo computacional también incrementa. Consecuentemente, el diseño de un

método que determine automáticamente el mejor número de grupos para cada imagen utilizando Ψ para segmentación, puede solucionar el problema de determinar manualmente C .

Tabla 4.1 Desempeño de los algoritmos FRFCM y SFFCM con la base de datos BSDS500 utilizando $2 \leq C \leq 10$ para la métrica PRI.

Algoritmo	C	PRI	Desviación estándar
FRFCM	2	0.60	0.1040
	3	0.68	0.0926
	4	0.70	0.0993
	5	0.71	0.1070
	6	0.72	0.1123
	7	0.72	0.1176
	8	0.72	0.1216
	9	0.72	0.1239
	10	0.72	0.1272
SFFCM	2	0.61	0.1175
	3	0.69	0.0995
	4	0.73	0.0998
	5	0.74	0.1056
	6	0.75	0.1100
	7	0.75	0.1160
	8	0.75	0.1187
	9	0.75	0.1216
	10	0.75	0.1241

4.3 Diseño de AENCIC

El método AENCIC trata de determinar el mejor número de grupos utilizando la RCI de cada imagen Ψ_i . AENCIC es descrito en las próximas secciones cubriendo el siguiente material. La relación entre Ψ y C para la segmentación de imágenes es analizada en la Sección 4.3.1. En la Sección 4.3.2 la información proveída por cada Ψ_i es relacionada con C , asumiendo que $2 \leq C \leq 10$. La justificación de este intervalo es porque 2 es considerado el mínimo número de grupos para segmentar una imagen (fondo y objeto de interés) y 10 porque se decidió incrementar intervalo comúnmente utilizado en la literatura de $2 \leq C \leq 6$ [88], [89], [91], ya que se encontró que algunas imágenes son segmentadas de una mejor manera con un valor alto de C , como lo fue mostrado en el análisis de la Figura 4.2.

4.3.1 Relación entre complejidad de imagen y número de grupos

El algoritmo FRFCM [88] fue utilizado para analizar la relación entre Ψ y C para las imágenes de la base de datos BSDS500. El mapa de calor mostrado en la Figura 4.3 es el

resultado de este análisis. El eje de las abscisas corresponde al valor de Ψ_i , en orden ascendente de las 200 imágenes de entrenamiento de BSDS500. El eje de las ordenadas representa el valor de C utilizado por el algoritmo FRFCM para cada imagen de entrenamiento, donde $2 \leq C \leq 10$. La escala de color indica el valor de la métrica PRI normalizada (PRIN) para cada imagen y es calculada mediante:

$$\text{PRIN}_i(C) = \frac{2\text{PRI}_i(C) - \min(\text{PRI}_i)}{\max(\text{PRI}_i) - \min(\text{PRI}_i)} - 1 \quad (4.1)$$

donde $1 \leq i \leq 200$ es el índice de cada imagen de entrenamiento de la base de datos. Una función de suavizado fue aplicada a los valores obtenidos por cada C del mapa de calor, con la finalidad de facilitar la observación de la tendencia de los datos entre Ψ y C .

En la Figura 4.3 se puede observar que hay muy pocas imágenes con valores máximos de PRI cuando $C = \{2, 3\}$. Esto es debido a la distribución de Ψ en la base de datos. Para relacionar el comportamiento de los datos mostrados en la Figura 4.3 con la distribución de Ψ , la distribución de Ψ es mostrada en la Figura 4.4. En esta figura se puede notar como la mayoría de las imágenes tienen una complejidad entre 0.4 y 0.6, es decir, no existe una gran cantidad de imágenes con baja complejidad como para ser correctamente segmentadas con $C = \{2, 3\}$.

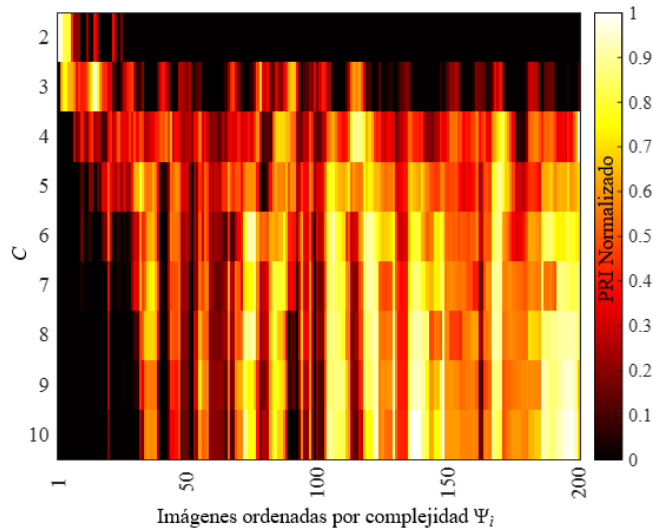


Figura 4.3 Mapa de calor que representa la relación entre la complejidad de imagen y número de grupos en términos de la métrica PRI normalizada (PRIN) del conjunto de entrenamiento de BSDS500.

Además, en la Figura 4.3 también se puede notar lo siguiente:

- Los valores máximos de PRI para $C = 2$ y $C = 3$ están relacionados a las imágenes con baja complejidad. Cuando la complejidad incrementa, los valores de PRI decremantan para $C = 2$ y $C = 3$.
- Para las primeras 30 imágenes con baja complejidad, si C incrementa, los valores de PRI decremantan. El mejor resultado para estas imágenes corresponde a $C = \{2,3,4\}$.
- Para las imágenes con alta complejidad, las últimas 20 imágenes de la Figura 4.3, los mejores valores de PRI se dan con $C = \{8,9,10\}$.
- La mayor concentración de valores altos de PRI empiezan en $C = 6$ y a partir de la imagen 100. Este comportamiento es debido a la distribución de complejidad de las imágenes en cuestión, ya que la mayoría de las imágenes se encuentran concentradas en un intervalo reducido de complejidad.

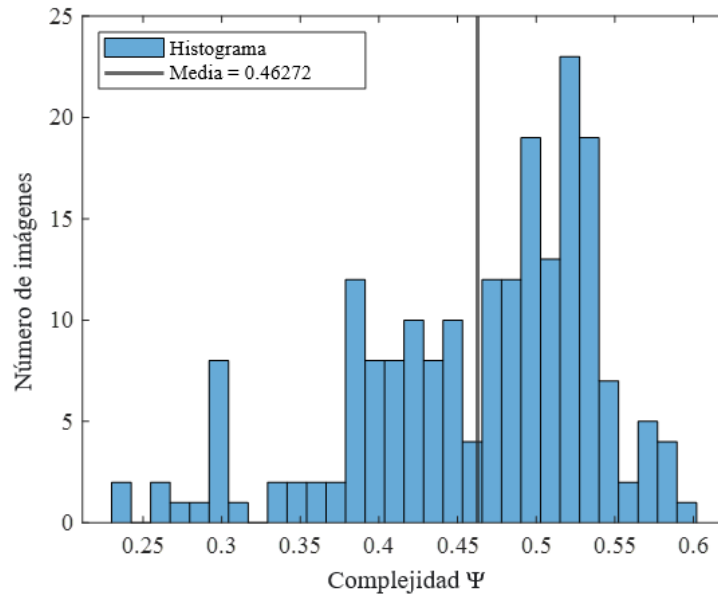


Figura 4.4 Histograma de complejidad de imagen del conjunto de entrenamiento de la base de datos BSDS500.

En general, la información de la Figura 4.3 indica que hay una correlación proporcional entre Ψ y C utilizado por los algoritmos de agrupamiento para obtener el mejor desempeño. Basado en este hallazgo, la siguiente sección describe un método para determinar automáticamente el valor de C para segmentar las imágenes utilizando Ψ .

4.3.2 Determinación automática del número de grupos

AENCIC determina automáticamente al número de grupos C necesarios para segmentar una imagen mediante intervalos de Ψ , considerando que hay una correlación directa entre Ψ y el valor de C . Ya que la distribución de Ψ_i del conjunto de entrenamiento no es uniforme, primero es necesario encontrar el dominio de los valores de la RCI para determinar los intervalos. El valor máximo y mínimo de Ψ_i del conjunto de imágenes de entrenamiento en BSDS500 son $\max(\Psi_{\text{Train}}) = 0.6009$ y $\min(\Psi_{\text{Train}}) = 0.2358$. Por lo tanto, los intervalos de complejidad son definidos dentro de un rango $Ra = 0.3651$. Luego, la distribución de los valores máximos de PRI de acuerdo a los valores de C es analizada. Esta distribución considerando el algoritmo FRFCM es ilustrada en la Figura 4.5. Ya que la distribución de Ψ_i no es uniforme, los límites de los intervalos son asignados proporcionalmente a el número de imágenes.

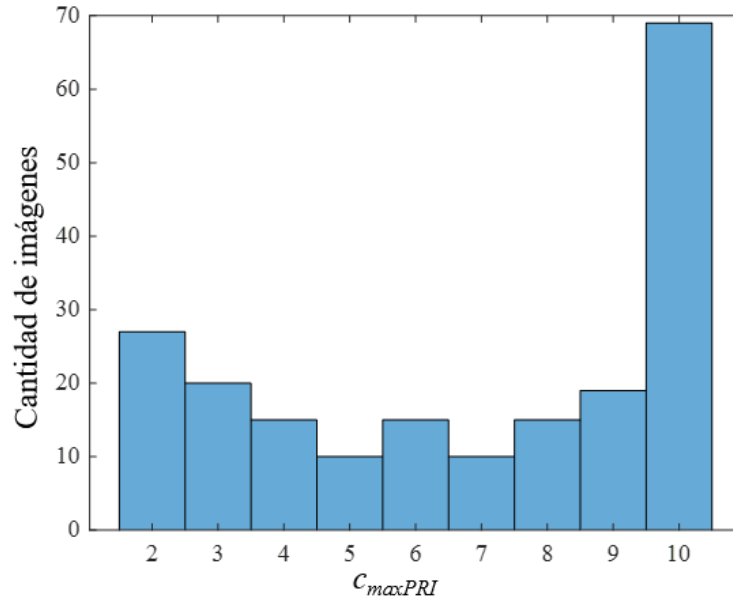


Figura 4.5 Cantidad de imágenes en la que se obtuvo el máximo PRI de acuerdo con el número de grupos utilizado en el algoritmo FRFCM.

El histograma mostrado en la Figura 4.5 tiene los siguientes valores en cada bin:

$$H(C) = \{27, 20, 15, 10, 15, 10, 15, 19, 69\} \quad (4.2)$$

Para establecer los límites de complejidad, primero se determinan los límites de Ψ de acuerdo con el número total de imágenes relacionadas a cada valor de C . Esto se lleva a cabo definiendo los límites del histograma $H(C)$ como $H_{\text{Lim}}(\hat{a})$,

$$H_{\text{Lim}}(\hat{a}) = \{H(2), H(2) + H(3), \dots, H(2) + H(3) + \dots + H(10)\} \quad (4.3)$$

donde $H_{\text{Lim}}(\hat{a})$ es el arreglo de los límites acumulados del número de imágenes donde C_{maxPRI} fue encontrado, para $1 \leq \hat{a} \leq 8$ se obtienen los siguientes valores:

$$H_{\text{Lim}}(\hat{a}) = \{27, 47, 62, 72, 87, 97, 112, 131\} \quad (4.4)$$

Los valores de $H_{\text{Lim}}(\hat{a})$ son traducidos a límites de complejidad mediante la definición de un paso Pa para dividir Ra . Ya que el conjunto de imágenes de entrenamiento tiene una cardinalidad de 200, el paso $Pa = Ra / 200 = 0.0018$. Pa puede ser interpretado como la resolución de complejidad sobre Ra para establecer un rango de complejidad para un valor de C dado basado en las imágenes que obtuvieron el valor máximo de PRI en ese valor de C . Por lo tanto, los imites de complejidad son:

$$\Psi_{\text{Lim}}(\hat{a}) = H_{\text{Lim}}(\hat{a}) Pa + \min(\Psi_{\text{Train}}) \quad (4.5)$$

y basado en Ψ_{Lim} , C es calculado mediante

$$C_i = \begin{cases} 2 & \text{if } \Psi_i < \Psi_{\text{Lim}}(1) \\ 3 & \text{if } \Psi_{\text{Lim}}(1) \leq \Psi_i < \Psi_{\text{Lim}}(2) \\ 4 & \text{if } \Psi_{\text{Lim}}(2) \leq \Psi_i < \Psi_{\text{Lim}}(3) \\ 5 & \text{if } \Psi_{\text{Lim}}(3) \leq \Psi_i < \Psi_{\text{Lim}}(4) \\ 6 & \text{if } \Psi_{\text{Lim}}(4) \leq \Psi_i < \Psi_{\text{Lim}}(5) \\ 7 & \text{if } \Psi_{\text{Lim}}(5) \leq \Psi_i < \Psi_{\text{Lim}}(6) \\ 8 & \text{if } \Psi_{\text{Lim}}(6) \leq \Psi_i < \Psi_{\text{Lim}}(7) \\ 9 & \text{if } \Psi_{\text{Lim}}(7) \leq \Psi_i < \Psi_{\text{Lim}}(8) \\ 10 & \text{if } \Psi_{\text{Lim}}(8) \leq \Psi_i \end{cases} \quad (4.6)$$

Los valores de los límites de complejidad para BSDS500 son:

$$\Psi_{\text{Lim}}(\hat{a}) = \{0.285, 0.322, 0.349, 0.367, 0.395, 0.413, 0.440, 0.475\} \quad (4.7)$$

Estos valores fueron determinados de acuerdo con el rango de complejidad y la distribución de máximos PRI cuando el valor de C cambia. Para propósitos ilustrativos, la Figura 4.6 muestra cuales valores de C fueron seleccionados para cada imagen mediante AENCIC de acuerdo con los valores de Ψ_{Lim} . Se puede observar que, en algunos casos, la determinación de C no es ideal. Sin embargo, el valor de C seleccionado es mejor que utilizar un valor fijo de C o que obtenerlo mediante otros métodos del estado del arte, lo cual será demostrado en la sección de resultados de este capítulo. Una vez que Ψ_{Lim} es determinado para cierta base de datos o algún tipo de imágenes, C es encontrado automáticamente mediante AENCIC e incorporado en el proceso de segmentar una imagen por un algoritmo de agrupamiento difuso como es mostrado en el diagrama de la Figura 4.7.

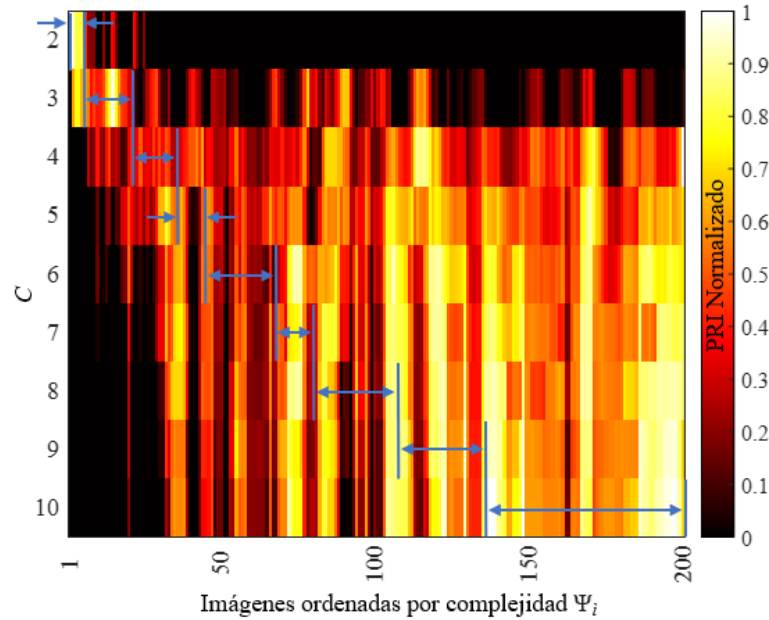


Figura 4.6 Mapa de calor que representa cual valor de C fue seleccionado para cada imagen mediante AENCIC de acuerdo con los valores de Ψ_{Lim} , en términos del PRI normalizado para las imágenes de entrenamiento de BSDS500.

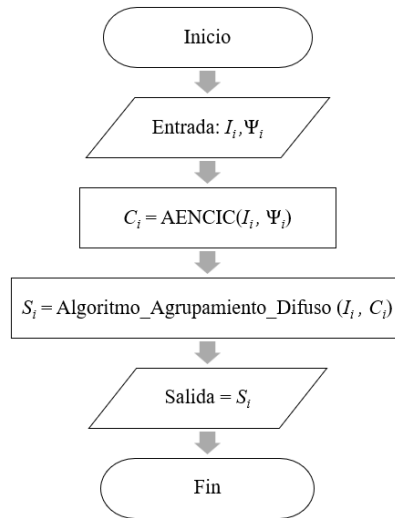


Figura 4.7 Diagrama de flujo que ilustra el proceso de segmentación de una imagen mediante AENCIC.

4.4 Experimentos y resultados

Esta sección muestra los resultados de determinar C mediante AENCIC considerando la subjetividad del criterio mediante el uso de Ψ . La Sección 4.4.1 describe los algoritmos de segmentación utilizados en los experimentos para evaluar a AENCIC y la configuración de parámetros utilizada. Los resultados de AENCIC son analizados en la Sección 4.4.2. Luego, AENCIC es comparado contra los métodos de AFCF [91] y FAS-SGC [117] para determinar C y también contra el número fijo de C , c_f , donde se obtienen los mejores resultados para cada algoritmo de segmentación por agrupamiento difuso, en las imágenes de BSDS500. Por último, la Sección 4.4.3 presenta un análisis de correlación entre los métodos para determinar C considerados en este capítulo.

4.4.1 Verificación de los códigos de los algoritmos de segmentación por agrupamiento y configuración de parámetros

Esta sección da una breve descripción de los algoritmos de segmentación por agrupamiento utilizados para evaluar el valor de C determinado por AENCIC. Los algoritmos son FRFCM, SFFCM, los tres algoritmos de segmentación utilizados por AFCF que son llamados $\text{AFCF}^{\text{SLIC}}$, AFCF^{LSC} y $\text{AFCF}^{\text{MMGR}}$, y también FAS-SGC. Estos algoritmos difusos se seleccionaron porque tienen los mejores resultados con respecto a los algoritmos del estado del arte. Son métodos de

segmentación no supervisados y requieren un bajo costo computacional. Todas esas características los hacen apropiados para diversas aplicaciones. Esta sección también describe la configuración de los parámetros involucrados en esos algoritmos y cómo es el comportamiento de los algoritmos.

4.4.1.1 FRFCM

FRFCM [88] es un algoritmo basado en agrupamiento difuso que utiliza un método de reconstrucción morfológica y filtrado de funciones de membresía para volverlo más rápido y robusto. Para la reproducción de este algoritmo se utilizó el código provisto por los autores, utilizando un radio de disco para la reconstrucción morfológica $rd = 3$ y un tamaño de ventana para el filtro de 3×3 . Se realizó una prueba similar a la reportada en [88] y los resultados fueron reproducidos correctamente, sólo con pequeñas variaciones causadas por la inicialización aleatoria de los centroides en el algoritmo de agrupamiento.

4.4.1.2 SFFCM

SFFCM [89] también es un algoritmo basado en agrupamiento difuso que incorpora el uso de superpíxeles mediante una reconstrucción de gradiente morfológica multi escala lo que lo vuelve eficiente y preciso. Los autores proporcionan el código para SFFCM, el cual se logró reproducir con éxito de acuerdo con los parámetros indicados en [89]. Estos parámetros fueron un radio de disco $rd = 3$ para la reconstrucción morfológica, 50 como número máximo de iteraciones, una ponderación exponencial de 2 y la condición de convergencia en 10^{-5} .

4.4.1.3 Algoritmos de segmentación AFCF

El trabajo reportado en [91] propone un método para determinar automáticamente C y es llamado AFCF (*Automatic Fuzzy Clustering Framework for image segmentation*). AFCF es un *framework* que segmenta imágenes con un valor de C definido automáticamente. El funcionamiento de AFCF puede ser descrito en 4 pasos, los cuales son:

1. Cálculo de super píxeles. Los super píxeles pueden ser obtenidos mediante cualquier método, pero los autores se limitan a los métodos: SLIC (*Simple Linear Iterative*

Clustering) [118], LSC (*Linear Spectral Clustering*) [125] y MMGR (*Multiscale Morphological Gradient Reconstruction*) [120]. Cabe mencionar que MMGR es propuesto también por los autores de AFCF.

2. Obtención de la gráfica de decisión. La gráfica de decisión es obtenida mediante el algoritmo DPA (*Density Peak Algorithm*) [121], el cual se refiere al producto de la densidad local de cada super píxel y la mínima distancia entre cada super píxel.
3. Determinación del número de grupos C . C es determinado al utilizar el algoritmo DBA (*Density Balance Algorithm*). DBA mapea la gráfica de decisión obtenida por DPA. Este mapeo incrementa la diferencia entre aquellos superpíxeles que pueden ser considerados como centroides. Por lo tanto, facilita la determinación de C .
4. Segmentación de la imagen. La segmentación se lleva a cabo mediante el algoritmo PEFC (*Prior Entropy-Based Fuzzy Clustering*) de acuerdo con el valor de C determinado en el paso 3.

Lei *et al.* [91], proponen tres algoritmos de segmentación de imágenes basados en AFCF. Estos tres algoritmos cambian la forma en la que se generan los superpíxeles, los cuales son llamados $\text{AFCF}^{\text{SLIC}}$, AFCF^{LSC} y $\text{AFCF}^{\text{MMGR}}$. Cada algoritmo segmenta las imágenes de diferente manera y obtienen su propio valor de C para las imágenes de BSDS500. Los autores comparten el código para la reproducción de resultados de $\text{AFCF}^{\text{MMGR}}$, y empleando el método de superpíxeles desarrollado por ellos se obtienen los mejores resultados de acuerdo con [91].

El código de $\text{AFCF}^{\text{MMGR}}$ fue utilizado para reproducir los resultados de $\text{AFCF}^{\text{SLIC}}$ y AFCF^{LSC} pero cambiando las funciones de superpíxeles utilizadas. Para el algoritmo SLIC se utilizó la función de superpíxeles provista por los autores en [118] y para LSC la función provista por Li en [119]. Además, ya que SLIC y LSC no son métodos adaptativos, las imágenes fueron divididas en 400 superpíxeles, como es indicado en [91].

Con la finalidad de verificar que los resultados obtenidos por estos 3 algoritmos, se trató de reproducir un experimento presentado en [91], en el cual muestran las segmentaciones de 5 imágenes de la base de datos BSDS500, las cuales son: 370036.jpg, 35070.jpg, 182053.jpg, 97010.jpg y 376020.jpg. Las segmentaciones reportadas en [91] y las obtenidas mediante esta

implementación utilizando los métodos $\text{AFCF}^{\text{SLIC}}$, AFCF^{LSC} , y $\text{AFCF}^{\text{MMGR}}$ son ilustradas en la Figura 4.8.

Los resultados mostrados en la Figura 4.8 para AFCF^{LSC} mediante la reproducción de resultados son muy similares a los reportados en [91]. Sin embargo, los resultados de segmentación de $\text{AFCF}^{\text{SLIC}}$ y $\text{AFCF}^{\text{MMGR}}$ varían insignificadamente y también varían en el número de grupos C que determinan. La Tabla 4.2 muestra la comparación del valor de C obtenido en [91] para las 5 imágenes segmentadas y el valor de C obtenido con AFCF en la reproducción de los resultados. Los números en negritas indican los casos donde el valor de C fue el mismo. De acuerdo con [91], $\text{AFCF}^{\text{SLIC}}$ y $\text{AFCF}^{\text{MMGR}}$ obtienen el mismo valor de C y AFCF^{LSC} varía solamente en las imágenes “97010.jpg” y “376020.jpg”. En la reproducción del experimento, solamente para las imágenes “370036.jpg” y “35070.jpg” se encontró el mismo valor de C que el reportado en [91]. AFCF^{LSC} fue el algoritmo donde se obtuvieron los resultados más similares entre lo deportado en [91] y la reproducción de resultados.

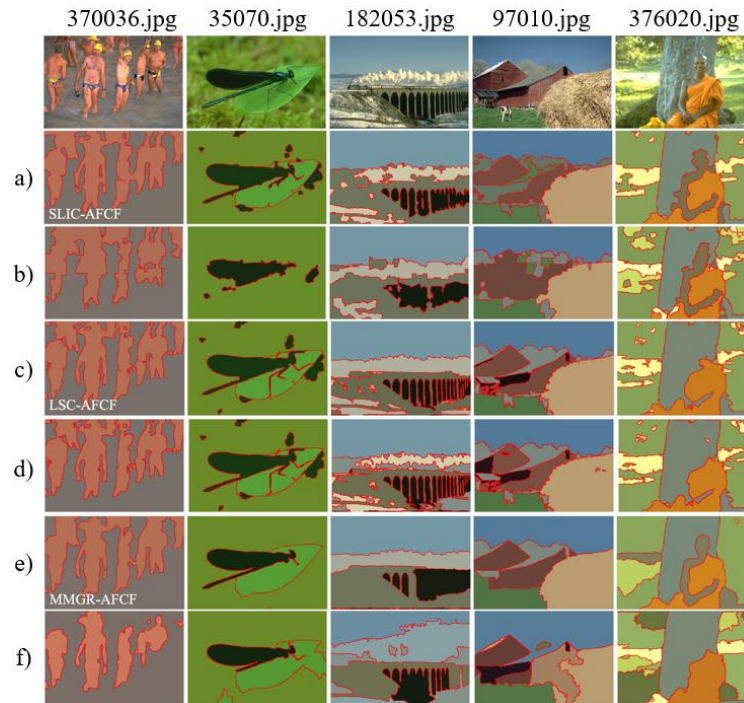


Figura 4.8 Comparación de resultados del experimento de la segmentación de imágenes reportada por AFCF en [91] y la reproducción de resultados. a) Segmentación reportada en [91] por $\text{AFCF}^{\text{SLIC}}$, b) Segmentación obtenida con $\text{AFCF}^{\text{SLIC}}$, c) Segmentación publicada en [91] por AFCF^{LSC} , d) Segmentación generada con AFCF^{LSC} , e) Segmentación documentada en [91] por $\text{AFCF}^{\text{MMGR}}$, f) Segmentación obtenida por $\text{AFCF}^{\text{MMGR}}$.

Tabla 4.2 Comparación del número de grupos C reportado en [91] y el obtenido reproduciendo el experimento.

Imagen	Reportado en [91]			Reproduciendo los resultados		
	AFCF ^{SLIC}	AFCF ^{LSC}	AFCF ^{MMGR}	AFCF ^{SLIC}	AFCF ^{LSC}	AFCF ^{MMGR}
370036.jpg	2	2	2	2	2	2
35070.jpg	3	3	3	3	3	3
182053.jpg	4	4	4	4	5	5
97010.jpg	5	6	5	3	6	6
376020.jpg	5	4	5	4	4	6

4.4.1.4 FAS-SGC

FAS-SGC (*Fast and Automatic Image Segmentation Using Supersixel-Based Graph Clustering*) [117] segmenta las imágenes mediante los superpíxeles obtenidos por el algoritmo AMR-RMR (*Adaptive Morphological Reconstruction based on Regional Minimum Removal*). FAS-SGC determina el valor de C utilizando EGC (*Eigenvalue Gradient Clustering*). Los autores de FAS-SGC no proveen el código del algoritmo. Sin embargo, es posible realizar una comparación parcial con los resultados reportados en [117], ya que se reporta el desempeño obtenido con ese valor de C para los algoritmos FRFCM y SFFCM.

4.4.2 Análisis de resultados

Esta sección describe un análisis de los resultados considerando todas las imágenes de BSDS500. El análisis de la Tabla 4.3 muestra los resultados de los algoritmos de agrupamiento FRFCM, SFFCM y los algoritmos de segmentación de AFCF sin la etapa que determina automáticamente el valor de C . Estas variaciones de AFCF son llamadas PEFC^{SLIC}, PEFC^{LSC} y PEFC^{MMGR}. Los algoritmos de segmentación son evaluados con seis formas distintas de determinar C . La primera forma es con un valor fijo $c_f = \{2, \dots, 10\}$ para todas las imágenes de BSDS500, el cual, de el mejor resultado en términos de PRI. Este intervalo es porque AENCIC determina C en este mismo rango y no sería una comparación justa utilizar un intervalo diferente. Las tres formas siguientes para determinar C son mediante los métodos AFCF^{SLIC}, AFCF^{LSC} y AFCF^{MMGR}. El siguiente método es el empleado por FAS-SGC, el cual es llamado EGC-Adaptive y sólo es mostrado para los algoritmos FRFCM y SFFCM ya que son los únicos mencionados en [117] y la métrica BDE no es presentada. La última forma para determinar el valor de C es mediante AENCIC. La métrica PRI es mejor si se obtiene un valor alto y las

métricas VI, GCE y BDE son mejores entre menor sea su valor. Esto es indicado en la Tabla 4.3 por $\uparrow$ y $\downarrow$ respectivamente.

Tabla 4.3 Resultados obtenidos en los algoritmos difusos variando el número de grupos C .

Métodos para obtener el valor de C		Algoritmo	PRI $\uparrow$	VI $\downarrow$	GCE $\downarrow$	BDE $\downarrow$
Valor fijo de C con el mejor PRI	$c_f = 9$	FRFCM	0.72	3.36	0.42	14.29
	$c_f = 8$	SFFCM	0.75	2.32	0.29	13.24
	$c_f = 5$	PEFC ^{SLIC}	0.74	2.81	0.37	14.23
	$c_f = 4$	PEFC ^{LSC}	0.74	2.65	0.35	13.55
	$c_f = 4$	PEFC ^{MMGR}	0.73	2.32	0.29	15.42
AFCF ^{SLIC}		FRFCM	0.70	2.71	0.33	14.09
		SFFCM	0.72	2.20	0.24	15.08
		PEFC ^{SLIC}	0.71	2.50	0.29	15.66
		PEFC ^{LSC}	0.71	2.46	0.29	14.80
		PEFC ^{MMGR}	0.71	2.25	0.25	15.22
AFCF ^{LSC}		FRFCM	0.70	2.69	0.33	13.94
		SFFCM	0.71	2.20	0.24	15.04
		PEFC ^{SLIC}	0.71	2.47	0.29	15.63
		PEFC ^{LSC}	0.71	2.43	0.28	14.95
		PEFC ^{MMGR}	0.71	2.23	0.24	15.47
AFCF ^{MMGR}		FRFCM	0.70	2.70	0.34	13.96
		SFFCM	0.72	2.19	0.24	14.59
		PEFC ^{SLIC}	0.72	2.49	0.30	15.03
		PEFC ^{LSC}	0.72	2.43	0.29	14.09
		PEFC ^{MMGR}	0.72	2.23	0.25	15.28
EGC-Adaptive		FRFCM	0.68	2.55	0.30	-
		SFFCM	0.69	2.21	0.23	-
AENCIC		FRFCM	0.73	3.22	0.41	14.11
		SFFCM	0.76	2.28	0.28	13.37
		PEFC ^{SLIC}	0.75	2.88	0.36	14.14
		PEFC ^{LSC}	0.75	2.82	0.36	13.51
		PEFC ^{MMGR}	0.73	2.35	0.28	16.68

Los resultados en términos de la métrica PRI para cada algoritmo con diferentes valores de C son gráficamente ilustrados en la Figura 4.9. La Figura 4.9 también muestra el valor máximo de PRI para cada algoritmo (maxPRI). maxPRI indica el desempeño obtenido al seleccionar manualmente el mejor resultado para cada imagen utilizando $2 \leq C \leq 10$. EGC-Adaptive no es mostrado para los métodos basados en PEFC (especificado con NaN) porque los resultados mostrados en [117] sólo mencionan a FRFCM y SFFCM, y el código no es compartido para obtener el valor de C para todas las imágenes mediante EGC-Adaptive. Es importante mencionar que AENCIC tiene el objetivo de determinar automáticamente el valor de C únicamente, y no es un nuevo método de agrupamiento para segmentar imágenes. Por lo tanto, AENCIC puede ser empleado por los algoritmos de agrupamiento de segmentación de imágenes para tratar de

obtener el mejor resultado como si C fuera seleccionado manualmente para cada imagen. El análisis de los resultados de la Tabla 4.3 y la Figura 4.9 es presentado en las siguientes secciones.

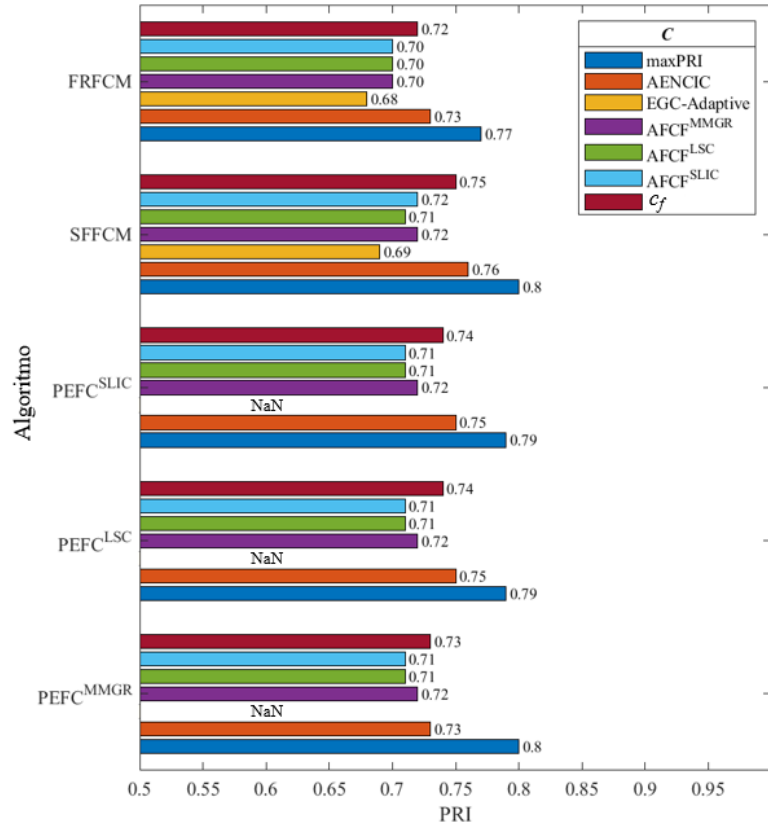


Figura 4.9 Resultados obtenidos en términos de PRI para cada algoritmo de segmentación de imágenes utilizando diferentes métodos para determinar C .

4.4.2.1 Resultados de FRFCM

Considerando la información de la Tabla 4.3 y la Figura 4.9, FRFCM utilizando $c_f = 9$ obtiene $PRI = 0.72$. Por su parte, si el valor de C es; C de $AFCE^{SLIC}$, C de $AFCE^{LSC}$ y C de $AFCE^{MMGR}$ de la reproducción de los experimentos, FRFCM tiene un promedio de $PRI = 0.70$. El desempeño de FRFCM en [91] es $PRI = 0.71$ con C de $AFCE^{MMGR}$ y en [117] utilizando C de EGC-Adaptive el $PRI = 0.68$. Por lo tanto, se puede concluir que utilizar el valor adaptivo de AFCE o FAS-SGC afecta el desempeño de FRFCM. De acuerdo con C de AENCIC, el PRI de FRFCM incrementa a 0.73. El mejor resultado para las métricas VI y GCE fue obtenido por C

de $AFCF^{LSC}$. Sin embargo, se observa que estas métricas mejoran cuando el valor de C se vuelve más pequeño. Esta mejora es debido a que hay muy pocas regiones en la segmentación y las diferencias entre S y GT disminuyen. El mejor resultado de la métrica BDE fue también para C de $AFCF^{LSC}$. Sin embargo, en este caso, la varianza de esta métrica considerando el cambio de C es muy pequeña y depende más del tipo de algoritmo de segmentación utilizado.

4.4.2.2 Resultados de SFFCM

En el caso de SFFCM, un valor de $c_f = 8$ para todas las imágenes obtiene un $PRI = 0.75$, para C de $AFCF^{SLIC}$ y C de $AFCF^{MMGR}$, $PRI = 0.72$, para C de $AFCF^{LSC}$ $PRI = 0.71$ y para C de EGC-Adaptive $PRI = 0.69$. Estos hallazgos sugieren que tampoco es conveniente utilizar un valor de C obtenido por los métodos adaptativos de AFCF o FAS-SGC con SFFCM. Además, los resultados reportados en [91] de SFFCM para C de $AFCF^{MMGR}$ obtienen un $PRI = 0.73$, lo cual es más bajo que utilizar $c_f = 8$ para todas las imágenes. Cuando C de AENCIC, $PRI = 0.76$, siendo un mejor desempeño que utilizar $c_f = 8$. El desempeño de ANECIC en las métricas VI y GCE es mejor que los obtenidos por c_f , y la variación en BDE es muy baja.

4.4.2.3 Resultados de las variantes de AFCF

$PEFC^{SLIC}$ y $PEFC^{LSC}$ obtienen un $PRI = 0.74$, con $c_f = 5$ para $PEFC^{SLIC}$ y $c_f = 4$ para $PEFC^{LSC}$. Si C es cambiado C de $AFCF^{SLIC}$ o C de $AFCF^{LSC}$, el desempeño de $PEFC^{SLIC}$ y $PEFC^{LSC}$ es $PRI = 0.71$ y para C de $AFCF^{MMGR}$ $PRI = 0.72$. Estos resultados indican que $PEFC^{SLIC}$ y $PEFC^{LSC}$ tienen mejor desempeño con c_f que utilizando AFCF, en el caso de la reproducción de los algoritmos de AFCF. Por su parte, en los resultados reportados en [91] $PEFC^{SLIC}$ tiene un $PRI = 0.71$ y $PEFC^{LSC}$ un $PRI = 0.73$ utilizando AFCF, lo cual sigue siendo más bajo que utilizando c_f . Finalmente, cuando $PEFC^{SLIC}$ y $PEFC^{LSC}$ trabajan con C de AENCIC, obtiene un $PRI = 0.75$, lo cual es mejor que utilizar c_f y el método AFCF.

En la reproducción de AFCF, $PEFC^{MMGR}$ obtuvo un $PRI = 0.73$ con $c_f = 4$. Para C de $AFCF^{SLIC}$ y C de $AFCF^{LSC}$ el PRI fue de 0.71, y cuando el experimento fue reproducido con el código compartido sin ninguna modificación utilizando $AFCF^{MMGR}$ [126] el PRI fue de 0.72. Además, cuando C de AENCIC es utilizado en $PEFC^{MMGR}$ el valor de $PRI = 0.73$ siendo igual

al PRI cuando $c_f = 4$. Los resultados reportados en [91] para $PEFC^{MMGR}$ utilizando C de $AFCF^{MMGR}$ obtiene $PRI = 0.76$, lo cual es igual a el resultado obtenido con $SFFCM$ para C de AENCIC.

Sin embargo, es difícil observar el efecto de AENCIC sólo analizando directamente el resultado de la métrica PRI. Para observar de una mejor manera el efecto de AENCIC, se fija una meta a alcanzar en cada algoritmo en términos de PRI. Esta meta se refiere al valor máximo de PRI que puede ser alcanzado por cada algoritmo de agrupamiento difuso solamente modificando el valor de C . Por ejemplo, la meta en $FRFCM$ es obtener un $PRI = 0.77$. Este valor de PRI para $FRFCM$ fue el valor máximo obtenido al variar manualmente el valor de C para cada imagen en una base de datos. Así, la meta de AENCIC es determinar el valor de C de forma automática para alcanzar el valor máximo de PRI, en este caso $PRI=0.77$. La Tabla 4.4 muestra el porcentaje de eficiencia de acuerdo con la meta de cada algoritmo de segmentación cuando el valor de C es determinado automáticamente por AENCIC y los otros métodos para determinar C del estado del arte. Por ejemplo, considerando a $FRFCM$, si un método para determinar C hace que $FRFCM$ obtenga un $PRI = 0.77$, entonces el porcentaje de eficiencia sería de 100% ya que se alcanzó la meta.

Tabla 4.4 Porcentaje de eficiencia alcanzada en cada algoritmo de segmentación de acuerdo con la métrica PRI mediante los diferentes métodos para determinar C .

Algoritmo		C				
		Fijo	$AFCF^{SLIC}$	$AFCF^{LSC}$	$AFCF^{MMGR}$	EGC-Adaptive
	$FRFCM$	93.51%	90.91%	90.91%	90.91%	88.31%
	$SFFCM$	93.75%	90.00%	88.75%	90.00%	86.25%
	$PEFC^{SLIC}$	93.67%	89.87%	89.87%	91.14%	NaN
	$PEFC^{LSC}$	93.67%	89.87%	89.87%	91.14%	NaN
	$PEFC^{MMGR}$	91.25%	88.75%	88.75%	90.00%	NaN
Media		93.17%	89.88%	89.63%	90.64%	87.28%
						94.19%

De acuerdo con los resultados mostrados en la Tabla 4.4, es posible darse cuenta de que el método AENCIC obtiene el mejor desempeño al determinar C en términos de PRI, ya que obtiene una eficiencia de 94.19% en promedio. Es importante mencionar que los métodos adaptativos $AFCF$ y EGC -Adaptive tiene una eficiencia menor que utilizar un valor fijo de C . Esto es porque $AFCF$ y EGC -Adaptive fueron diseñados para trabajar solamente con un algoritmo de segmentación.

De esta manera, AENCIC demuestra ser un método que ayuda a los algoritmos de agrupamiento difuso a trabajar automáticamente mediante la determinación de C con un desempeño superior a otros métodos del estado del arte. Esto es debido a que utilizar C obtenido por AENCIC hace que los algoritmos de agrupamiento trabajen aproximadamente al 94% de su máximo desempeño. Por ejemplo, comparado con el valor de C obtenido por AFCF que hace trabajar a los algoritmos aproximadamente al 90%, tal y como es observado en la Tabla 4.4.

4.4.3 Análisis de correlación entre los valores de C determinados

El análisis de correlación entre C y c_{maxPRI} es llevado a cabo para mejorar el análisis de los resultados sobre el método AENCIC. Este análisis consiste en obtener el coeficiente de correlación entre los valores C de cada método adaptativo, y el valor de c_{maxPRI} cuando el máximo PRI es obtenido para cada imagen, de cada algoritmo de segmentación en el rango de $2 \leq C \leq 10$. Estos coeficientes de correlación son mostrados en la Tabla 4.5. Ya que no se tiene suficiente información ni el código de FAS-SGC, no fue posible obtener los valores de C de EGC-Adaptive. Por lo tanto, FAS-SGC no es considerado en este análisis.

En la Tabla 4.5 se puede observar que para los métodos basados en AFCF y el valor de c_{maxPRI} , región sombreada en la Tabla 4.5, el valor más alto de correlación es para C de AFCF^{MMGR} y c_{maxPRI} de FRFCM con un valor de 0.21. La correlación entre la tendencia de como C es determinado automáticamente por cada método de AFCF y c_{maxPRI} para cada algoritmo es relativamente baja. Otro punto para comentar es que los coeficientes de correlación de C de AENCIC son mayores que los coeficientes de C obtenido de AFCF cuando son correlacionados con los valores de c_{maxPRI} . Estos valores son remarcados con negritas en la Tabla 4.5, donde la mejor correlación se da en c_{maxPRI} de FRFCM y AENCIC con un coeficiente de 0.41. Hay una buena correlación entre los valores adaptativos basados en AFCF, entre C de AFCF^{SLIC} y C de AFCF^{LSC} obtienen un valor de 0.63. Sin embargo, la correlación de C de AFCF^{MMGR} obtiene un coeficiente de 0.35 con C de AFCF^{SLIC} y con C de AFCF^{LSC}. Por lo tanto, el método de superpíxeles utilizado en AFCF afecta en la determinación del valor de C para cada imagen, y no puede ser apto para ser utilizado en diferentes algoritmos de segmentación. Sólo puede ser utilizado para el algoritmo que fue diseñado. En contraste, AENCIC está basado en el criterio

humano y no sólo en la información espacial de las imágenes, por lo que puede obtener un mejor valor de C para cada imagen y mejorar el desempeño de diferentes algoritmos de segmentación.

Tabla 4.5 Coeficientes de correlación entre C obtenido por cada método adaptativo y C correspondiente al máximo PRI de cada algoritmo de segmentación.

	C de AFCF ^{SLIC}	C de AFCF ^{LSC}	C de AFCF ^{MMGR}	C de AENCIC
C de AFCF ^{SLIC}	-	0.63	0.35	0.24
C de AFCF ^{LSC}	0.63	-	0.35	0.21
C de AFCF ^{MMGR}	0.35	0.35	-	0.26
C de AENCIC	0.24	0.21	0.26	-
C_{maxPRI} de FRFCM	0.17	0.14	0.21	0.41
C_{maxPRI} de SFFCM	0.16	0.16	0.19	0.38
C_{maxPRI} de SLIC	0.15	0.14	0.19	0.36
C_{maxPRI} de LSC	0.12	0.11	0.18	0.39
C_{maxPRI} de MMGR	0.07	0.11	0.18	0.34

4.5 Conclusiones

Este capítulo presentó a AENCIC, un nuevo método para determinar automáticamente el número de grupos, C , utilizado por los algoritmos de agrupamiento difuso que segmentan imágenes de entornos reales. AENCIC está basado en el concepto de RCI, Ψ , la cual es calculada mediante el algoritmo difuso de la Sección 3.2.1. Ψ fue considerada para establecer una relación entre la cantidad de información semántica de las imágenes, y el mejor valor de C requerido por los algoritmos de agrupamiento para segmentación. Esta idea trata de simular la percepción humana del agrupamiento para segmentar una imagen. Para llevar a cabo esta simulación, AENCIC utiliza un esquema basado en intervalos de Ψ estableciendo una relación entre Ψ y el valor de C necesario para segmentar cada imagen.

Los resultados demostraron que AENCIC representa una buena alternativa para determinar C ya que mejora el desempeño de los algoritmos de agrupamiento para segmentación, Sección 4.4.2. Esta mejora fue probada contra el uso de un número fijo de grupos y otros métodos del estado del arte para determinar C , donde AENCIC ayudó a alcanzar cerca del 95% del máximo desempeño de los algoritmos de segmentación por agrupamiento, mientras que otros algoritmos como AFCF sólo alcanza el 90%. Esta mejora fue posible ya que AENCIC imita el criterio humano y ayuda a lidiar con la variabilidad de los GT generados por humanos. De esta manera, AENCIC obtiene un coeficiente de correlación de 0.41 con respecto al mejor valor de C

determinado manualmente para cada imagen, mientras que otros algoritmos del estado del arte como AFCF sólo obtienen 0.21. Además, AENCIC es un método automático, que no necesita de la intervención humana para determinar C , lo que ayuda a que muchos algoritmos de segmentación puedan trabajar sin esa intervención del humano.

CAPÍTULO 5. ENFOQUE DIFUSO PARA LA EVALUACIÓN DE LA SEGMENTACIÓN DE IMÁGENES BASADO EN LA COMPLEJIDAD DE IMAGEN

Como fue mencionado en el Capítulo 4, la segmentación de imágenes es una tarea que ha tenido gran impacto en el área de procesamiento digital de imágenes y visión por computadora [89]–[91]. De acuerdo con la literatura, cuando la tarea de segmentación de imágenes no tiene como objetivo ser utilizada para una aplicación final, se considera como un problema mal definido. Es un problema mal definido ya que existen múltiples soluciones correctas para cada imagen [122]. A través de los años, muchos investigadores se han centrado en proponer algoritmos de segmentación de imágenes sin una aplicación final cada vez más eficientes, robustos y confiables. Estos algoritmos tienen la finalidad de ser utilizados en múltiples aplicaciones, sin embargo, no se ha realizado mucha investigación para proponer nuevas maneras de evaluar la segmentación de imágenes [124]. Para el caso de la evaluación de la segmentación como un problema mal definido, se utiliza una evaluación directa. La evaluación directa consiste en medir el desempeño del algoritmo comparándolo con múltiples segmentaciones humanas. Por su parte, cuando la segmentación de imágenes es evaluada de acuerdo con su desempeño en una aplicación final, se le llama una evaluación indirecta [127]. Este capítulo se centra en la evaluación directa de los algoritmos de segmentación de imágenes.

Uno de los problemas principales de las métricas existentes de evaluación directa en el estado del arte, es que inducen a los investigadores a buscar el desempeño más alto, sin tomar en cuenta que es prácticamente imposible obtener un desempeño perfecto en las métricas. Esto es debido a la comparación con múltiples *Ground Truth* (GT) generados por humanos para cada imagen. Por su parte, muchas de las métricas de evaluación carecen de una interpretación objetiva debido a los rangos de trabajo que tienen y que algunas no cuentan con límites superiores. Además, la técnica que utilizan las métricas para tomar en cuenta la variación del criterio humano no siempre es la más adecuada. Cuando se evalúan directamente los algoritmos de segmentación, la variación del criterio humano es tomada en cuenta mediante la obtención de un promedio contra múltiples GT por imagen. Sin embargo, esto implica que se pierda la generalización del comportamiento humano al segmentar imágenes, ya que la cantidad de humanos que generan los GT puede ser muy pequeña para representar todas esas múltiples

soluciones correctas que puede tener una imagen [127]. Lo cual, lleva a la necesidad de un nuevo enfoque de evaluación de la segmentación de imágenes. Este enfoque necesita generalizar las variaciones en las segmentaciones realizadas por humanos para cada imagen independiente, de acuerdo con una aproximación de la cantidad de soluciones correctas que puede tener una imagen. Esto, con la finalidad de determinar de una manera más adecuada la calidad de una segmentación hecha por un algoritmo de acuerdo con el criterio humano.

Debido a la gran problemática y subjetividad que representa la evaluación directa de la segmentación de imágenes, este capítulo se centra en proponer un enfoque difuso basado en evaluación directa, cuya contribución se divide en dos partes principales. La primera se enfoca en fusificar las métricas de evaluación de segmentación de imágenes del estado del arte. Esto se logra tomando en cuenta el rango de desempeño humano generalizado por imagen en función de la respuesta a la complejidad de imagen, RCI. Esta fusificación facilitará la interpretación de cada métrica y permitirá observar el grado de pertenencia respecto a lo que se considera desempeño humano. La segunda, propone una agregación de métricas difusas para saber si la segmentación que se evalúa es correcta o no, de acuerdo con su consistencia al criterio humano. Los resultados obtenidos demuestran que el enfoque difuso propuesto de evaluación de segmentación de imágenes es robusto, confiable y fácil de interpretar.

La estructura de este capítulo es de la siguiente manera. La Sección 5.1 muestra las desventajas de las métricas de evaluación existentes y la variación del criterio humano. La Sección 5.2 presenta la metodología del enfoque de evaluación difuso propuesto. La Sección 5.3 contiene los resultados y la Sección 5.4 las conclusiones obtenidas en este capítulo.

5.1 Métricas de evaluación y variación del criterio humano

De acuerdo con el estado del arte sobre segmentación de imágenes, se puede decir que una segmentación correcta es aquella que tiene buena consistencia con segmentaciones humanas [127]. La mayoría de las métricas de evaluación de segmentación de imágenes, utilizan un esquema rígido cuyo objetivo es identificar cuando dos segmentaciones son consistentes, considerando que la mejor segmentación para cierta imagen es cuando se obtiene el más alto desempeño en las métricas. Sin embargo, es prácticamente imposible conseguir ese nivel de

desempeño por la variación implícita en las segmentaciones humanas. Por lo tanto, utilizar un esquema rígido en las métricas es poco realista para conocer el grado de desempeño alcanzado por algún algoritmo de segmentación de imágenes. Por ejemplo, es posible que un algoritmo obtenga resultados de segmentación similares a los de una persona que realizó un GT con altos índices de desempeño en las métricas, pero si se compara contra las segmentaciones de otros humanos su desempeño no estaría garantizado.

Por lo tanto, en las siguientes secciones, se presentan las desventajas de las métricas para la evaluación de la segmentación y el por qué deben de ser procesadas mediante el enfoque difuso propuesto, se analiza la variación de las métricas al evaluar segmentaciones humanas y se describe la motivación del trabajo en utilizar la RCI para generalizar el rango de variación humana en cada una de las métricas.

5.1.1 Métricas de evaluación

Las cuatro métricas descritas en la Sección 1.2.1.2, presentan características interesantes, donde dependiendo de la aplicación, algunas métricas son mejores que otras para realizar la evaluación de la segmentación de imágenes [124]. Sin embargo, cuando son utilizadas en la evaluación directa presentan algunas desventajas. Una de las desventajas del PRI es que tiene un pequeño rango dinámico, lo que provoca que los valores obtenidos en la evaluación de las segmentaciones sean comúnmente similares [124]. Además, el PRI es conocido por tener una mayor sensibilidad que especificidad. lo que implica que no se obtienen valores de 0 para evaluaciones donde S y GT son máximamente diferentes, y comúnmente se obtienen valores altos aunque se evalúen segmentaciones muy diferentes al GT [128]. Por su parte, la interpretación de la métrica VI carece de significado cuando se realiza una evaluación con múltiples GT [92], como en el caso de la evaluación directa, además, su medición no es acertada cuando se trabaja con imágenes segmentadas que contienen una cantidad pequeña de regiones [129]. Al igual que VI , GCE tampoco realiza una evaluación acertada cuando S o GT tienen una cantidad pequeña de regiones. Además, no castiga la sobre segmentación de las imágenes [124]. Aunado a esto, GCE carece de significado perceptual cuando S contiene una etiqueta diferente para cada pixel o cuando todos los pixeles tienen la misma etiqueta, ya que en estos casos se

obtiene un valor de 0, indicando que la segmentación en términos de GCE es perfecta. En cuanto a BDE, una de sus principales desventajas es que tiene sesgo hacia la sobre segmentación y no cuenta con un límite superior.

De acuerdo con lo anterior, estas métricas presentan diversas situaciones que vuelven complicada la evaluación directa de la segmentación de imágenes, debido principalmente a su difícil interpretación por carencia de límites superiores. Por lo cual, es necesario mejorar la manera en la que se lleva a cabo la evaluación directa, facilitando la interpretación de las métricas para que se utilicen de una manera más intuitiva. El enfoque de evaluación directa de segmentación de imágenes propuesto mediante la fusificación de las métricas resuelve el problema de interpretación de las métricas y mediante su agregación ayuda a determinar de una manera intuitiva el desempeño de los algoritmos de segmentación de imágenes.

5.1.2 La RCI y la variación en el criterio humano al segmentar imágenes

A continuación, se analiza la variación que existe entre las segmentaciones realizadas por humanos en términos de las métricas $M = \{PRI, VI, GCE, BDE\}$ y en función de la complejidad de imagen. Esto, con la finalidad de generalizar un rango de desempeño humano y así poder determinar cuándo un algoritmo cumple correctamente con la tarea de segmentación, dicho análisis se presenta a continuación.

El análisis realizado consiste en dos puntos principales. El primero se enfoca en observar la variación de la segmentación realizada por los humanos para cada imagen en cada una de las métricas. La variación humana para cada una de las métricas es representada por V_{Mi} . El segundo punto estudia cómo V_{Mi} varía en función de la respuesta a la complejidad de imagen Ψ . La complejidad de imagen se utiliza porque es una medida relativa y subjetiva de la cantidad de información semántica que puede ser percibida mediante el sentido de la vista en una imagen, para un conjunto determinado de imágenes. Por lo tanto, puede ser directamente relacionada con la cantidad de interpretaciones que puede tener una imagen. Por ejemplo, una imagen con complejidad baja $\Psi \downarrow$ tiende a tener menor cantidad de interpretaciones (soluciones correctas a la tarea de segmentación) que una imagen con complejidad alta $\Psi \uparrow$. En la Figura 5.1a se muestra una imagen con $\Psi \downarrow$, así como algunos de sus GT generados por humanos de BSDS500 para esta

imagen, se puede inferir que para esta imagen existe una baja cantidad de interpretaciones o soluciones correctas. Por otro lado, en el caso de la imagen en la Figura 5.1b con $\Psi \uparrow$, el número de soluciones correctas incrementa considerablemente, desde realizar una segmentación por conjuntos de objetos hasta el caso de asignar una región diferente a cada uno de los pimientos de la imagen.

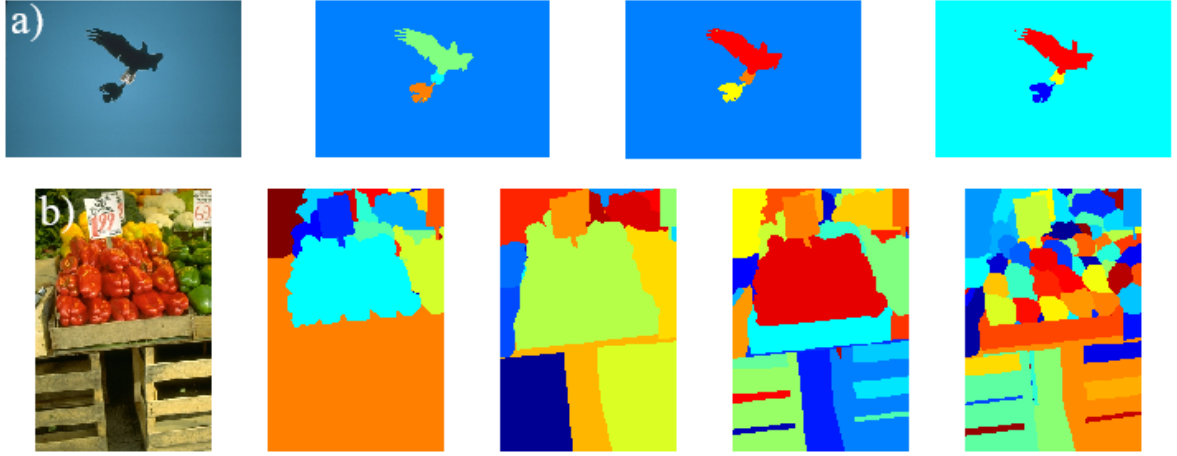


Figura 5.1 a) Imagen de baja complejidad con sus respectivos GT, b) Imagen de alta complejidad con sus respectivos GT.

Para el análisis de V_{Mi} en función de Ψ se utilizó la base de datos BSDS500 que cuenta con K número de GT por cada imagen I_i generados por humanos. El promedio de K en BSDS500 es de 5 GT por imagen. V_{Mi} fue obtenido para cada I_i de BSDS500 mediante un promedio del valor obtenido por cada M entre todos los GT de una imagen, es decir:

$$V_{Mi} = \frac{M \left(GT_{i\varpi, \hat{e}}, GT_{i\varpi, \ddot{e}} \right)}{K_i} \quad \hat{e} \neq \ddot{e} \quad (5.1)$$

donde $\hat{e}$ y $\ddot{e}$ son índices que representan personas que realizaron el GT y ϖ es el índice del número de GT generados para una imagen i . Por lo tanto, V_{Mi} indica la variación que existe en los GT generados por los humanos para cierta métrica M por cada imagen i .

Para encontrar la tendencia de V_{Mi} en función de Ψ , los resultados fueron ordenados de menor a mayor Ψ y luego procesados con una función de suavizado. Este resultado se muestra en la Figura 5.2 para las cuatro métricas en cuestión.

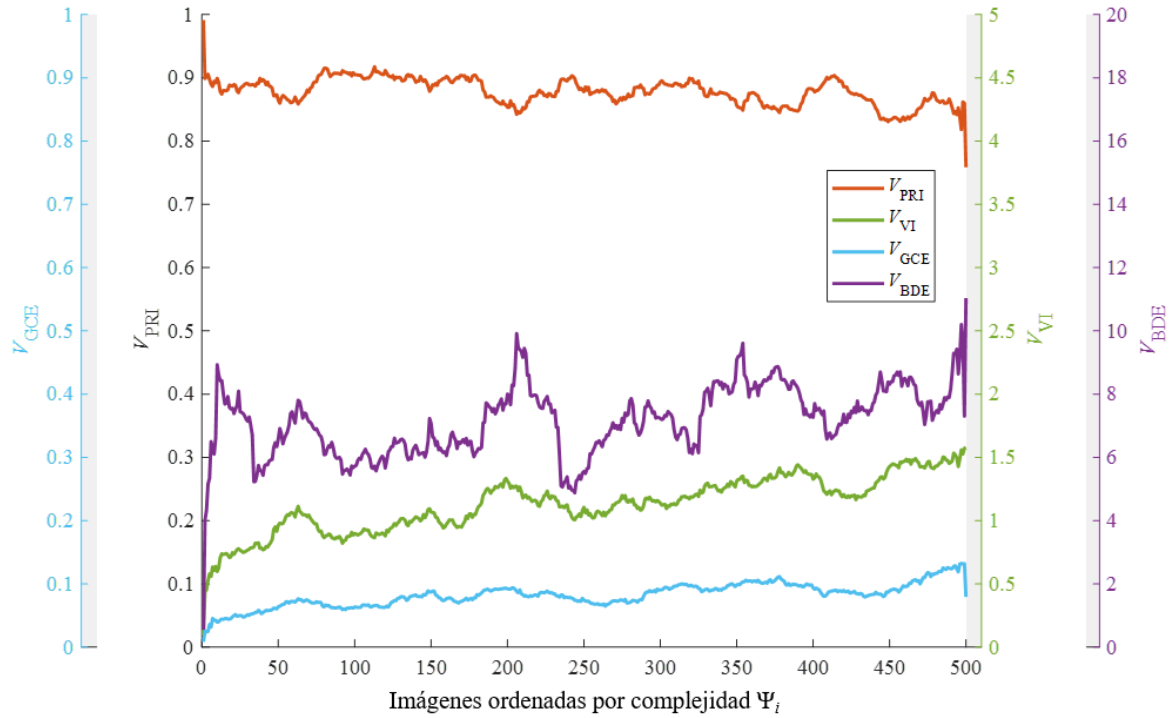


Figura 5.2 Variación humana de acuerdo con las métricas M en función de la complejidad de imagen Ψ .

En la Figura 5.2 se pueden observar 2 tipos de comportamiento de V_{Mi} en función de Ψ . Las métricas que indican un mejor desempeño entre mayor sea su valor, son denotadas por el conjunto $M\uparrow = \{PRI\}$ y las métricas de error cuyo desempeño es mejor entre más bajo sea su valor, son denotadas por el conjunto $M\downarrow = \{VI, GCE, BDE\}$. De acuerdo con Ψ , V_{PRI} decae conforme Ψ_i incrementa. Para V_{VI} , V_{GCE} , y V_{BDE} , el valor aumenta conforme Ψ_i incrementa. Esto debido a que a mayor Ψ , mayor número de interpretaciones tiene la imagen segmentada, y por lo tanto se genera una mayor variación en las segmentaciones humanas que impacta en el decaimiento del valor de la métrica PRI y el aumento del error de las métricas VI, GCE y BDE.

Como se observa en la Figura 5.2, las métricas VI, GCE y BDE no tienen un límite superior, por lo que su interpretación y comparación se vuelve complicada. Uno de los objetivos de este

trabajo es otorgar un significado a las métricas de evaluación de los algoritmos de segmentación de imágenes para que puedan ser interpretadas con mayor facilidad. La significancia aquí propuesta para las métricas M , radica en analizar y evaluar el grado de consistencia de los valores de las métricas respecto al desempeño humano, en función de la RCI. La RCI es un parámetro que permite establecer una razón de cambio del número posible de interpretaciones de segmentación y el rango de variación humana en la generación de GT. De manera que, a mayor complejidad de imagen, mayor cantidad de interpretaciones, mayor variación humana y por ende más difícil obtener un desempeño perfecto en las métricas. Por ello, en este trabajo se propone hacer un cambio en la forma en la que se realizan las evaluaciones de algoritmos de segmentación, para una evaluación más consistente que le exija un mayor desempeño a los algoritmos de segmentación cuando se trabaja con imágenes simples que con imágenes complejas.

5.2 Metodología del enfoque difuso propuesto

El enfoque difuso para la evaluación de la segmentación de imágenes en función de la RCI se realizó mediante las cuatro métricas más populares de BSDS500: PRI, VI, GCE y BDE. Se sabe que existen otras métricas mejor evaluadas en términos de meta-métricas [124][130]. Sin embargo, la finalidad de este trabajo no es proponer nuevas métricas sino mejorar la interpretación y utilización de las métricas del estado del arte para poder realizar una evaluación más adecuada que facilite una correcta segmentación. Por lo tanto, consideramos que las métricas PRI, VI, GCE y BDE son suficientes para demostrar los beneficios y correcto funcionamiento del enfoque de evaluación propuesto. Además, el enfoque difuso propuesto es capaz de adoptar otras métricas existentes en la literatura de ser necesario. De esta manera, la Figura 5.3 muestra el esquema general del método propuesto, donde, cada métrica M primero se fusifica para obtener uM_H y posteriormente las métricas difusas son agregadas para obtener $u^{\omega}SEG_H$ que cuantifica la consistencia de las segmentaciones con el criterio humano.

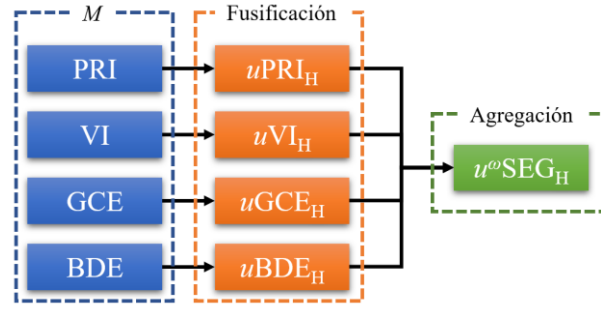


Figura 5.3 Esquema general del método de evaluación difuso propuesto.

5.2.1 Fusificación de métricas de acuerdo con la complejidad de imagen

El primer paso del enfoque de evaluación difuso propuesto se centra en determinar una función de umbral de acuerdo con Ψ y con la variación humana para cada métrica $M = \{\text{PRI}, \text{VI}, \text{GCE}, \text{BDE}\}$. Esta función de umbral representa el valor a partir del cual el valor obtenido de M en una segmentación corresponde a un valor similar al desempeño humano.

Para encontrar esta función de umbral, se realizó un ajuste por medio de regresión lineal en V_M , utilizando un filtro de media móvil. Por ejemplo, en la Figura 5.4 se muestra el ajuste lineal de V_{PRI} para la métrica PRI. Este ajuste lineal representa la tendencia de V_{PRI} en función de Ψ . Para el punto A mostrado en la Figura 5.4, se tiene $\Psi = 0.1890$ y $V_{\text{PRI}} = 0.8977$ y para el punto B, $\Psi = 0.7180$ y $V_{\text{PRI}} = 0.8587$. Por lo que, la función de umbral, VF_M , se obtiene por:

$$VF_M = \alpha_M \Psi + \beta_M \quad (5.2)$$

donde, para el caso de la métrica PRI,

$$\alpha_{\text{PRI}} = \frac{V_{\text{PRI}}^B - V_{\text{PRI}}^A}{\Psi^B - \Psi^A} = \frac{0.8587 - 0.8977}{0.7180 - 0.1890} = -0.0737 \quad (5.3)$$

y

$$\beta_{\text{PRI}} = V_{\text{PRI}}^A - \alpha_{\text{PRI}} \Psi^A = 0.9116 \quad (5.4)$$

Así, VF_M se utiliza para determinar una región R_M . Donde, a manera general, todos los valores de $M\uparrow$ que superen $VF_{M\uparrow}$ serán considerados dentro de R_M . Para el caso de $M\downarrow$, serían todos los valores que queden por debajo de $VF_{M\downarrow}$. Entonces, R_M es una región de valores de M que representa cuando una segmentación es correcta de acuerdo con M , ya que si se encuentra dentro de R_M se tendría un valor igual o mejor al desempeño humano. El procedimiento para la obtención de VF_M para las métricas VI, GCE y BDE, es el mismo que para la métrica PRI pero estas tendrán una pendiente α_M positiva.

Como no siempre se cuenta con suficientes GTs por imagen para generalizar la variación humana, R_M es definida de manera lineal, para así lograr generalizar el comportamiento de V_{Mi} para todas las imágenes utilizando como base a Ψ .

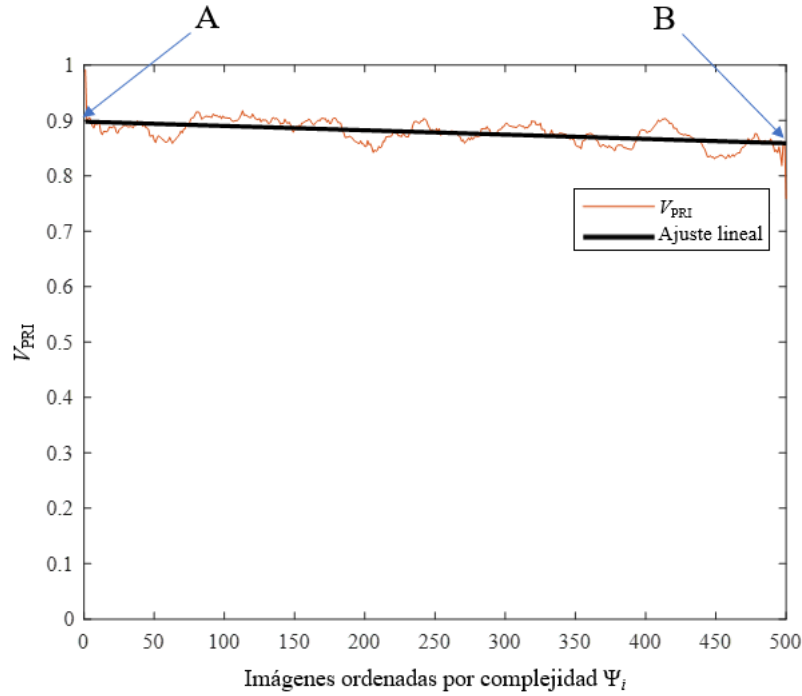


Figura 5.4 Ajuste lineal de V_{PRI} en función de Ψ .

Una vez obtenidas las VF_M y R_M , se fusifica cada métrica M , uM_H , con la finalidad de obtener un grado de pertenencia a R_M . Para el enfoque propuesto, se eligieron las funciones *S-Shaped* y *Z-Shaped* debido a su propiedades no lineales y que son comúnmente utilizadas para modelar variables lingüísticas [131]. La función *S-Shaped* es utilizada para fusificar las métricas $M\uparrow$

mientras que *Z-Shaped* es utilizada para las métricas $M_{\downarrow}$. Así, la función $uM_{\uparrow H}$ es representada por:

$$uM_{\uparrow H} = \begin{cases} 0, & M_{\uparrow} < a_M \\ 2 \left(\frac{M_{\uparrow} - a_M}{VF_{M_{\uparrow}} - a_M} \right)^2, & a_M \leq M_{\uparrow} < \frac{a_M + VF_{M_{\uparrow}}}{2} \\ 1 - 2 \left(\frac{M_{\uparrow} - VF_{M_{\uparrow}}}{VF_{M_{\uparrow}} - a_M} \right)^2, & \frac{a_M + VF_{M_{\uparrow}}}{2} \leq M_{\uparrow} < VF_{M_{\uparrow}} \\ 1, & VF_{M_{\uparrow}} \leq M_{\uparrow} \end{cases} \quad (5.5)$$

donde a_M representa el punto a partir del cual $uM_{\uparrow H} = 0$. Para el caso de $uPRI_H$, $a_{PRI} = 0$. De esta manera, en la Figura 5.5 se muestra gráficamente la función $uPRI_H$ donde se especifica la localización de los valores VF_{PRI} y a_{PRI} .

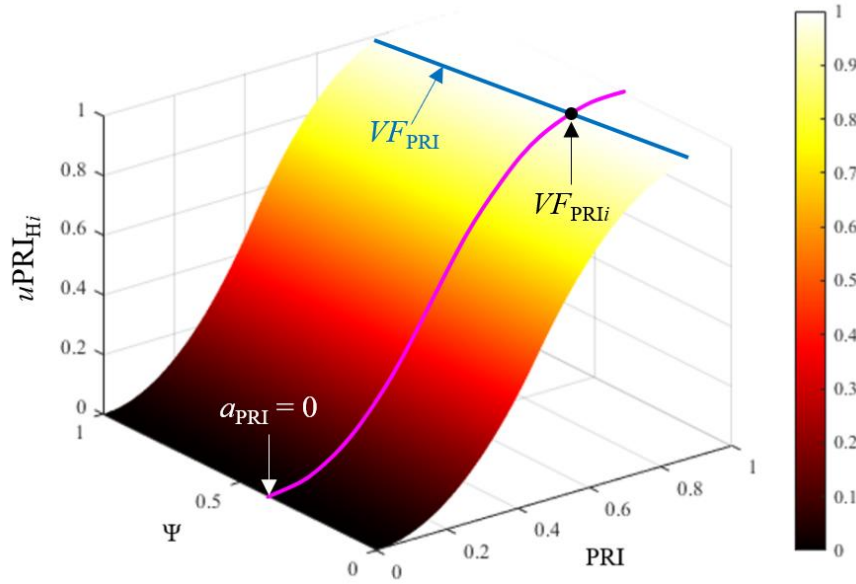


Figura 5.5 Representación de la función $uPRI_H$, donde se especifica la localización de los valores VF_{PRI} y a_{PRI} .

Las métricas $M_{\downarrow}$ difusas son representadas por $uM_{\downarrow H}$, la cual es definida mediante:

$$uM_{\downarrow H} = \left\{ \begin{array}{ll} 1, & M_{\downarrow} < VF_{M_{\downarrow}} \\ 1 - 2 \left(\frac{M_{\downarrow} - VF_{M_{\downarrow}}}{a_M - VF_{M_{\downarrow}}} \right)^2, & VF_{M_{\downarrow}} \leq M_{\downarrow} < \frac{a_M + VF_{M_{\downarrow}}}{2} \\ 2 \left(\frac{M_{\downarrow} - a_M}{a_M - VF_{M_{\downarrow}}} \right)^2, & \frac{a_M + VF_{M_{\downarrow}}}{2} \leq M_{\downarrow} < a_M \\ 0, & a_M \leq M_{\downarrow} \end{array} \right\} \quad (5.6)$$

A diferencia de $uPRI_H$ donde $a_{PRI} = 0$. Para las métricas VI, GCE y BDE se tiene $a_{VI} = 5$, $a_{GCE} = 1$, $a_{BDE} = 40$. Estos valores fueron definidos experimentalmente, escogiendo valores en los que las métricas ya no representan un desempeño aceptable. Cabe mencionar que estos valores son adecuados para trabajar con las imágenes de BSDS500 que poseen una resolución específica, ya que en algunos casos como BDE su límite superior es afectado de acuerdo con la resolución de las imágenes.

Las métricas M difusas uM_H expresan el grado en que una segmentación pertenece al rango de desempeño humano R_M de acuerdo con cada métrica M . Para ejemplificar esto, en la Figura 5.6 se muestra gráficamente las funciones uM_H . Donde se puede observar la superficie desde una vista superior de las funciones de membresía *S-Shaped* y *Z-Shaped* en función de cada métrica M y la RCI, Ψ . De acuerdo con la Figura 5.6, las métricas de evaluación se vuelven más estrictas con los algoritmos de segmentación cuando se trata de imágenes simples que con imágenes complejas. Esto es debido a que de acuerdo con $VF_{M_{\uparrow}}$, existe una menor cantidad de interpretaciones en imágenes con baja Ψ y es esperado que exista una menor variación respecto a los GTs, y conforme incrementa Ψ se tiene que ser menos exigente, ya que incluso entre humanos existe una mayor variación. Por lo tanto, mediante este enfoque de evaluación es posible determinar que tanto una segmentación realizada por un algoritmo de segmentación de imágenes se acerca o se encuentra dentro del desempeño humano de acuerdo con cada métrica. Además, este enfoque difuso facilita la interpretación de las métricas, por ejemplo, en el caso de BDE, una segmentación de $BDE = 30$ carece de significado si no se tiene conocimiento del límite superior o de la resolución de las imágenes y no se puede saber si es una segmentación

buena o mala. Por lo que, el enfoque propuesto, toma en cuenta estas variables y las utiliza para poder obtener un resultado con mayor significado, ya que $uBDE_H: S \rightarrow [0,1]$, siendo 1 el valor máximo de pertenencia al desempeño humano.

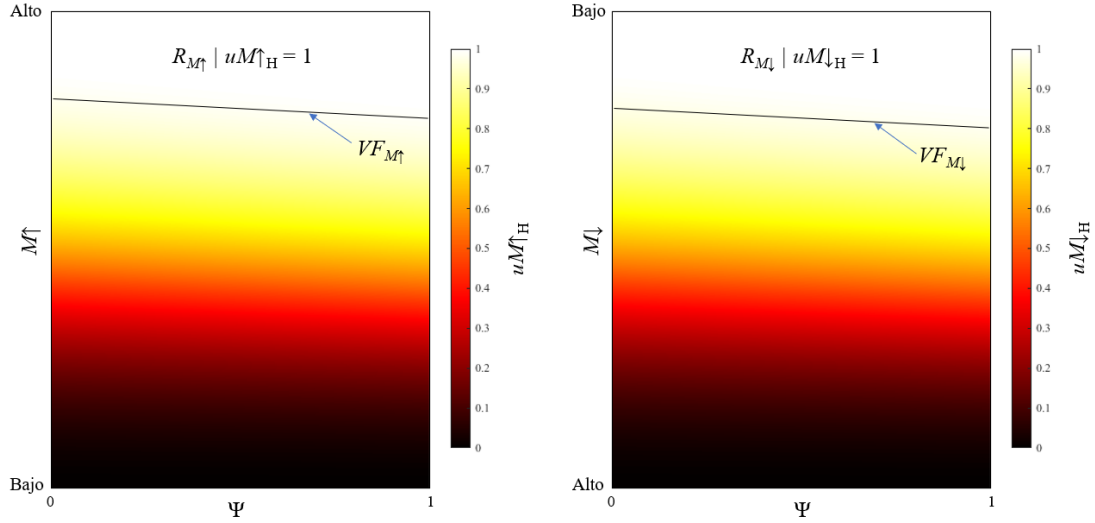


Figura 5.6 Representación gráfica de las funciones uM_H .

5.2.2 Agregación de las métricas difusas uM_H

La siguiente contribución presentada en este capítulo es un esfuerzo por contestar la pregunta, ¿Qué es una buena segmentación? [127]. Debido a que la respuesta más acertada en el estado del arte menciona que una buena segmentación es aquella que tiene una considerable consistencia con las segmentaciones realizadas por humanos, se trata entonces de cuantificar la consistencia al desempeño humano mediante la agregación de diversas métricas de evaluación. Donde una buena segmentación será aquella en la que en todas las métricas obtengan un desempeño similar al del humano.

Partiendo de la fusificación de las métricas descrita en la sección anterior, es posible realizar su agregación debido a que todas se encuentran dentro del mismo rango $[0,1]$, es decir, se tiene un grado de pertenencia al desempeño humano en cada una de las métricas M . El procedimiento de agregación de métricas se inspira en la métrica *F-Measure* [132], que utiliza una media armónica ponderada. Esta media armónica ponderada fue elegida debido a su sensibilidad con

los valores pequeños y por su flexibilidad para otorgar mayor o menor peso en la combinación de métricas, ya que según [133], las métricas tienen diferente contribución dependiendo de la base de datos y los tipos de imágenes. Debido a que el enfoque propuesto no tiene una aplicación específica en cierta base de datos o tipo de imágenes, a todas las métricas se les da el mismo peso o ponderación, el cual es representado por $\omega_M = 1$. No obstante, este valor puede ser modificado de acuerdo con las necesidades que presente la evaluación.

La función de agregación de las métricas M , es definida como:

$$u^{\omega}SEG_H = \frac{\omega_{PRI} + \omega_{VI} + \omega_{GCE} + \omega_{BDE}}{\frac{\omega_{PRI}}{uPRI_H} + \frac{\omega_{VI}}{uVI_H} + \frac{\omega_{GCE}}{uGCE_H} + \frac{\omega_{BDE}}{uBDE_H}} \quad (5.7)$$

De esta manera, se logra obtener un sólo valor en el rango $[0,1]$ que representa el grado de pertenencia a una segmentación humana, la cual puede ser considerada como una aproximación a que tan buena es la segmentación de una imagen.

5.3 Resultados

Esta sección presenta los resultados obtenidos mediante el enfoque de evaluación difuso propuesto. Primero, se comparan y analizan las evaluaciones de algoritmos de segmentación de imágenes realizadas por medio de M , uM_H y $u^{\omega}SEG_H$, Sección 5.3.1. Luego, la Sección 5.3.2 compara a M , uM_H y $u^{\omega}SEG_H$ mediante dos meta-métricas. Las meta-métricas son encargadas de medir el desempeño de las métricas de evaluación, respecto al criterio humano.

5.3.1 Evaluación de algoritmos del estado del arte mediante las diferentes métricas y enfoque propuesto

Para demostrar el correcto funcionamiento del enfoque propuesto en este capítulo, se evaluaron siete algoritmos populares del estado del arte mediante las métricas M para analizar su comportamiento. Luego, los algoritmos fueron evaluados mediante uM_H . Por último, para cada algoritmo se calculó la agregación de las métricas mediante el uso de $u^{\omega}SEG_H$, para poder observar cómo es su comportamiento al evaluar la segmentación de imágenes.

Los siete algoritmos de segmentación de imágenes utilizados fueron SFFCM [89], FRFCM [88], RSSFCA [90], MeanShift [134], MMGR, SLIC-MMGR y LSC-MMGR [91]. Estos algoritmos fueron elegidos de acuerdo con su popularidad y novedad en el estado del arte. Se utilizó la base de datos BSDS500 para obtener las segmentaciones de cada algoritmo y sus resultados. Para los algoritmos basados en agrupamiento (SFFCM, FRFCM, RSSFCA, MMGR, SLIC-MMGR y LSC-MMGR) se segmentaron todas las imágenes de la base de datos utilizando un número de grupos $2 \leq C \leq 10$ para cada imagen, y con los parámetros descritos por cada autor en los artículos donde se presentan. Luego, para cada métrica se seleccionó el mejor resultado de cada imagen. En el caso del algoritmo MeanShift, al no ser un algoritmo basado en agrupamiento se obtuvieron los resultados directamente para cada imagen de acuerdo con los parámetros utilizados en [134]. Estos resultados son mostrados en la Tabla 5.1 para cada algoritmo de segmentación y se encuentran ordenados del mejor al peor desempeño en cada métrica.

Tabla 5.1 Algoritmos de segmentación de imágenes evaluados por M y ordenados del mejor al peor desempeño en cada métrica.

Algorithm	PRI	Algorithm	VI	Algorithm	GCE	Algorithm	BDE
SFFCM	0.802	SFFCM	1.886	RSSFCA	0.034	LSC-MMGR	10.888
MMGR	0.798	MMGR	1.913	MMGR	0.107	SFFCM	11.411
LSC-MMGR	0.792	RSSFCA	1.919	SFFCM	0.133	MMGR	11.681
SLIC-MMGR	0.792	LSC-MMGR	2.108	LSC-MMGR	0.152	SLIC-MMGR	11.702
MeanShift	0.785	SLIC-MMGR	2.152	SLIC-MMGR	0.154	FRFCM	11.754
RSSFCA	0.782	FRFCM	2.304	MeanShift	0.157	MeanShift	12.808
FRFCM	0.767	MeanShift	2.731	FRFCM	0.194	RSSFCA	12.844

En la Tabla 5.1 se muestra el ranking de los algoritmos para cada métrica. Se observa que cada métrica posiciona de manera distinta a los algoritmos de segmentación de imágenes. Por ejemplo, SFFCM es posicionado en primer lugar por la métrica PRI y VI, y para las métricas GCE y BDE ocupa el tercer y segundo lugar respectivamente. Además, aunque se observa algo de consistencia en las posiciones (por ejemplo, SFFCM se mantiene en los primeros lugares y MeanShift en los últimos), también se presenta un caso crítico, donde el algoritmo RSSFCA es posicionado en primer lugar por la métrica GCE, pero en último lugar por la métrica BDE. Por lo tanto, de acuerdo con estas métricas de BSDS500 es complicado tomar una decisión de que algoritmo es mejor, o cual se adapta mejor a la aplicación que se le desee dar al algoritmo.

Una de las desventajas más importantes de utilizar las métricas M para evaluar los algoritmos de segmentación es que carecen de un significado práctico que ayude a interpretar los resultados. Por ejemplo, en la Tabla 5.1 los resultados de $GCE = 0.034$ por parte de RSSFCA y $GCE = 0.194$ de FRFCM, no significan que RSSFCA es cerca de 6 veces mejor que FRFCM o que utilizar RSSFCA garantice una buena segmentación. Otro ejemplo similar sucede con la métrica BDE, donde es bastante complicado dar un significado a los valores obtenidos. Se sabe que un valor de cero en la métrica BDE significa una segmentación perfecta, sin embargo, es prácticamente imposible alcanzar ese valor cuando se realiza una comparación con múltiples GT generados por humanos. Por lo que, al fusificar las métricas se facilita en gran medida la interpretación de estas, ya que se toma en cuenta el rango de desempeño humano y se limitan sus valores a un rango de $[0,1]$. En la Tabla 5.2 se muestra uM_H para cada algoritmo de segmentación obtenido mediante los resultados de las métricas M en función de la complejidad de imagen Ψ .

Tabla 5.2 Algoritmos de segmentación de imágenes evaluados por uM_H , ordenados del mejor al peor desempeño.

Algoritmo	$uPRI_H$	Algoritmo	uVI_H	Algoritmo	$uGCE_H$	Algoritmo	$uBDE_H$
SFFCM	0.965	SFFCM	0.885	RSSFCA	0.993	LSC_MMGR	0.925
SLIC_MMGR	0.965	MMGR	0.877	MMGR	0.983	SLIC_MMGR	0.917
LSC_MMGR	0.964	RSSFCA	0.867	SFFCM	0.980	SFFCM	0.914
MMGR	0.964	LSC_MMGR	0.816	MeanShift	0.977	MMGR	0.908
RSSFCA	0.957	SLIC_MMGR	0.805	SLIC_MMGR	0.973	FRFCM	0.905
FRFCM	0.946	FRFCM	0.760	LSC_MMGR	0.971	RSSFCA	0.895
MeanShift	0.937	MeanShift	0.611	FRFCM	0.947	MeanShift	0.885

Usando de ejemplo los mismos algoritmos RSSFCA y FRFCM discutidos en la Tabla 5.1, ahora en la Tabla 5.2, RSSFCA tiene un $uGCE_H = 0.993$, que representa un valor muy cercano al desempeño humano en términos de GCE. Por otro lado, FRFCM obtiene un $uGCE_H = 0.947$, disminuyendo cerca de un 5% en desempeño contra RSSFCA, no como en el caso de utilizar M directamente que se observaba una diferencia casi seis veces mayor.

Un punto para destacar cuando se utiliza uM_H , es que el lugar de los algoritmos en el ranking varía en comparación a M . Por ejemplo, entre PRI y $uPRI_H$ cinco algoritmos cambiaron su posición, sólo SFFCM y LSC_MMGR se mantuvieron en el mismo lugar. Un efecto similar se observa entre GCE y $uGCE_H$, y entre BDE y $uBDE_H$. Esta variación es debida a la inclusión de

Ψ para obtener uM_H . Recodando que la variación en el desempeño humano se incrementa conforme Ψ incrementa. Lo cual es modelado mediante la función VF_M .

Para lograr una evaluación con significado, realista y objetiva que facilite la selección de los algoritmos de segmentación de imágenes para resolver una tarea específica, en la Tabla 5.3 se muestra el ranking generado por $u^{\omega}SEG_H$ de los algoritmos de segmentación de imágenes, utilizando $\omega_M = 1$. El valor obtenido por $u^{\omega}SEG_H$ puede ser interpretado como una aproximación a que tan correcta es la segmentación de una imagen considerando el criterio de segmentación humano, ya que este valor se refiere a un grado de pertenencia a la región R_M . Además, $u^{\omega}SEG_H$ se puede extender para combinar más métricas del estado del arte y cada una de ellas puede ser ponderada de acuerdo con la aplicación que realizará el algoritmo de segmentación de imágenes.

Es importante resaltar que el ranking generado en la Tabla 5.3 coincide con el ranking de la métrica VI mostrado en la Tabla 5.2. Sin embargo, esto ocurre únicamente para el caso particular de la base de datos BSDS500, no es un comportamiento generalizado. Lo anterior, se abordará en la Sección 5.3.2 mediante la evaluación de $u^{\omega}SEG_H$ por medio de meta-métricas para demostrar su robustez y confiabilidad sobre las métricas M utilizando otra base de datos.

Tabla 5.3 Algoritmos de segmentación de imágenes evaluados por $u^{\omega}SEG_H$, utilizando $\omega_M = 1$, ordenados del mejor al peor desempeño

Algoritmo	$u^{\omega}SEG_H$
SFFCM	0.894
MMGR	0.884
RSSFCA	0.863
LSC_MMGR	0.855
SLIC_MMGR	0.841
FRFCM	0.810
MeanShift	0.756

La Figura 5.7 muestra un ejemplo de evaluación de dos segmentaciones de una imagen de BSDS500, Figura 5.7a, realizadas por un algoritmo con diferentes parámetros para demostrar el uso del enfoque propuesto. La primera segmentación, Figura 5.7b, se considera correcta de acuerdo con el criterio humano. Por su parte, la segunda segmentación, Figura 5.7c, presenta un error que difícilmente cometería un humano, donde una región uniforme (fondo de la imagen) es fragmentada en dos regiones. Para las segmentaciones de la Figura 5.7b y Figura 5.7c se muestran los resultados obtenidos por las métricas M , las métricas difusas uM_H y la agregación

de las métricas difusas $u^{\omega}SEG_H$. Si las segmentaciones de la Figura 5.7b y Figura 5.7c son comparadas sólo utilizando M , el resultado de la Figura 5.7b sería una mejor segmentación que el de la Figura 5.7c. Aunque resultaría complicado interpretar esos valores de M para llegar a esa conclusión, y más complicado saber en qué grado es mejor. Esto es debido principalmente a la carencia de límites superiores en VI, GCE y BDE. En cambio, cuando las segmentaciones son comparadas con las métricas difusas uM_H propuestas, es más sencillo determinar que la segmentación de la Figura 5.7b es mejor que la de la Figura 5.7c y en qué grado es mejor. Esto es debido a que las métricas difusas uM_H propuestas representan un grado de pertenencia al desempeño humano, donde se puede observar que la segmentación de la Figura 5.7b toda $uM_H = 1$ mientras que la segmentación de la Figura 5.7c sólo obtiene un desempeño aceptable en uVI_H y $uGCE_H$. Estas métricas resultaron altas debido a que uVI_H y $uGCE_H$ no castigan la sobre segmentación presente en el fondo [13][14].

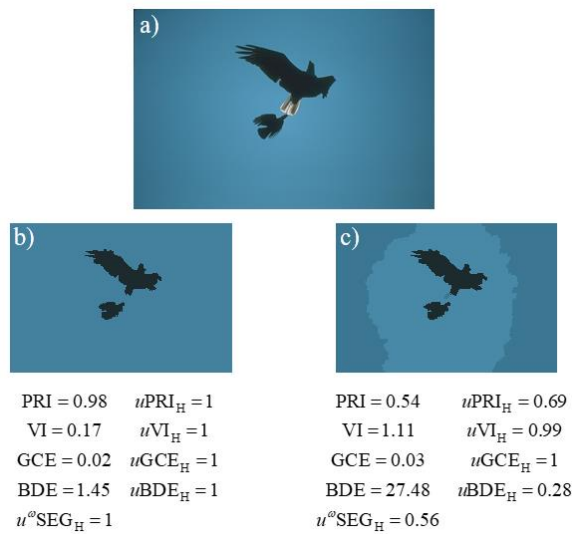


Figura 5.7 Ejemplo de evaluación de la segmentación de una imagen. a) Imagen original b) Segmentación correcta c) Segmentación incorrecta.

Para complementar la evaluación se utiliza la agregación de las métricas difusas uM_H indicada como $u^{\omega}SEG_H$. De manera que $u^{\omega}SEG_H$ representa el grado de pertenencia final con respecto al desempeño humano. $u^{\omega}SEG_H$ determina qué tan coherente es una segmentación considerando el criterio humano. Se puede observar que en la Figura 5.7b se obtiene un valor

de $u^{\omega} \text{SEG}_H = 1$ mientras que en la Figura 5.7c $u^{\omega} \text{SEG}_H = 0.56$. Por lo tanto, se puede determinar fácil e intuitivamente que la segmentación de la Figura 5.7b es mejor que la de la Figura 5.7c. Lo que indicaría que la segmentación de la Figura 5.7b es más similar a la que haría un humano y la de la Figura 5.7c sería una segmentación que difícilmente realizaría un humano.

5.3.2 Evaluación y análisis por medio de meta-métricas

Con la finalidad de realizar una evaluación cuantitativa de la mejora obtenida mediante el enfoque propuesto, en esta sección se muestran los resultados obtenidos por medio de dos meta-métricas. Una propuesta por Peng *et al.* en [129] y otra llamada SIHD (*Swapped-Image Human Discrimination*) [130]. Las meta-métricas tienen la finalidad de evaluar una hipótesis aceptada que se hace sobre los resultados de una segmentación y analizar qué tan coherente es una medición con dicha hipótesis [130].

5.3.2.1 Meta-métrica Peng

La meta-métrica propuesta por Peng *et al.* en [129] evalúa que tan coherente es una métrica con el criterio humano al decidir cuál de dos segmentaciones de una imagen es mejor [124]. Para esto, se obtiene una tasa de exactitud, la cual está dada por la cantidad de veces que una métrica toma la misma decisión que el criterio humano al comprar pares de segmentaciones entre el total de pares comparados. La base de datos utilizada por Peng *et al.* [129] consiste en dos partes: Parte_A y Parte_B. La parte Parte_A fue generada por los autores de [129] conteniendo un total de 200 imágenes, mientras que la parte Parte_B es un extracto de 300 imágenes de la base de datos BSDS500 [92]. Es importante mencionar que los GT de Parte_A fueron generados por personas distintas a las que hicieron los GT de Parte_B y con instrucciones diferentes.

De esta manera, la Tabla 5.4 presenta los resultados obtenidos en términos de la tasa de exactitud para $u^{\omega} \text{SEG}_H$ y cada métrica M . Para la Parte_A el resultado obtenido por $u^{\omega} \text{SEG}_H$ supera a los de M obteniendo un resultado de 79.50%, ya que $u^{\omega} \text{SEG}_H$ toma en cuenta la media armónica de todas las métricas difusas, lo que ayuda a que se tome una mejor decisión para esta base de datos. Para el caso de la Parte_B, $u^{\omega} \text{SEG}_H$ obtiene un resultado de 79.00% el cual es

sólo superado ligeramente por la métrica PRI. En el caso de la Parte_B el resultado de $u^{\omega}\text{SEG}_H$ es afectado ya que toma en cuenta a las métricas VI y GCE que obtienen muy bajo resultado. De acuerdo con [129], el resultado obtenido por VI y GCE es muy bajo para la Parte_B, ya que la cantidad de regiones generadas en las segmentaciones humanas es baja, por lo que estas métricas, VI y GCE se ven afectadas, y no realizan una selección correcta conforme al criterio humano. Esto demuestra la inestabilidad de VI y GCE ante la manera en que se generan los GT. Por otro lado, $u^{\omega}\text{SEG}_H$ si se mantiene estable incluso cuando se cambia la manera en la que se generan los GT. Otra ventaja del enfoque propuesto es que puede reducir el impacto que tienen las métricas en la evaluación. Por ejemplo, para este caso donde VI y GCE no son muy buena opción para evaluar la segmentación, se puede reducir el valor de ω_{VI} y ω_{GCE} , lo que permite seguir tomándolas en cuenta, pero con un menor impacto.

Tabla 5.4 Resultados obtenidos mediante la meta-métrica propuesta por Peng *et al.* [129].

Parte	$u^{\omega}\text{SEG}_H$	PRI	VI	GCE	BDE
Parte_A	79.50%	77.00%	74.50%	75.50%	78.50%
Parte_B	79.00%	80.33%	62.00%	49.33%	78.33%

5.3.2.2 Meta-métrica SIHD

La meta-métrica SIHD [130] se basa en la hipótesis de que dado un par de GT, una métrica debe de ser capaz de decir si vienen de la misma imagen con un aceptable refinamiento o de diferentes imágenes con discrepancias inaceptables. Para obtener SIHD, primero se calcula M entre todos los pares de GT seleccionados de la base de datos BSDS500. En este caso se seleccionaron 500 pares de GT correspondientes a la misma imagen y 500 pares de imágenes diferentes. Luego, para obtener el índice SIHD se utiliza un clasificador que coloca un umbral para separar los dos tipos de pares. SIHD es definido como el porcentaje de clasificaciones correctas de acuerdo con el umbral encontrado por el clasificador. En este caso se utiliza el método de Bayes Risk como clasificador.

En la Figura 5.8 se muestran las distribuciones de los pares comparados para el cálculo de SIHD en las métricas M . Las gráficas muestran los valores medios de cada distribución y el umbral de decisión encontrado por el clasificador. Se puede observar que el SIHD más alto corresponde a la métrica BDE, donde VI tiene un valor muy cercano a esta. Para este caso, la

métrica PRI es la que obtuvo menor desempeño con un SIHD de 0.87, es decir, que PRI es menos capaz de distinguir cuando dos segmentaciones tienen grandes discrepancias entre sí.

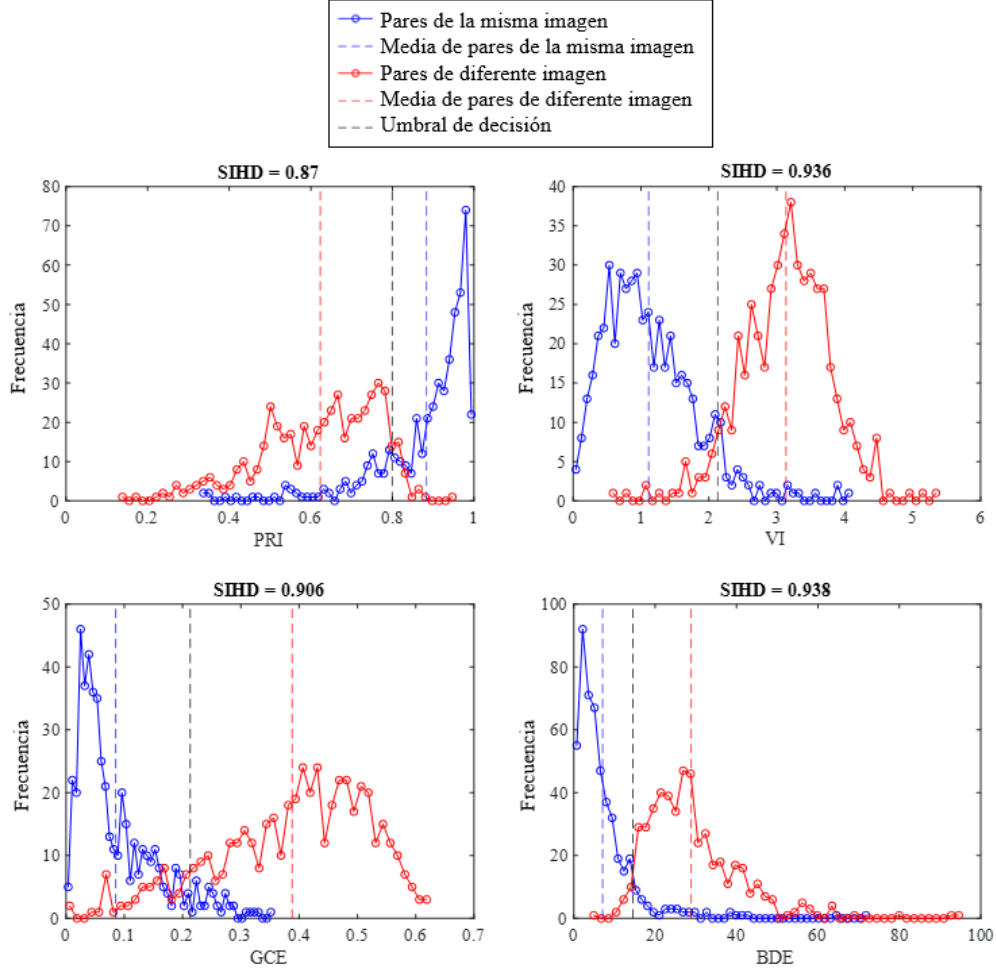


Figura 5.8 Gráficas de distribución de pares comparados por medio de M para el cálculo de SIHD.

A continuación, en la Figura 5.9 se muestra la gráfica de distribución para obtener SIHD del enfoque propuesto $u^{\omega}\text{SEG}_H$, el cual obtiene un valor de SIHD = 0.971, superando considerablemente el valor obtenido por las métricas M . En esta gráfica se observa un comportamiento muy diferente a las de M , esto es debido a que una gran cantidad de pares provenientes de la misma imagen obtienen un desempeño de 1. Esto es esperado ya que se están comparando GT de la misma imagen realizados por humanos, por lo tanto, los valores obtenidos se encuentran dentro del rango de variación humano. Por otro lado, el valor medio de $u^{\omega}\text{SEG}_H$

para los pares de GT que provienen de diferentes imágenes se encuentra cerca de $u^o\text{SEG}_H = 0.6$. Estos resultados indican que el enfoque propuesto es capaz de generar una mejor discriminación que las métricas M , sobre si los pares comparados provienen o no de una misma imagen.

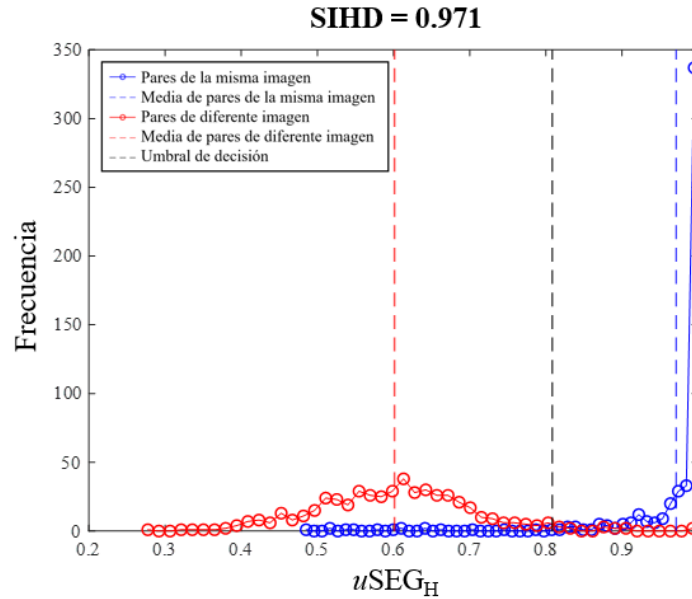


Figura 5.9 Gráfica de distribución pares comparados por medio de $u^o\text{SEG}_H$ para el cálculo de SIHD.

Debido a que el enfoque propuesto se centra en realizar una evaluación de la segmentación de imágenes basándose en diversas métricas, es capaz de realizar una evaluación robusta, ya que toma en cuenta diferentes aspectos de las segmentaciones mediante la agregación de métricas, algo a lo que es muy difícil llegar utilizando cada métrica por separado. Al agregar las métricas difusas uM_H se facilita en gran parte la interpretación, ya que, incorpora conocimiento sobre el rango de trabajo de cada una de las métricas M y el rango de desempeño humano. Además, cuando algunas métricas no son tan eficientes para evaluar cierta base de datos, el enfoque propuesto permite la ponderación de las métricas, pudiendo ser modificadas de acuerdo con el problema que se desee resolver mediante los algoritmos de segmentación imágenes.

5.4 Conclusiones

Este capítulo presentó un nuevo enfoque difuso basado en la complejidad de imagen para evaluar la tarea de segmentación de imágenes. Primero, se fusificó cada métrica para obtener

funciones de membresía al desempeño humano de acuerdo con un rango de desempeño humano. Luego, se propuso una función de agregación de las métricas difusas utilizando la media armónica ponderada. Esto, con la finalidad de obtener una aproximación a que tan consistente es una segmentación con el criterio humano.

Los resultados obtenidos demostraron lo siguiente para el enfoque propuesto. El enfoque propuesto otorga un mejor significado a las métricas. Facilita la interpretación de las métricas. Permite la ponderación de cada métrica para poder decidir cuales tendrán mayor contribución en el resultado final. Combate el problema de la carencia de límites superiores en algunas métricas y resuelve la falta de estandarización en sus rangos.

Además, la evaluación por medio de meta-métricas demostró la robustez y confiabilidad del enfoque propuesto para trabajar con distintas bases de datos generadas por personas e instrucciones diferentes. Esto, debido a que obtuvo un desempeño consistente de 0.795 y 0.79 en la meta-métrica propuesta por Peng para la base de datos Parte_A y Parte_B, respectivamente. Por su parte, en la meta-métrica SIHD se obtuvo un desempeño de 0.971 superando considerablemente el valor de las métricas del estado del arte PRI, VI, GCE y BDE. Por lo tanto, el enfoque propuesto establece una mejora considerable respecto a la forma actual de evaluar la segmentación de imágenes por medio de métricas, ya que resuelve algunos de sus problemas y ayuda a tomar una mejor decisión sobre que algoritmos son idóneos para resolver algún problema específico, considerando que tan cercanos están sus desempeños al del humano.

CAPÍTULO 6. RESULTADOS Y CONCLUSIONES GENERALES

Esta tesis presentó un estudio para determinar la respuesta a la complejidad de imagen y su aplicación en los algoritmos de segmentación de imágenes. Primero, se describieron los métodos existentes en la literatura para determinar la respuesta a la complejidad de imagen, RCI. Estos métodos fueron analizados y clasificados de acuerdo con su enfoque, computacional o humano. Luego, se presentaron las dos bases de datos relacionadas a la complejidad de imagen utilizadas en este trabajo de tesis, RSIVL y SAVOIAS. Después, se describió la tarea de segmentación de imágenes y como es llevada a cabo por medio de algoritmos de agrupamiento difuso. Además, se señalaron algunos de los algoritmos de segmentación de imágenes utilizados en las aplicaciones de la complejidad de imagen y la base de datos BSDS500, una de las bases de datos más populares para la evaluación de los algoritmos de segmentación de imágenes.

Una vez realizado el análisis del estado del arte sobre la RCI, se propuso una nueva definición de la RCI de manera conceptual y operacional, debido a la falta de una definición estándar aceptada en la literatura. La RCI fue definida como una medida relativa y subjetiva de la cantidad de información semántica percibida mediante el sentido de la vista, dentro de un conjunto determinado de imágenes. El resultado de proponer esta nueva definición en su forma conceptual y operacional fue tener una base sólida para el trabajo de investigación. Además, se presentó un análisis de la subjetividad del criterio humano ante la complejidad de imagen. Este análisis sirvió como preprocesamiento de la base de datos RSIVL, donde se descartaron un total de 14 evaluadores de la base de datos debido a sus evaluaciones incongruentes de la RCI. También, algunas evaluaciones por imagen de los sujetos restantes fueron descartadas debido a que fueron consideradas atípicas. Este análisis tuvo la finalidad de asegurar que los datos fueran confiables y pudieran ser utilizados para el diseño de dos algoritmos que determinan la RCI.

Después de proponer una nueva definición de la RCI y preprocesar la información de las bases de datos, se desarrollaron dos métodos computacionales para estimar la RCI. Uno basado en lógica difusa y otro en redes neuronales. Estos algoritmos simulan como el humano mapea del espacio multidimensional de la percepción de la complejidad de imagen al espacio unidimensional de la respuesta a la complejidad de imagen. Lo anterior, mediante la extracción

de 11 características visuales simples (contraste, correlación, energía, homogeneidad, factor de frecuencia, densidad de bordes, grado de compresión, número de regiones, coloración, número de colores y armonía de colores) que luego son procesadas por el algoritmo PCA para así mapearlas al espacio unidimensional de la RCI. En el caso del algoritmo basado en lógica difusa se obtuvo un resultado en términos del coeficiente de correlación con el criterio humano de 0.857 al ser probado en la base de datos RSIVL. Por su parte, el algoritmo basado en redes neuronales se correlación con el criterio humano de la base de datos RSIVL con un coeficiente de 0.871. Al ser mejor el resultado obtenido por la red neuronal, se decidió utilizarla para trabajos posteriores, por lo que también fue evaluada en la base de datos SAVOIAS obtenido un coeficiente de correlación promedio en todas sus categorías de 0.833. Mediante la comparación con otros métodos del estado del arte, los dos métodos propuestos demostraron ser superiores en términos del coeficiente de correlación con el criterio humano.

Referente al uso de la RCI en los algoritmos de segmentación, se propusieron dos aplicaciones. La primera aplicación consistió en determinar automáticamente el número de grupos necesario para segmentar una imagen mediante algoritmos de agrupamiento difuso. Este método fue nombrado AENCIC. AENCIC relaciona directamente la complejidad de imagen con el número de grupos necesarios para segmentar una imagen de una manera óptima. AENCIC ayudó a alcanzar un poco más del 94% de la eficiencia de los algoritmos de segmentación por agrupamiento difuso de acuerdo con el número de grupos, mientras que al utilizar otros algoritmos como AFCF sólo se logra una eficiencia en los algoritmos de segmentación cerca del 90%, estos resultados fueron mostrados en la Tabla 4.4. Esta mejora fue posible ya que AENCIC imita el criterio humano y ayuda a lidiar con la variabilidad de las segmentaciones generadas por humanos.

La segunda aplicación consistió en un nuevo enfoque difuso para la evaluación de la segmentación de imágenes. El cual, juzga a los algoritmos de una manera estricta cuando se trata de imágenes de baja complejidad y relaja la evaluación cuando se trata de imágenes de alta complejidad. Esto debido a que las imágenes de baja complejidad no tienen muchas interpretaciones semánticas. Por su parte, cuando se trata de imágenes de alta complejidad, normalmente tienen muchas interpretaciones semánticas, por lo que su evaluación debe ser

menos castigada ya que es un trabajo en el cual los humanos cometen más errores. El correcto funcionamiento del enfoque de evaluación difuso propuesto obtuvo buenos resultados al realizar evaluaciones de diferentes algoritmos de segmentación del estado del arte. Obtuvo un desempeño de 0.795 y 0.971 al ser evaluado por las meta-métricas Peng y SIHD respectivamente. Estos resultados superan a los obtenidos por las métricas más populares del estado del arte; PRI, VI, GCE y BDE. De esta manera, el enfoque propuesto establece una mejora considerable respecto a la forma actual de evaluar la segmentación de imágenes por medio de métricas, ya que resuelve algunos de sus problemas y ayuda a tomar una mejor decisión sobre que algoritmos son idóneos para resolver algún problema específico, considerando que tan cercanos están sus desempeños al del humano.

REFERENCIAS

- [1] M. Yamaguchi and R. W. Proctor, "Multidimensional vector model of stimulus-response compatibility," *Psychol. Rev.*, vol. 119, no. 2, pp. 272–303, 2012, doi: 10.1037/a0026620.
- [2] C. H. Hsu, C. Y. Lee, and A. Marantz, "Effects of visual complexity and sublexical information in the occipitotemporal cortex in the reading of Chinese phonograms: A single-trial analysis with MEG," *Brain Lang.*, vol. 117, no. 1, pp. 1–11, 2011, doi: 10.1016/j.bandl.2010.10.002.
- [3] A. Aksentijevic, A. Mihailovic, and D. T. Mihailovic, "Time for change: Implementation of Aksentijevic-Gibson complexity in psychology," *Symmetry (Basel)*, vol. 12, no. 6, pp. 1–32, 2020, doi: 10.3390/SYM12060948.
- [4] L. Dai, K. Zhang, X. S. Zheng, R. R. Martin, Y. Li, and J. Yu, "Visual complexity of shapes: a hierarchical perceptual learning model," *Vis. Comput.*, vol. 38, no. 2, pp. 419–432, 2022, doi: 10.1007/s00371-020-02023-z.
- [5] J. Rossignac, "Shape complexity," *Vis. Comput.*, vol. 21, no. 12, pp. 985–996, 2005, doi: 10.1007/s00371-005-0362-7.
- [6] J. G. Snodgrass and M. Vanderwart, "A standardised set of 260 pictures: normal for name agreement, familiarity and visual complexity," *J. Exp. Psychol. Hum. Learn. Mem.*, vol. 6, no. 2, pp. 174–215, 1980.
- [7] V. Chikhman, V. Bondarko, M. Danilova, A. Goluzina, and Y. Shelepin, "Complexity of images: experimental and computational estimates compared," *Perception*, vol. 41, no. 6, pp. 631–647, 2012, doi: 10.1068/p6987.
- [8] M. I. Chacón, A. D. Corral S., and R. Sandoval R., "A Fuzzy Approach on Image Complexity Measure," *Comput. y Sist.*, vol. 10, no. 3, pp. 268–284, 2007, [Online]. Available: <http://www.scielo.org.mx/pdf/cys/v10n3/v10n3a6.pdf>
- [9] J. Huo, "An image complexity measurement algorithm with visual memory capacity and an EEG study," in *SAI Computing Conference*, 2016, pp. 264–268. doi: 10.1109/SAI.2016.7555993.
- [10] C. Heaps and S. Handel, "Similarity and features of natural textures," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 25, no. 2, pp. 299–320, 1999, doi: 10.1037/0096-

- 1523.25.2.299.
- [11] G. Ciocca, S. Corchs, and F. Gasparini, “Complexity Perception of Texture Images,” in *New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops*, 2015, vol. 9281, pp. 119–126. doi: 10.1007/978-3-319-23222-5.
 - [12] M. I. Chacón and A. D. Corral S., “Image complexity measure: A human criterion free approach,” in *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, 2005, pp. 241–246. doi: 10.1109/NAFIPS.2005.1548541.
 - [13] S. H. Yoon, J. Lim, and Y. G. Ji, “Assessment model for perceived visual complexity of automotive instrument cluster,” *Appl. Ergon.*, vol. 46, no. Part A, pp. 76–83, 2015, doi: 10.1016/j.apergo.2014.07.005.
 - [14] R. Xu *et al.*, “ApproxNet: Content and Contention-Aware Video Analytics System for Embedded Clients,” *Trans. Sens. Networks*, vol. 1, no. 1, pp. 1–27, 2021.
 - [15] E. Saraee, M. Jalal, and M. Betke, “Visual complexity analysis using deep intermediate-layer features,” *Comput. Vis. Image Underst.*, vol. 195, no. April, pp. 1–17, 2020, doi: 10.1016/j.cviu.2020.102949.
 - [16] S. Wang, C. Xue, J. Zhang, and J. Shao, “Interface Design of GIS System Based on Visual Complexity,” in *International Conference on Applied Human Factors and Ergonomics*, 2019, vol. 972, pp. 713–722. doi: 10.1007/978-3-319-60492-3.
 - [17] R. Grover, D. K. Yadav, D. K. Chauhan, and S. Kamya, “Adaptive Steganography via Image Complexity Analysis using 3D Color Texture Feature,” in *2018 3rd International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH)*, 2018, vol. 3, pp. 125–129. doi: 10.1109/CIPECH.2018.8724225.
 - [18] D. E. Schoth, J. Wu, J. Zhang, X. Guo, and C. Liossi, “Eye-movement behaviours when viewing real-world pain-related images,” *Eur. J. Pain (United Kingdom)*, vol. 23, no. 5, pp. 945–956, 2019, doi: 10.1002/ejp.1363.
 - [19] W. Xiaotian, M. Wanchao, Z. Kai, L. Shaoyi, and Y. Jie, “Complexity Estimation of Infrared Image Sequence for Automatic Target Track,” *J. Northwest. Polytech. Univ.*, vol. 37, no. 4, pp. 664–672, 2019, doi: 10.1051/jnwpu/20193740664.
 - [20] S. Mishra, P. Liang, A. Czajka, D. Z. Chen, and X. S. Hu, “CC-NET: Image complexity

- guided network compression for biomedical image segmentation,” in *International Symposium on Biomedical Imaging*, 2019, pp. 57–60. doi: 10.1109/ISBI.2019.8759448.
- [21] P. Li, Y. Yang, W. Zhao, and M. Zhang, “Evaluation of image fire detection algorithms based on image complexity,” *Fire Saf. J.*, vol. 121, no. 103306, pp. 1–11, 2021, doi: 10.1016/j.firesaf.2021.103306.
- [22] B. Xiao, J. Duan, X. Liu, Y. Zhu, and H. Wang, “Evaluation of Image Complexity Based on SVOR,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 7, pp. 1–18, 2018, doi: 10.1142/S0218001418540204.
- [23] M. A. Abdelwahab, A. M. Iliyasu, and A. S. Salama, “Leveraging the Potency of CNN for Efficient Assessment of Visual Complexity of Images,” *2019 9th Int. Conf. Image Process. Theory, Tools Appl. IPTA 2019*, pp. 1–8, 2019, doi: 10.1109/IPTA.2019.8936095.
- [24] I. A. Siradjuddin, A. Sakinah, and M. K. Sophan, “Combination of feature engineering and feature learning approaches for classification on visual complexity images,” *Int. J. Innov. Comput. Inf. Control*, vol. 17, no. 3, pp. 991–1005, 2021, doi: 10.24507/ijicic.17.03.991.
- [25] S. E. Corchs, G. Ciocca, E. Bricolo, and F. Gasparini, “Predicting complexity perception of real world images,” *PLoS One*, vol. 11, no. 6, pp. 1–22, 2016, doi: 10.1371/journal.pone.0157986.
- [26] A. D. Corral S., “Modeling edge perception using fuzzy logic,” Master Thesis, 2003.
- [27] P. Machado, J. Romero, M. Nadal, A. Santos, J. Correia, and A. Carballal, “Computerized measures of visual complexity,” *Acta Psychol. (Amst.)*, vol. 160, pp. 43–57, 2015, doi: 10.1016/j.actpsy.2015.06.005.
- [28] M. Ivanovici, “A Naive Complexity Measure for Color Texture Images,” in *2017 International Symposium on Signals, Circuits and Systems (ISSCS)*, 2017, pp. 1–4.
- [29] M. I. of Technology, “Vision Texture,” 2002. <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html> (accessed Apr. 14, 2018).
- [30] J. Perki and A. Hyvarinen, “Modelling image complexity by independent component analysis, with application to content-based image retrieval,” in *Artificial Neural*

- Networks–ICANN 2009*, 2009, pp. 704–714. doi: 10.1007/978-3-642-04277-5_71.
- [31] X. Guo, T. Kurita, C. Muraki Asano, and A. Asano, “Visual complexity assessment of painting images,” in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, 2013, pp. 388–392.
- [32] K. J. Yoon and I. S. Kweon, “Color image segmentation considering the human sensitivity for color pattern variations,” in *Photonics Boston, SPIE, Page (s)*, 2001, vol. 4572, pp. 269–278. doi: 10.1117/12.444191.
- [33] J. Ning, L. Zhang, D. Zhang, and C. Wu, “Interactive image segmentation by maximal similarity based region merging,” *Pattern Recognit.*, vol. 43, no. 2, pp. 445–456, 2010, doi: 10.1016/j.patcog.2009.03.004.
- [34] J. Rigau, M. Feixas, and M. Sbert, “An Information-Theoretic Framework for Image Complexity,” in *Computational Aesthetics in Graphics, Visualization and Imaging*, 2005, pp. 177–184. doi: 10.2312/COMPAESTH/COMPAESTH05/177-184.
- [35] A. Cavalcante *et al.*, “Measuring streetscape complexity based on the statistics of local contrast and spatial frequency,” *PLoS One*, vol. 9, no. 2, pp. 1–13, 2014, doi: 10.1371/journal.pone.0087097.
- [36] M. Cardaci, V. Di Gesù, M. Petrou, and M. E. Tabacchi, “A fuzzy approach to the evaluation of image complexity,” *Fuzzy Sets Syst.*, vol. 160, no. 10, pp. 1474–1484, 2009, doi: 10.1016/j.fss.2008.11.017.
- [37] V. Di Gesù and S. Roy, “Fuzzy measures for image distance,” 2000.
- [38] T. D. Pham and H. Yan, “A regularity statistic for images,” *Chaos, Solitons and Fractals*, vol. 106, no. 1, pp. 227–232, 2018, doi: 10.1016/j.chaos.2017.11.033.
- [39] A. Forsythe, M. Nadal, N. Sheehy, C. J. Cela-Conde, and M. Sawey, “Predicting beauty: Fractal dimension and visual complexity in art,” *Br. J. Psychol.*, vol. 102, no. 1, pp. 49–70, 2011, doi: 10.1348/000712610X498958.
- [40] E. Saraee, M. Jalal, and M. Betke, “SAVOIAS: A Diverse, Multi-Category Visual Complexity Dataset,” *arXiv:1810.01771v1*, 2018, [Online]. Available: <http://arxiv.org/abs/1810.01771>
- [41] D. Comaniciu and P. Meer, “Mean Shift: A Robust Approach Toward Feature Space Analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 1–18, 2002, doi:

- 10.1109/34.1000236.
- [42] D. Hasler and S. E. Suesstrunk, "Measuring colourfulness in natural images," in *Electronic Imaging 2003*, 2003, pp. 87–95. doi: 10.1117/12.477378.
 - [43] M. Solli and R. Lenz, "Color harmony for image indexing," in *ICCV Workshops 2009*, 2009, pp. 1885–1892. doi: 10.1109/ICCVW.2009.5457512.
 - [44] R. Rosenholtz, Y. Li, and L. Nakano, "Measuring visual clutter," *J. Vis.*, vol. 7, no. 2, pp. 1–22, 2007, doi: 10.1167/7.2.17.
 - [45] Y.-Q. Chen, J. Duan, Y. Zhu, X.-F. Qian, and B. Xiao, "Research on the image complexity based on neural network," in *Proceedings of the 2015 International Conference on Machine Learning and Cybernetics*, 2015, pp. 295–300.
 - [46] F. Yaghmaee and M. Jamzad, "Estimating watermarking capacity in gray scale images based on image complexity," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 851920, pp. 1–9, 2010, doi: 10.1155/2010/851920.
 - [47] A. Conci and F. Aquino, "Fractal coding based on image local fractal dimension," *Comput. Appl. Math.*, vol. 24, no. 1, pp. 83–98, 2005, doi: 10.1590/S0101-82052005000100005.
 - [48] M. Jamzad and F. Yaghmaee, "Achieving higher stability in watermarking according to image complexity," *Sci. Iran. J.*, vol. 13, no. 4, pp. 404–412, 2006, [Online]. Available: http://www.plan.sid.ir/en/VEWSSID/J_pdf/95520060405.pdf
 - [49] M. P. Da Silva, V. Courboulay, and P. Estrailier, "Image Complexity Measure Based on Visual Attention," in *IEEE International Conference On Image Processing*, 2011, pp. 3281–3284. doi: 10.1109/ICIP.2011.6116371.
 - [50] J. Huo, "A measurement method for the mismatch between the image target and salient points as a metric for image complexity," in *Science and Information Conference, SAI 2015*, 2015, pp. 645–649. doi: 10.1109/SAI.2015.7237210.
 - [51] J. Huo, "Image Complexity and Visual Working Memory Capacity," *Emerg. Trends Adv. Technol. Comput. Intell.*, vol. 647, no. 1, pp. 301–314, 2016, doi: 10.1007/978-3-319-33353-3.
 - [52] H. C. Purchase, E. Freeman, and J. Hamer, "An Exploration of Visual Complexity," in *International Conference on Theory and Application of Diagrams*, 2012, vol. 7352, pp.

- 200–213. doi: 10.1007/978-3-642-31223-6_22.
- [53] J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, 1986, doi: 10.1109/TPAMI.1986.4767851.
 - [54] X. Guo, C. M. Asano, A. Asano, T. Kurita, and L. Li, “Analysis of texture characteristics associated with visual complexity perception,” *Opt. Rev.*, vol. 19, no. 5, pp. 306–314, 2012, doi: 10.1007/s10043-012-0047-1.
 - [55] X. Guo, C. M. Asano, A. Asano, and T. Kurita, “Visual complexity perception and texture image characteristics,” in *2011 International Conference on Biometrics and Kansei Engineering*, 2011, pp. 260–265. doi: 10.1109/ICBAKE.2011.13.
 - [56] K. Reinecke *et al.*, “Predicting users’ first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 2049–2058. doi: 10.1145/2470654.2481281.
 - [57] L. Sun, T. Yamasaki, and K. Aizawa, “Relationship Between Visual Complexity and Aesthetics: Application to Beauty Prediction of Photos,” in *Computer Vision - ECCV 2014 Workshops*, 2015, vol. 8925, pp. 20–34. doi: 10.1007/978-3-319-16199-0.
 - [58] J. A. Martín H., M. Santos, and J. de Lope, “Orthogonal variant moments features in image analysis,” *Inf. Sci. (Ny)*, vol. 180, no. 6, pp. 846–860, 2010, doi: 10.1016/j.ins.2009.08.032.
 - [59] N. D. B. Bruce, N. D. B. Bruce, J. K. Tsotsos, and J. K. Tsotsos, “Saliency, attention, and visual search: An information theoretic approach,” *J. Vis.*, vol. 9, no. 2009, pp. 1–24, 2009, doi: 10.1167/9.3.5.Introduction.
 - [60] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, “A coherent computational approach to model bottom-up visual attention,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 802–817, 2006, doi: 10.1109/TPAMI.2006.86.
 - [61] S. E. Corchs, G. Ciocca, E. Bricolo, and F. Gasparini, “Imaging and Vision Laboratory, Department of Informatics, Systems and Communication, University of Milano-Bicocca,” 2016. <http://www.ivl.disco.unimib.it/activities/complexity-perception-in-images> (accessed May 02, 2018).
 - [62] H. Sheik, Z. Wang, L. Cormack, and A. Bovik, “LIVE Image Quality Assessment

- Database Release 2,” 2006. <http://live.ece.utexas.edu/research/quality/> (accessed May 02, 2018).
- [63] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004, doi: 10.1109/TIP.2003.819861.
- [64] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “Image Quality Assessment Algorithms,” *Image Process. IEEE Trans.*, vol. 15, no. 11, pp. 3441–3452, 2006, doi: 10.1109/TPCG.2004.1314471.
- [65] S. Corchs, F. Gasparini, and R. Schettini, “No reference image quality classification for JPEG-distorted images,” *Digit. Signal Process. A Rev. J.*, vol. 30, no. 1, pp. 86–100, 2014, doi: 10.1016/j.dsp.2014.04.003.
- [66] S. E. Corchs, G. Ciocca, E. Bricolo, and F. Gasparini, “Imaging and Vision Laboratory, Department of Informatics, Systems and Communication, University of Milano-Bicocca,” 2014. <http://www.ivl.disco.unimib.it/activities/image-quality> (accessed May 02, 2018).
- [67] “MIT Media Lab, Vision Texture,” 2002. <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/> (accessed May 02, 2018).
- [68] C. Cusano, P. Napoletano, and R. Schettini, “Evaluating color texture descriptors under large variations of controlled lighting conditions,” *J. Opt. Soc. Am.*, vol. 33, no. 1, pp. 17–30, 2016, doi: 10.1364/JOSAA.33.000017.
- [69] S. E. Corchs, G. Ciocca, E. Bricolo, and F. Gasparini, “RawFoot DB - Imaging and Vision Laboratory - Department of Systems, Informatics and Communication - University of Milan-Bicocca, Italy,” 2016. <http://projects.ivl.disco.unimib.it/rawfoot> (accessed May 02, 2018).
- [70] G. Ramanarayanan, K. Bala, J. a. Ferwerda, and B. Walter, “Dimensionality of Visual Complexity in Computer Graphics Scenes,” in *Proceedings of SPIE 6806, Human Vision and Electronic Imaging XIII*, 2008, pp. 1–12. doi: 10.1117/12.767029.
- [71] L. Palumbo, A. D. J. Makin, and M. Bertamini, “Examining visual complexity and its influence on perceived duration,” *J. Vis.*, vol. 14, no. 14, pp. 1–18, 2014, doi: 10.1167/14.14.3.doi.

- [72] A. Forsythe, G. Mulhern, and M. Sawey, “Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing,” *Behav. Res. Methods*, vol. 40, no. 1, pp. 116–129, 2008, doi: 10.3758/BRM.40.1.116.
- [73] B. Bonev, L. L. Chuang, and F. Escolano, “How do image complexity, task demands and looking biases influence human gaze behavior?,” *Pattern Recognit. Lett.*, vol. 34, no. 7, pp. 723–730, 2013, doi: 10.1016/j.patrec.2012.05.007.
- [74] S. J. White, M. Hirotsu, and S. P. Liversedge, “Eye movement behaviour during reading of Japanese sentences: Effects of word length and visual complexity,” *Read. Writ.*, vol. 25, no. 5, pp. 981–1006, 2012, doi: 10.1007/s11145-010-9289-0.
- [75] S. P. Roth, A. N. Tuch, E. D. Mekler, J. A. Bargas-Avila, and K. Opwis, “Location matters, especially for non-salient features-An eye-tracking study on the effects of web object placement on different types of websites,” *Int. J. Hum. Comput. Stud.*, vol. 71, no. 3, pp. 228–235, 2013, doi: 10.1016/j.ijhcs.2012.09.001.
- [76] H. Cho *et al.*, “Cortical responses and shape complexity of stereoscopic image - A simultaneous EEG/MEG study,” *NeuroSignals*, vol. 24, no. 1, pp. 102–112, 2016, doi: 10.1159/000442617.
- [77] L. Madrid, “Diseño de algoritmos para extracción de características de señal P300 en interfaces cerebro computadora,” Instituto Tecnológico de Chihuahua, 2018.
- [78] E. Donchin and L. A. Farwell, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, 1988.
- [79] M. I. Chacón, L. E. Aguilar D., and A. Delgado S., “Fuzzy adaptive edge definition based on the complexity of the image,” in *IEEE International Conference on Fuzzy Systems*, 2002, vol. 2, pp. 675–678. doi: 10.1109/FUZZ.2001.1009045.
- [80] G. Ciocca, S. Corchs, and F. Gasparini, “A complexity-based image analysis to investigate interference between distortions and image contents in Image Quality Assessment,” in *CCIW 2017: Computational Color Imaging*, 2017, vol. 10213, pp. 1–15. doi: 10.1007/978-3-319-56010-6.
- [81] F. Garcia-Lamont, J. Cervantes, A. López, and L. Rodríguez, “Segmentation of images by color features: A survey,” *Neurocomputing*, vol. 292, no. 1, pp. 1–27, 2018, doi:

- 10.1016/j.neucom.2018.01.091.
- [82] R. Klette, *Concise Computer Vision*. London: Springer, 2014. doi: 10.1007/978-1-4471-6320-6.
 - [83] P. Kohli, A. Osokin, and S. Jegelka, “A Principled Deep Random Field Model for Image Segmentation,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1971–1978. doi: 10.1109/CVPR.2013.257.
 - [84] J. Sevak, A. Kapadia, J. Chavda, A. Shah, and M. Rahevar, “Survey on Semantic Image Segmentation Techniques,” in *International Conference on Intelligent Sustainable Systems (ICISS)*, 2017, pp. 306–313. doi: 10.1109/ISS1.2017.8389420.
 - [85] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, “A survey on deep learning techniques for image and video semantic segmentation,” *Appl. Soft Comput.*, vol. 70, no. 1, pp. 41–65, 2018, doi: 10.1016/j.asoc.2018.05.018.
 - [86] W. Khan, “Image Segmentation Techniques: A Survey,” *J. Image Graph.*, vol. 1, no. 4, pp. 166–170, 2014, doi: 10.12720/joig.1.4.166-170.
 - [87] H. S. Park and C. H. Jun, “A simple and fast algorithm for {K}-medoids clustering,” *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, 2009, doi: 10.1016/j.eswa.2008.01.039.
 - [88] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and A. K. Nandi, “Significantly Fast and Robust Fuzzy C-Means Clustering Algorithm Based on Morphological Reconstruction and Membership Filtering,” *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3027–3041, 2018, doi: 10.1109/TFUZZ.2018.2796074.
 - [89] T. Lei, X. Jia, Y. Zhang, S. Liu, H. Meng, and A. K. Nandi, “Superpixel-Based Fast Fuzzy C-Means Clustering for Color Image Segmentation,” *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1753–1766, 2019, doi: 10.1109/TFUZZ.2018.2889018.
 - [90] X. Jia, T. Lei, X. Du, S. Liu, H. Meng, and A. K. Nandi, “Robust Self-Sparse Fuzzy Clustering for Image Segmentation,” *IEEE Access*, vol. 4, no. 8, pp. 1–14, 2020, doi: 10.1109/ACCESS.2020.3015270.
 - [91] T. Lei, P. Liu, X. Jia, X. Zhang, H. Meng, and A. K. Nandi, “Automatic Fuzzy Clustering Framework for Image Segmentation,” *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 2078–2092, 2020, doi: 10.1109/tfuzz.2019.2930030.

- [92] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, 2011, doi: 10.1109/TPAMI.2010.161.
- [93] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, 2001, pp. 416–423.
- [94] W. M. Rand, “Objective criteria for the evaluation of clustering methods,” *J. Am. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971, doi: 10.1080/01621459.1971.10482356.
- [95] M. Meilă, “Comparing clusterings - An axiomatic view,” in *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*, 2005, pp. 577–584.
- [96] X. Wang, Y. Tang, S. Masnou, and L. Chen, “A Global/Local Affinity Graph for Image Segmentation,” *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1399–1411, 2015, doi: 10.1109/TIP.2015.2397313.
- [97] H. Bansal, L. Bhagat, S. Mittal, and A. Tiwari, “Image Correction and Identification of Ishihara Test Images for Color Blind Individual,” in *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*, 2021, pp. 997–1005. doi: 10.1007/978-981-16-0733-2_70.
- [98] R. M. Haralick and K. Shanmugam, “Textural Features for Image Classification,” *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973, doi: 10.1109/TSMC.1973.4309314.
- [99] S. Corchs, F. Gasparini, and R. Schettini, “Grouping strategies to improve the correlation between subjective and objective image quality data,” in *Electronic Imaging International Society for Optics and Photonics*, 2013, vol. 8653, pp. 1–8. doi: 10.1117/12.2006227.
- [100] M. Artese, G. Ciocca, and I. Gagliardi, “Good 50x70 Project: A portal for Cultural And Social Campaigns,” in *Society for Imaging Science and Technology*, 2014, pp. 213–218.
- [101] E. Odhiambo Omuya, G. Onyango Okeyo, and M. Waema Kimwele, “Feature Selection for Classification using Principal Component Analysis and Information Gain,” *Expert Syst. Appl.*, vol. 174, no. 114765, pp. 1–12, Jul. 2021, doi: 10.1016/j.eswa.2021.114765.

- [102] K. M. Z. Othman and A. M. Salih, “Scaled conjugate gradient ann for industrial sensors calibration,” *Bull. Electr. Eng. Informatics*, vol. 10, no. 2, pp. 680–688, 2021, doi: 10.11591/eei.v10i2.2738.
- [103] N. Gauvrit, F. Soler-Toscano, and A. Guida, “A preference for some types of complexity comment on ‘perceived beauty of random texture patterns: A preference for complexity,’” *Acta Psychol. (Amst)*, vol. 174, no. 1, pp. 48–53, 2017, doi: 10.1016/j.actpsy.2017.01.007.
- [104] X. Guo, Y. Qian, L. Li, and A. Asano, “Assessment model for perceived visual complexity of painting images,” *Knowledge-Based Syst.*, vol. 159, no. 1, pp. 110–119, Nov. 2018, doi: 10.1016/j.knosys.2018.06.006.
- [105] L. Madrid-Herrera, M. I. Chacon-Murguia, D. A. Posada-Urrutia, and J. A. Ramirez-Quintana, “Human Image Complexity Analysis Using a Fuzzy Inference System,” in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2019, pp. 1–6. doi: 10.1109/FUZZ-IEEE.2019.8858966.
- [106] A. Gartus and H. Leder, “Predicting perceived visual complexity of abstract patterns using computational measures: The influence of mirror symmetry on complexity perception,” *PLoS One*, vol. 12, no. 11, pp. 1–30, 2017, doi: 10.1371/journal.pone.0185276.
- [107] B. Zhou, S. Xu, and X. X. Yang, “Computing the color complexity of images,” in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015*, 2015, no. 12, pp. 1898–1902. doi: 10.1109/FSKD.2015.7382237.
- [108] X. Chen and L. Pan, “A Survey of Graph Cuts/Graph Search Based Medical Image Segmentation,” *IEEE Rev. Biomed. Eng.*, vol. 11, no. 1, pp. 112–124, 2018, doi: 10.1109/RBME.2018.2798701.
- [109] F. Kulwa *et al.*, “A State-of-the-Art Survey for Microorganism Image Segmentation Methods and Future Potential,” *IEEE Access*, vol. 7, no. 1, pp. 100243–100269, 2019, doi: 10.1109/access.2019.2930111.
- [110] G. Liu, Y. Zhang, and A. Wang, “Incorporating Adaptive Local Information Into Fuzzy Clustering for Image Segmentation,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3990–4000, 2015.

- [111] F. Zhao, F. Liu, C. Li, H. Liu, R. Lan, and J. Fan, “Coarse–fine surrogate model driven multiobjective evolutionary fuzzy clustering algorithm with dual memberships for noisy image segmentation,” *Appl. Soft Comput.*, vol. 112, no. 107778, pp. 1–17, 2021, doi: 10.1016/j.asoc.2021.107778.
- [112] J. Miao, X. Zhou, and T. Z. Huang, “Local segmentation of images using an improved fuzzy C-means clustering algorithm based on self-adaptive dictionary learning,” *Appl. Soft Comput.*, vol. 91, no. 106200, pp. 1–15, 2020, doi: 10.1016/j.asoc.2020.106200.
- [113] D. Marutho, S. Hendra Handaka, E. Wijaya, and Muljono, “The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News,” in *International Seminar on Application for Technology of Information and Communication*, 2018, pp. 533–538. doi: 10.1109/ISEMANTIC.2018.8549751.
- [114] C. Subbalakshmi, R. Sayal, and H. S. Saini, “Cluster Validity Using Modified Fuzzy Silhouette Index on Large Dynamic Data Set,” *Comput. Intell. Data Mining. Adv. Intell. Syst. Comput.*, vol. 990, no. 1, pp. 1–14, 2020, doi: 10.1007/978-981-13-8676-3.
- [115] C. Yuan and H. Yang, “Research on K-Value Selection Method of K-Means Clustering Algorithm,” *Multidiscip. Sci. J.*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.
- [116] M. S. Yang and Y. Nataliani, “Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters,” *Pattern Recognit.*, vol. 71, no. 1, pp. 45–59, 2017, doi: 10.1016/j.patcog.2017.05.017.
- [117] X. Jia, T. Lei, P. Liu, D. Xue, H. Meng, and A. K. Nandi, “Fast and Automatic Image Segmentation Using Superpixel-Based Graph Clustering,” *IEEE Access*, vol. 8, no. 1, pp. 211526–211539, 2020, doi: 10.1109/ACCESS.2020.3039742.
- [118] A. Radhakrishna, S. Appu, S. Kevin, L. Aurelien, F. Pascal, and S. Sabine, “SLIC Superpixels Compared to State-of-the-Art Superpixel Methods,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, 2012, doi: 10.1109/TPAMI.2012.120.
- [119] J. Chen, Z. Li, and B. Huang, “Linear spectral clustering superpixel,” *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3317–3330, 2017, doi: 10.1109/TIP.2017.2651389.
- [120] T. Lei, X. Jia, T. Liu, S. Liu, H. Meng, and A. K. Nandi, “Adaptive Morphological Reconstruction for Seeded Image Segmentation,” *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5510–5523, 2019, doi: 10.1109/TIP.2019.2920514.

- [121] J. Hou, A. Zhang, and N. Qi, “Density peak clustering based on relative density relationship,” *Pattern Recognit.*, vol. 108, no. 1, pp. 1–16, 2020, doi: 10.1016/j.patcog.2020.107554.
- [122] R. Unnikrishnan, C. Pantofaru, and M. Hebert, “Toward objective evaluation of image segmentation algorithms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 929–944, 2007, doi: 10.1109/TPAMI.2007.1046.
- [123] W. Wu and J. M. Keller, “Sequential possibilistic local information one-means clustering for image segmentation,” in *IEEE International Conference on Fuzzy Systems*, 2020, vol. 2020-July. doi: 10.1109/FUZZ48607.2020.9177576.
- [124] M. Simfukwe, B. Peng, and T. Li, “Fusion of measures for image segmentation evaluation,” *Int. J. Comput. Intell. Syst.*, vol. 12, no. 1, pp. 379–386, 2019, doi: 10.2991/ijcis.2018.125905654.
- [125] Z. Li and J. Chen, “Superpixel segmentation using Linear Spectral Clustering,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1356–1363. doi: 10.1109/CVPR.2015.7298741.
- [126] T. Lei, “Shaanxi University of Science and Technology - Artificial Intelligence and Machine Vision Laboratory,” 2020. <https://aimv.sust.edu.cn/lwgcg.htm> (accessed Nov. 08, 2020).
- [127] Z. Wang, E. Wang, and Y. Zhu, *Image segmentation evaluation: a survey of methods*, vol. 53, no. 8. Springer Netherlands, 2020. doi: 10.1007/s10462-020-09830-9.
- [128] S. Rovetta, F. Masulli, and A. Cabri, “The ‘Probabilistic Rand Index’: A Look from Some Different Perspectives,” *Neural Approaches to Dyn. Signal Exch.*, vol. 151, no. 1, pp. 95–105, 2019, doi: 10.1007/978-981-13-8950-4_10.
- [129] B. Peng, L. Zhang, X. Mou, and M. H. Yang, “Evaluation of Segmentation Quality via Adaptive Composition of Reference Segmentations,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 10, pp. 1929–1941, 2017, doi: 10.1109/TPAMI.2016.2622703.
- [130] J. Pont-Tuset and F. Marques, “Supervised Evaluation of Image Segmentation and Object Proposal Techniques,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1465–1478, 2016, doi: 10.1109/TPAMI.2015.2481406.
- [131] M. Versaci and F. C. Morabito, “Image Edge Detection: A New Approach Based on

- Fuzzy Entropy and Fuzzy Divergence,” *Int. J. Fuzzy Syst.*, vol. 23, no. 4, pp. 918–936, 2021, doi: 10.1007/s40815-020-01030-5.
- [132] R. Soleymani, E. Granger, and G. Fumera, “F-measure curves: A tool to visualize classifier performance under imbalance,” *Pattern Recognit.*, vol. 100, no. 107146, pp. 1–19, 2020, doi: 10.1016/j.patcog.2019.107146.
- [133] Y. H. Nai *et al.*, “Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset,” *Comput. Biol. Med.*, vol. 134, no. 104497, pp. 1–12, 2021, doi: 10.1016/j.combiomed.2021.104497.
- [134] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002, doi: 10.1109/34.1000236.