



UNIVERSIDAD MICHOACANA
DE SAN NICOLÁS DE HIDALGO

DIVISIÓN DE ESTUDIOS DE POSGRADO DE LA FACULTAD
DE INGENIERÍA ELÉCTRICA

FUSIÓN DE IMÁGENES MULTIFOCO POR MEDIO DEL
FILTRADO DE REGIONES DE ALTA Y BAJA FRECUENCIA

TESIS

QUE PARA OBTENER EL GRADO DE MAESTRO EN CIENCIAS EN INGENIERÍA ELÉCTRICA

PRESENTA:

Ing. Roberto Iván Orozco Cuevas

ASESOR:

Dr. Félix Calderón Solorio
Doctorado en ciencias computacionales

Morelia Michoacán, Agosto del 2013

Resumen

Las herramientas que hacen uso de lentes ópticos, tales como cámaras fotográficas o microscopios, presentan una limitación en común inherente a las propiedades geométricas de los lentes llamada profundidad de campo. Esta se define el rango de distancias al lente en el cual los objetos se encuentran enfocados y objetos fuera de este rango aparecen borrosos. Esta limitación impide obtener una imagen donde todos los objetos en la escena estén enfocados ya que varios de los objetos de interés pueden encontrarse a diferentes distancias del lente. El objetivo de la fusión de imágenes es la de obtener la unión de dos o más imágenes tomadas de la misma escena, donde cada imagen está enfocada a una distancia distinta en la escena, produciendo una imagen con toda la escena enfocada. El resultado de la fusión es entonces una imagen fusionada que contiene múltiples puntos de enfoque o bien un rango de profundidad de campo extendido. En este trabajo se presenta un método con el cual se supera esta limitación, basándose en la selección de regiones de alto enfoque en las imágenes, por medio de un filtro pasa altas. Se comparó este filtro con otras funciones que son usadas como filtros de frecuencia y que se pueden encontrar asociadas a otros métodos de fusión de imágenes multifoco. Después, se logró mejorar la imagen fusionada, haciendo uso de un algoritmo de segmentación basado en campos aleatorios de Markov. Dicha segmentación, la cual se realiza sobre un mapa de decisión que define las regiones de alto enfoque seleccionadas de las imágenes originales, determina para cada píxel y por medio de sus vecinos, la probabilidad de corresponder a la región de alto o bajo enfoque. También se presenta una comparación del método propuesto con otros métodos de fusión de imágenes, usando conjuntos de imágenes de casos reales así como de imágenes multifoco generadas de forma artificial. Las pruebas experimentales del método, presentaron resultados que superan a métodos de fusión usados con frecuencia en el estado del arte, por lo que se presenta un avance en la solución del problema de fusión de imágenes multifoco.

Abstract

Tools that make use of optical lenses, such as photographic cameras or microscopes, exhibit a common limitation due to the geometric properties of these lenses, which is called depth of field. This defines a range of distances to the lens in which objects in the scene are in focus and objects outside this range are out of focus or without sharpness. This limitation, hinders the obtention of an image where all the objects in the scene are in focus, because several of these objects may be at different distances from the lens. The purpose of image fusion is to obtain the union of two or more images taken from the same scene, where each image is focused at a different distance in the scene, producing an image where all the scene is in focus, thus simulating an extended depth of field. In this thesis a method to overcome this limitation is presented, based on the selection of high focus regions in the image by filtering high and low frequencies using a high pass filter. The proposed filter is compared with other functions that are used as frequency filters and can be found in other multifocus image fusion methods. After this, the fused image was improved, by using a segmentation algorithm which is based on Markov random fields. This segmentation algorithm is used on a decision map that describes the high focus regions selected from the source images, then it uses the neighbors of a each pixel or regions in this map to define the probability of said pixels belonging to a high or low focus region. A comparison of the proposed method and other image fusion methods is presented, using sets of multifocus images of real cases as well as artificially generated multifocus images. The results obtained in the experimental tests showed that the proposed method outperformed other fusion methods, commonly used in multifocus image fusion, so the proposed method provides an advance in the multifocus image fusion problem.

Contenido

Resumen	III
Abstract	V
Contenido	VII
Lista de Figuras	IX
Lista de Tablas	XIII
Lista de Símbolos	XV
1. Introducción.	1
1.1. Definición del problema	7
1.1.1. Motivación	8
1.2. Objetivos de la Tesis	9
1.3. Descripción de Capítulos	9
2. Estado del arte.	11
2.1. Métodos basados en selección de regiones	11
2.1.1. División en bloques	12
2.2. Métodos basados en aprendizaje	17
2.3. Métodos basados en descomposición multi escala	20
2.3.1. Pirámide Laplaciana	20
2.3.2. Transformada Discreta Wavelet	23
2.4. Conclusiones	27
3. Selección de regiones de alto enfoque, por medio de frecuencia.	29
3.1. Fusión de imágenes usando un filtro pasa altas	30
3.2. Filtros de frecuencia asociados a los métodos de fusión de imágenes	34
3.2.1. Laplaciano de una Gaussiana	35
3.2.2. Diferencia de Gaussianas	37
3.2.3. Wavelets Haar	41
3.3. Filtro pasa altas propuesto	42
3.4. Medición de la calidad de la fusión	43
3.5. Medición de la calidad en imágenes sintéticas	44
3.5.1. Índice universal de calidad	44
3.5.2. Suma de diferencias cuadradas	45
3.6. Medición de la calidad en imágenes sin referencia	46

3.6.1. Entropía	46
3.6.2. Desviación estandar	46
3.6.3. Promedio de gradiente	47
3.7. Desempeño de mediciones de calidad en la fusión	47
3.8. Generación de imágenes de prueba	48
3.8.1. Mapa de decisión resultante	50
3.8.2. Coeficiente Tanimoto	52
3.9. Pruebas experimentales con imágenes sintéticas	53
3.9.1. Conclusiones del capítulo	60
4. Segmentación del mapa de decisión.	61
4.1. Campos aleatorios de Markov	61
4.1.1. Segmentación controlada por entropía	64
4.2. Resultados	70
4.2.1. Segmentación del mapa de decisión obtenido con la fusión de imágenes sintéticas	70
4.2.2. Segmentación del mapa obtenido con la fusión de imágenes reales	71
4.3. Conclusiones	73
5. Resultados.	75
5.1. Conclusiones del capítulo	88
6. Conclusiones y trabajos futuros	89
6.1. Conclusiones Generales	89
6.2. Trabajos Futuros	90
Glosario	93
Referencias	95

Lista de Figuras

1.1.	Cambio de enfoque conforme se acerca el sensor a la escena.	2
1.2.	Dos imágenes cada una representando un plano de enfoque distinto.	2
1.3.	Formación de una imagen con un lente. La luz ambiental en la escena que es capturada y refractada por el lente, forma una imagen invertida de la escena sobre el material fotosensible del dispositivo.	3
1.4.	Diagrama correspondiente a la ecuación del lente y a la formación de la imagen por el lente, [Jenkins y White, 1957].	4
1.5.	Formación del círculo de confusión al cambiar la distancia del lente a la escena. En las tres figuras, a la derecha se tiene una vista lateral y frontal de la superficie del arreglo de sensores CCD/CMOS, en la que se proyecta la imagen reflejada por el lente.	5
1.6.	Proceso de digitalización de una imagen [Woods, 2008].	6
1.7.	Diferentes fotografías del interior de un reloj, en las imágenes (a) hasta (c), fusionadas en una imagen completamente enfocada en (d).	8
2.1.	$ H(k) $	14
2.2.	Cambio en nivel de nitidez o enfoque de la imagen de Lena, al aumentar el valor de σ en el filtro Gaussiano.	15
2.3.	Proceso de fusión de imágenes usando división en bloques.	16
2.4.	Perceptrón con P salidas, cada salida es una suma ponderada de las entradas x_j con los pesos w_{ij}	19
2.5.	Formación de la pirámide de Laplace a partir de una pirámide Gaussiana.	22
2.6.	Proceso de fusión de imágenes usando la pirámide Laplaciana.	22
2.7.	Proceso de descomposición de una imagen con los wavelets Haar.	25
2.8.	Múltiples bandas de frecuencia y resoluciones obtenidas con la transformada discreta wavelet. a) imagen original, b) descomposición wavelet de un nivel y c) descomposición wavelet de dos niveles.	25
3.1.	a) Imagen original a filtrar, b) y d) magnitud de la transformada de Fourier de los filtros, c) y e) sus respectivas respuestas.	35
3.2.	Laplaciano de una Gaussiana con $\sigma = 1.0$, y la magnitud de su transformada de Fourier.	37

3.3. Diferencia de Gaussianas con $\sigma = 1.0$, $\tau = 1.1$ y la magnitud de su transformada de Fourier	40
3.4. Comparación de las gráficas del Laplaciano de una Gaussiana, y diferentes niveles de aproximación a este, usando diferencia de Gaussianas. La línea punteada corresponde a la diferencia de Gaussianas, en todas las gráficas se usa un valor de $\sigma = 1.0$	40
3.5. Magnitud de las transformadas de Fourier de los wavelets Haar.	41
3.6. Filtro pasa altas propuesto, con $\sigma = 1.0$ y la magnitud de su TDF.	43
3.7. Aquí se muestra cómo el emborronado de una imagen disminuye la medición de calidad de la misma.	48
3.8. Generación de imágenes multi-foco sintéticas, a) imagen inicial I , b) I_g resultante de la convolución de la imagen I con el filtro Gaussiano g , c) mapa de decisión M_o , con el que se eligen pixeles de las imágenes I e I_g , d) y e) imágenes sintéticas resultantes.	50
3.9. Conjunto de imágenes sintéticas usadas en las pruebas.	53
3.10. Resultados de la comparación de fusiones, se comparan los filtros variando el valor del ancho de banda filtrado contra la calidad de la fusión.	55
3.11. Calidad contra frecuencia filtrada.	55
3.12. Calidad contra frecuencia filtrada.	56
3.13. Mapas de decisión binarios M_r , de los mejores resultados de fusión según las medidas de calidad: (a) a (d) Entropía, (e) a (h) Desviación estandar, (i) a (l) Suma de diferencias cuadradas.	57
3.14. Mapas de decisión binarios M_r , de los mejores resultados de fusión según las medidas de calidad: (a) a (d) Índice universal de calidad, (e) a (h) Promedio de gradiente, (i) a (l) Coeficiente de Tanimoto.	58
3.15. Mapa de decisión obtenido de la fusión de las imágenes sintéticas de la figura 3.9, usando el filtro pasa altas (3.13) propuesto, con un valor de $\sigma = 0.4$	59
4.1. Componentes de un campo aleatorio de Markov a) Sistemas de vecinos de diferentes órdenes, b) Un sistema de vecinos de primer orden, c) Formas de los cliques de primer, segundo y tercer orden.	63
4.2. Mapas de decisión obtenidos en la fusión de imágenes usando el Algoritmo 3.	70
4.3. Resultados de la segmentación para diferentes valores de λ	71
4.4. Resultados de la segmentación para diferentes valores de λ	72
4.5. Selección de las regiones de alta frecuencia en las imágenes originales, usando el mapa de decisión segmentado, el mapa usado corresponde a el filtro pasa altas propuesto, con un valor de $\sigma = 0.3$ y la segmentación usando un valor de $\lambda = 3.0$	72
5.1. Fusión de imágenes con el Laplaciano de una Gaussiana (3.5) y segmentación.	76
5.2. Fusión de imágenes con una diferencia de Gaussianas (3.7) y segmentación.	77
5.3. Fusión de imágenes con el wavelet Haar HH (3.11) y segmentación.	78
5.4. Fusión de imágenes con el filtro propuesto (3.13) y segmentación.	79
5.5. Resultados de fusión usando el promedio de las imágenes, a) resultado de la fusión, b) acercamiento al reloj izquierdo, c) acercamiento al reloj derecho.	80

5.6. Resultados de fusión usando la transformada discreta wavelet, a) resultado de la fusión, b) acercamiento al reloj izquierdo, c) acercamiento al reloj derecho.	81
5.7. Resultados de fusión usando la pirámide Laplaciana, a) resultado de la fusión, b) acercamiento al reloj izquierdo, c) acercamiento al reloj derecho.	82
5.8. Resultados de fusión usando el método propuesto, a) resultado de la fusión, b) acercamiento al reloj izquierdo, c) acercamiento al reloj derecho.	83
5.9. Resultados de los diferentes métodos de fusión.	85
5.10. Comparación del método propuesto y el método de pirámide Laplaciana. . .	86
5.11. Resultados de los diferentes métodos de fusión.	87

Lista de Tablas

2.1. Valores de frecuencia espacial de la imagen Lena, convolucionada con una Gaussiana de diferentes valores de σ	15
---	----

Lista de Símbolos

$A(I)$	Función que mide el nivel de actividad de una imagen I .
B	Conjunto de variables aleatorias que forman un campo aleatorio de Markov.
b	Uno de los estados o configuraciones en las que se puede encontrar el campo aleatorio de Markov.
c_i	i -ésima característica extraída de una imagen.
\mathcal{C}	Clique en un grafo.
d	Índice del d -ésimo bloque en la división de una imagen.
$E(b)$	Función que calcula el nivel de energía en el estado b del campo aleatorio B .
\mathcal{F}	Transformada discreta de Fourier.
G_z	z -ésimo nivel en la pirámide Gaussiana.
I	Imagen en forma de función matricial.
$I(i, j)$	Pixel o elemento de una imagen en el renglón i , columna j .
\hat{I}_a	Imagen I_a convolucionada con un filtro de frecuencia.
\hat{I}_b	Imagen I_b convolucionada con un filtro de frecuencia.
I_a	Imagen con profundidad de campo centrada en una distancia a .
I_b	Imagen con profundidad de campo centrada en una distancia b distinta de a .
I_c	Imagen compuesta o fusionada obtenida con un método de fusión de imágenes multifoco.
$I_{a,d}$	d -ésimo bloque de la división de I_a .
$I_{b,d}$	d -ésimo bloque de la división de I_b .
I_c	Imagen compuesta o fusionada obtenida con un método de fusión de imágenes multifoco.
K	Número total de regiones producidas en la segmentación de una imagen.
\mathcal{L}	Lagrangiano de una función a minimizar.
L_z	z -ésimo nivel en la pirámide Laplaciana.
m	Índice de elementos en una señal discreta en el dominio de tiempo.
M	Número de renglones que componen una imagen.

M_o	Mapa de decisión binario que determina las regiones que se toman de las imágenes, I , y la versión emborronada I_g , para formar las imágenes sintéticas.
M_r	Mapa de decisión binario que determina las regiones seleccionadas de las imágenes en el método de fusión.
n	Índice de elementos en una señal discreta en el dominio de tiempo.
$n_g(s)$	Ruido con distribución normal.
N	Número de columnas que componen una imagen.
$\mathcal{N}_b(s)$	Conjunto de vecinos del sitio s en el campo aleatorio de Markov
P	Numero de perceptrones o salidas en una red neuronal.
$Pr(a)$	Función de probabilidad de una variable aleatoria a .
Q	Numero de entradas o tamaño del vector de entrada en una red neuronal.
R_k	k -ésima región producida en la segmentación de una imagen.
s	Sitio o coordenada en la imagen correspondiente al campo aleatorio de Markov.
t	Sitio o coordenada en la imagen correspondiente al campo aleatorio de Markov.
u	Índices de elementos en una señal discreta en el dominio de Fourier.
v	Índices de elementos en una señal discreta en el dominio de Fourier.
\mathbf{v}	vector de características extraídas de una imagen.
\mathbf{v}_a	vector de características extraídas de la imagen I_a .
\mathbf{v}_b	vector de características extraídas de la imagen I_b .
\mathbf{w}	Vector de pesos en una red neuronal.
\mathbf{x}	Vector de entrada en una red neuronal.
ξ	Constante de difusión de calor.
$x(n)$	Señales discretas en el dominio de tiempo.
x_i	i -ésimo elemento en el vector de entrada de una red neuronal.
$y(n)$	Señales discretas en el dominio de tiempo.
y_i	i -ésima salida resultante de la red neuronal.
Z	Número de niveles en la pirámide Laplaciana.
Z_b	Constante de normalización en la función de energía del modelo de segmentación.
γ_s	Multiplicador de Lagrange correspondiente a un Lagrangiano.
$\delta(i, j)$	Función delta de Dirac.
λ	Constante de control del nivel de suavidad en el modelo de segmentación.
μ	Constante de control de la entropía en el modelo de segmentación.
σ	Varianza de una función Gaussiana.
τ	Factor de escala entre las Gaussianas de la diferencia de Gaussianas.
$\bar{\theta}$	Vector de parámetros θ_k , correspondientes a las medias en el modelo de segmentación.
$[\dots]_{\downarrow 2}$	Submuestreo de una imagen I descartando renglones y columnas pares.
$[\dots]_{\uparrow 2}$	Escalamiento de una imagen, insertando renglones y columnas cero en las coordenadas i, j pares.
$[\dots]_{\downarrow 2ren}$	Submuestreo de una imagen, descartando solo los renglones pares.
$[\dots]_{\downarrow 2col}$	Submuestreo de una imagen, descartando solo las columnas pares.
$[\dots]_{\uparrow 2ren}$	Escalamiento de una imagen, donde se insertan renglones de ceros en coordenadas i par.
$[\dots]_{\uparrow 2col}$	Escalamiento de una imagen, donde se insertan columnas de ceros en coordenadas j par.

Capítulo 1

Introducción.

Los dispositivos o herramientas que dependen del uso de lentes ópticos, cómo cámaras, o microscopios, tienen una limitación en común conocida cómo profundidad de campo. La cual se define cómo el rango de distancias del lente a los objetos en la escena, en la cual dichos objetos aparecen en forma nítida y enfocados. Los objetos en la escena que no se encuentren dentro de este rango estarán borrosos o desenfocados. Los factores que determinan la longitud de la profundidad de campo son: el nivel de curvatura del lente, la distancia del objeto al lente y en el caso de cámaras fotográficas el tamaño de la apertura del iris en la cámara, que cambia la cantidad de luz que llega al lente [Allen y Triantaphillidou, 2010, Langford et al., 2012]. Por estos factores es difícil tener un dispositivo con una profundidad de campo que abarque a todos los objetos en la escena. Al enfocar un objeto en la escena cambiando la posición de la profundidad de campo, se deja fuera de ésta a los otros objetos que se encuentran a una distancia distinta.

En el diagrama de la Figura 1.1 se ejemplifica este problema. En la Figura 1.1(a) se enfoca la cámara en el objeto más cercano al lente, con lo que la profundidad de campo se centra en el cono, dejando fuera de este rango al cilindro y a la esfera. En la Figura 1.1(b) se ajusta el enfoque de la cámara acercándola al cilindro y lo presenta de forma clara y enfocado, pero aislando al cono y a la esfera del rango de profundidad de campo. Por último, en la Figura 1.1(c) se enfoca el lente en la esfera, acercando aun más la cámara y de la misma forma se deja a los otros objetos fuera del rango de profundidad de campo.

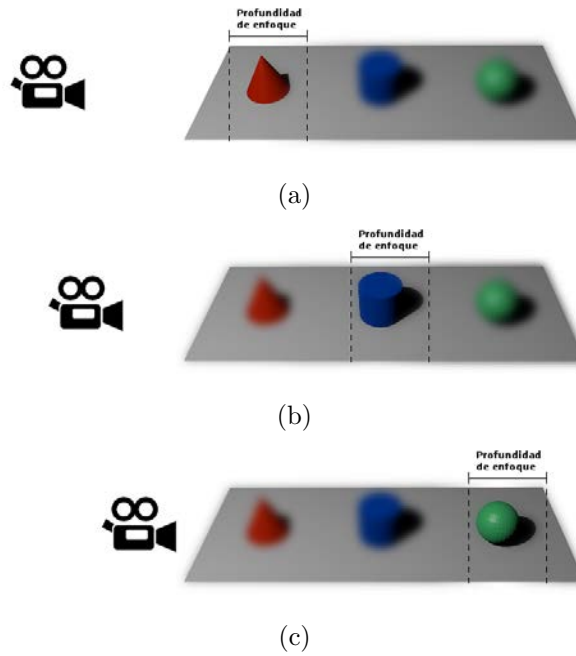


Figura 1.1: Cambio de enfoque conforme se acerca el sensor a la escena.

En la Figura 1.2(a) se muestra un ejemplo real del problema mostrando un reloj perteneciente a un plano de enfoque con la profundidad de campo centrada en el fondo de la escena. Al ajustar el enfoque de la cámara cambiando el tamaño de la apertura del iris se mueve la posición de la profundidad de campo, centrándola en el reloj más cercano a la cámara cómo se ve en la Figura 1.2(b).



(a) Enfoque en el fondo. (b) Enfoque en el reloj delantero.

Figura 1.2: Dos imágenes cada una representando un plano de enfoque distinto.

Los límites de la profundidad de campo o la distancia más cercana o más lejana a en la que un objeto se puede considerar nítido o enfocado se pueden determinar por medio del círculo de confusión [Langford et al., 2012]. La formación de una imagen, por ejemplo en una cámara fotográfica, es lograda gracias a la refracción de la luz que pasa a través de un lente convergente y afecta el plano o material fotosensible en las cámaras analógicas o bien el dispositivo CCD/CMOS en cámaras digitales. En el diagrama de la Figura 1.3 se muestra de forma simplificada este fenómeno. Cada parte o punto del sujeto de interés refleja luz en todas direcciones, pero para la formación de la imagen solo importa la luz que el lente de la cámara alcanza a recibir.

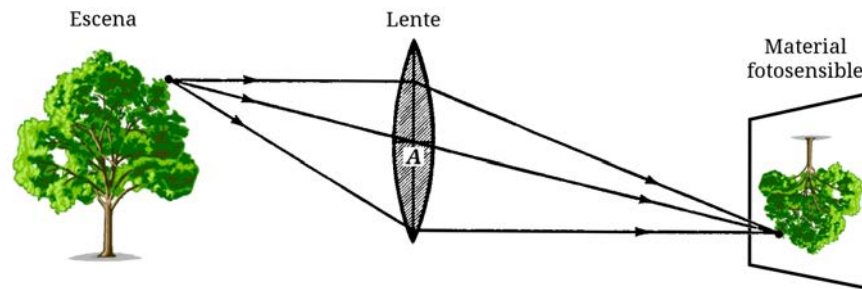


Figura 1.3: Formación de una imagen con un lente. La luz ambiental en la escena que es capturada y refractada por el lente, forma una imagen invertida de la escena sobre el material fotosensible del dispositivo.

Los lentes convergentes tienen una propiedad llamada longitud focal, la cual determina la distancia a la cual los rayos de luz que son paralelos entre si y perpendiculares al eje vertical del lente, son enfocados en el mismo punto en el lado opuesto del lente. Esta longitud focal es determinada por el radio de curvatura del lente. En la Figura 1.4 se presenta un diagrama que muestra como la longitud focal f determina la posición en la que se forma una imagen, dependiendo de la distancia del objeto al lente y el valor de longitud focal. La fórmula que determina estas distancias es llamada ecuación del lente [Jenkins y White, 1957], definida como

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f},$$

donde s es la distancia del objeto en la escena al centro del lente y s' es la distancia del objeto en la reflexión al centro del lente. En el diagrama se observa como un objeto en O a una distancia s del centro del lente, produce una reflexión O' , en el lado opuesto a una distancia s' del lente. Si se coloca un objeto frente al lente en la posición del punto focal F , se obtiene un reflejo en el lado opuesto del lente a la misma distancia.

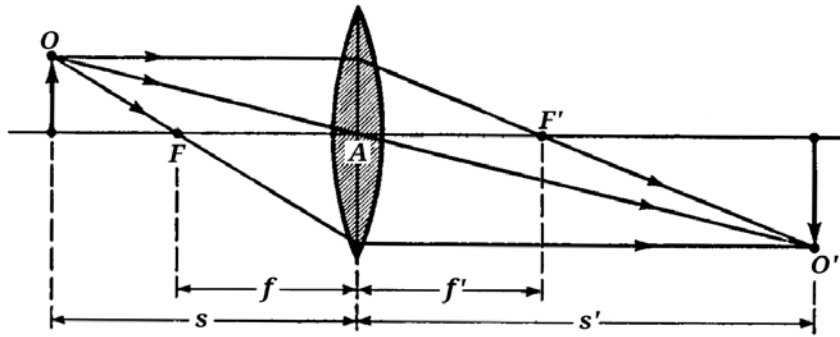


Figura 1.4: Diagrama correspondiente a la ecuación del lente y a la formación de la imagen por el lente, [Jenkins y White, 1957].

De la misma forma que se muestra en la Figura 1.4, la luz que proviene de la escena es refractada en forma cónica sobre el material fotosensible de la cámara. Este cono de luz a la vez forma un círculo sobre el material, cuyo diámetro depende de la distancia del lente al objeto, la longitud focal del lente, y el tamaño de apertura del iris, y se le conoce como círculo de confusión. La profundidad de campo tiene límites definidos por este círculo. En el diagrama de la Figura 1.5 se muestra cómo el diámetro del círculo de confusión determina estas distancias. El efecto de este círculo es más evidente en dispositivos sensores como los CCD, los cuales están compuestos por arreglos de múltiples sensores individuales como se muestra en el lado derecho de la Figura 1.5. En el diagrama se muestra una escena compuesta por un punto de luz, la escena es refractada sobre el material fotosensible. Cambiando el tamaño de apertura en la cámara o la distancia del lente al objeto de interés, se cambia proporcionalmente el diámetro del círculo de confusión.

En 1.5(a) se coloca el punto de luz a una distancia alejada del lente, afectando múltiples sensores. En 1.5(b) se coloca éste punto de luz a una distancia correspondiente

al punto focal del lente o el centro de la profundidad de campo y la luz refractada afecta el menor número de sensores. Centrando el punto de luz en esta posición se tiene el círculo de confusión con el diámetro más pequeño posible producido por la refracción de la luz en la escena, con este diámetro del círculo de confusión se tiene el mejor enfoque [Langford et al., 2012]. En 1.5(c) se acerca de nuevo el punto de luz al lente, afectando más sensores en el dispositivo.

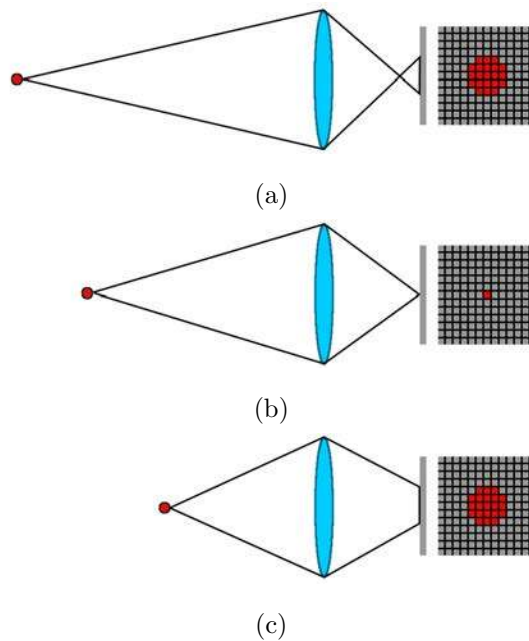


Figura 1.5: Formación del círculo de confusión al cambiar la distancia del lente a la escena. En las tres figuras, a la derecha se tiene una vista lateral y frontal de la superficie del arreglo de sensores CCD/CMOS, en la que se proyecta la imagen reflejada por el lente.

Cambiando la distancia del lente al objeto o el tamaño de la apertura del iris se cambia el tamaño del círculo de confusión y provoca que éste afecte un mayor número de sensores en el CCD Figuras 1.5(a) y 1.5(c). Esto es equivalente a desenfocar el objeto de interés al cambiar la ubicación de la profundidad de enfoque. En éste caso, el diámetro permisible de este círculo de confusión para un buen enfoque lo define la resolución del sensor en la cámara. Los sensores CCD usados en cámaras y dispositivos digitales para formar imágenes, se conforman por arreglos de sensores, los cuales han remplazado al material de filmación

de las cámaras. La forma en la que un arreglo de sensores funciona se muestra en la Figura 1.6, donde se muestra cómo el dispositivo captador refleja la luz ambiental de una escena o de alguna fuente de luz sobre el arreglo de sensores del CCD/CMOS.

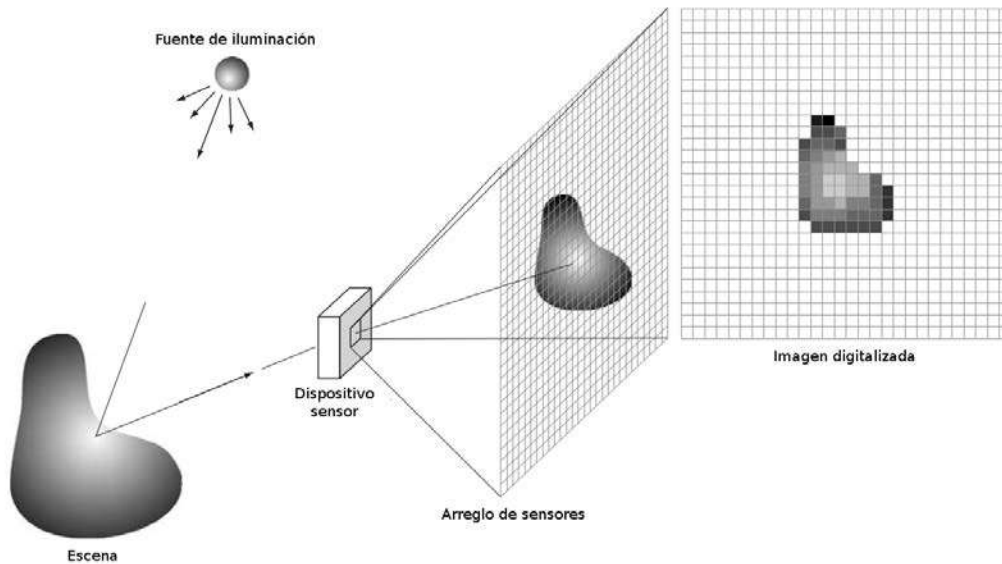


Figura 1.6: Proceso de digitalización de una imagen [Woods, 2008].

El arreglo de sensores se encarga de transformar la luz capturada en una señal eléctrica continua que es muestreada; es decir, se miden valores de la señal a intervalos iguales en las coordenadas vertical y horizontal y después se discretiza la señal muestreada en una imagen digital. El proceso de discretización consiste en tomar la amplitud de la señal en cada muestra, correspondiente a la intensidad de luz que fue capturada en la escena, y asignar un valor correspondiente a la escala de color que se esté usando, e.g., la escala de grises con valores de 0 a 255, comúnmente usada en imágenes digitales [Woods, 2008].

Una imagen digital puede ser definida cómo una matriz I , en la cual cada elemento de la matriz corresponde a un píxel $I(i, j)$ donde i y j son las coordenadas espaciales de la imagen. El valor del píxel de la imagen I en las coordenadas espaciales (i, j) es una cantidad cuyo significado físico está determinado por la fuente de iluminación de la imagen. Una imagen digital está compuesta por un número finito de elementos, conocidos cómo píxeles, cada uno de los cuales tiene una locación y valor particular. Por ejemplo, el valor

de la imagen en el origen es $I(0,0)$ y la próxima coordenada a lo largo del primer renglón es $I(0,1)$. La representación de una imagen como un arreglo numérico de tamaño $M \times N$ se define de la siguiente forma:

$$I = \begin{bmatrix} I(0,0) & I(0,1) & \cdots & I(0,N-1) \\ I(1,0) & I(1,1) & \cdots & I(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ I(M-1,0) & I(M-1,1) & \cdots & I(M-1,N-1) \end{bmatrix}$$

1.1. Definición del problema

El objetivo de la fusión de imágenes es tomar múltiples imágenes, cada una de las cuales contiene un plano de enfoque distinto tomado de la misma escena, definir las áreas con buen enfoque en cada imagen y por último extraerlas y combinarlas en una sola imagen que simule un rango de profundidad de campo infinito por lo que toda la escena estará enfocada. En la imagen de la Figura 1.7 se ejemplifica el objetivo de la fusión. En las Figuras 1.7(a) a 1.7(c) se tiene la misma escena en todas las fotografías en este caso el mecanismo interno de un reloj de pulsera. En cada fotografía se centra la profundidad de campo sobre diferentes componentes del mecanismo, y en cada imagen se presentan detalles del reloj que no se pueden observar en el resto de las imágenes. En la Figura 1.7(d) se tiene la fusión de las imágenes con toda la escena enfocada.

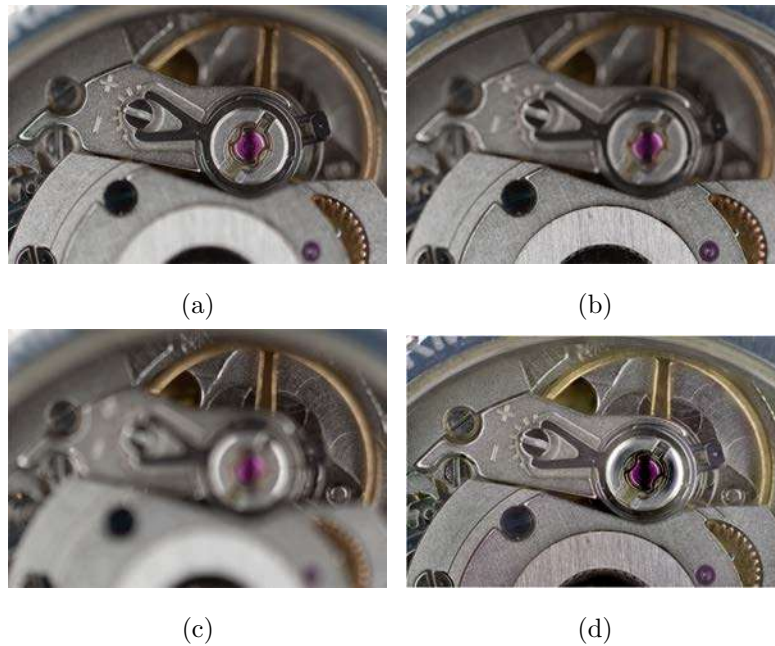


Figura 1.7: Diferentes fotografías del interior de un reloj, en las imágenes (a) hasta (c), fusionadas en una imagen completamente enfocada en (d).

1.1.1. Motivación

De forma práctica el rango limitado de profundidad de campo se presenta de forma más notoria en casos como los de tele detección y de microscopía. En el primer caso, se utilizan de lentes que tienen una profundidad de campo posicionada a una distancia muy lejana al dispositivo, lo que le permite enfocar objetos que se encuentran a largas distancias. Por las propiedades físicas del lente e.g., su curvatura o distancia focal, las cuales le permiten enfocar objetos lejanos, no se puede tener una profundidad de campo muy amplia que abarque a distintos objetos de interés en la escena. Esto representa un problema en sistemas de vigilancia. Este mismo problema se presenta en el caso de microscopía, más específicamente en el campo de medicina como en el análisis o detección de células. Al tratar con escenas microscópicas, el rango de profundidad de campo disminuye junto con el tamaño de la escena.

1.2. Objetivos de la Tesis

En este trabajo se avanza en la solución del problema de limitación de la profundidad de campo por medio de la fusión de imágenes. El método de fusión que se propone utiliza un filtro de frecuencia con el que se definen regiones de alto enfoque en las imágenes que se desean fusionar y en base a este filtro se realiza la composición de la imagen fusionada. También se presenta un método para mejorar el resultado de la fusión utilizando un algoritmo de segmentación de imágenes basado en campos aleatorios de Markov.

1.3. Descripción de Capítulos

- En el Capítulo 2 se presenta una revisión de la literatura de los métodos dominantes en la resolución del problema de fusión de imágenes, principalmente: selección de regiones, descomposición multiescala y aprendizaje.
- En el Capítulo 3 se presenta el método de fusión de imágenes propuesto en esta tesis, el cual está basado en selección de regiones por medio de frecuencia, y se presentan resultados experimentales de su funcionamiento.
- En el Capítulo 4 se describe el algoritmo de segmentación basado en campos aleatorios de Markov, usado para mejorar el método que se propone, incluyendo también resultados experimentales para corroborar esta mejora.
- En el Capítulo 5 se presentan los resultados obtenidos con el método de fusión de imágenes multifoco propuesto, así como una comparación con otros métodos encontrados en el estado del arte.
- Por último, en el Capítulo 6 se presentan las conclusiones derivadas de este trabajo.

Capítulo 2

Estado del arte.

Los esquemas dominantes en la fusión de imágenes multifoco se dividen en tres categorías: métodos basados selección de regiones, métodos basados en descomposiciones multiescala y métodos basados en aprendizaje. En este capítulo se presenta una explicación detallada de estos métodos.

2.1. Métodos basados en selección de regiones

Consisten en hacer la fusión de forma directa sobre las imágenes originales al combinarlas usando pesos apropiados para los píxeles o subregiones en la imagen. Por ejemplo, un método simple es obtener el promedio de las imágenes a fusionar. Teniendo dos imágenes, I_a e I_b , y usando

$$I_c(i, j) = \frac{I_a(i, j) + I_b(i, j)}{2}, \quad (2.1)$$

se consigue la imagen fusionada I_c . Lo que es equivalente a asignar el mismo peso a cada píxel. Ya que el resultado es una combinación de las regiones de alto y bajo enfoque de las imágenes originales, la imagen fusionada presenta tanto los detalles de alto enfoque, como los de bajo enfoque. Este método se usa normalmente como una referencia para comparar otros métodos.

Entre otros métodos de selección de regiones se encuentran el propuesto por Hariharan et al., en [Hariharan et al., 2007], donde las imágenes a fusionar se subdividen o segmentan en regiones, tomando como criterio de segmentación el gradiente de las imágenes, calculado usando aproximaciones con máscaras de Sobel. Al terminar este paso, se toman las regiones que presentan un gradiente más alto entre las imágenes originales, para formar la imagen fusionada. De forma similar Li y Yang, en [Li y Yang, 2008], proponen tomar el promedio (2.1) de las imágenes originales a fusionar para después segmentarlo usando un algoritmo de cortes de grafos. Al finalizar la segmentación, se realiza un corte de las imágenes originales con la forma de cada región obtenida en la segmentación, se mide el enfoque en cada corte y se selecciona el que presenta mayor enfoque para formar la imagen resultante. De la misma forma que Zaveri et al., en [Zaveri et al., 2009], hacen uso de la segmentación con cortes de grafos, pero realizando esta segmentación de forma directa sobre las imágenes originales.

2.1.1. División en bloques

Este es otro método de selección de regiones, usado para fusionar imágenes propuesto por Li et al., en [Li et al., 2001], así como por Goshtasby, en [Goshtasby, 2006]. En ambos trabajos se dividen cada una de las imágenes a fusionar en bloques de píxeles de tamaños iguales. A cada uno de estos bloques se le asigna un valor o peso que califica el nivel de enfoque que contiene; a este nivel lo llaman nivel de actividad. En estos trabajos el nivel de actividad usado es el propuesto por Eskicioglu y Fisher, en [Eskicioglu y Fisher, 1995], quienes proponen la frecuencia espacial (2.2) como un método para establecer la calidad de una imagen. Para una imagen I de tamaño $M \times N$ se puede medir la frecuencia espacial como:

$$S_f = \sqrt{F_r^2 + F_c^2} \quad (2.2)$$

$$F_r = \sqrt{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=1}^{N-1} (I(i, j) - I(i, j-1))^2}$$

$$F_c = \sqrt{\frac{1}{MN} \sum_{i=1}^{M-1} \sum_{j=0}^{N-1} (I(i, j) - I(i-1, j))^2}$$

Donde F_r es la frecuencia espacial por renglones en la imagen y F_c es la frecuencia espacial por columnas. Una frecuencia espacial alta equivale a una imagen con mejor enfoque y una frecuencia espacial baja indica una imagen borrosa o de bajo enfoque. Analizando esta medición se observa que F_r realiza una operación equivalente a convolucionar cada columna en la imagen, con el filtro de convolución pasa altas

$$h(n) = [1, -1], \quad (2.3)$$

y después sumar los valores obtenidos de cada convolución. De la misma forma, para F_c , que es la suma de las convoluciones de cada renglón con (2.3). Se puede ver entonces que F_r y F_c están midiendo las frecuencias en la imagen por medio de un filtro pasa altas, cuyo comportamiento puede ser analizado en el dominio de frecuencias al obtener su transformada de Fourier seguida de la magnitud de la transformada de Fourier:

$$H(k) = \frac{1}{N_F} \sum_{n=0}^{N_F-1} h(n) e^{-j(\frac{2\pi}{N_F})nk}$$

$$H(k) = \frac{1}{N_F} \left[(1) e^{-j(\frac{2\pi}{N_F})(0)k} + (-1) e^{-j(\frac{2\pi}{N_F})(1)k} \right]$$

$$H(k) = \frac{1}{N_F} \left[1 - e^{-j(\frac{2\pi}{N_F})k} \right]$$

La magnitud de la transformada de Fourier de (2.3) es

$$|H(k)| = \frac{1}{N_F} \sqrt{\left(1 - \cos \frac{2\pi k}{N_F}\right)^2 + \left(\operatorname{sen} \frac{2\pi k}{N_F}\right)^2}$$

$$|H(k)| = \frac{1}{N_F} \sqrt{1 - 2 \cos \frac{2\pi k}{N_F} + \cos^2 \frac{2\pi k}{N_F} + \operatorname{sen}^2 \frac{2\pi k}{N_F}}$$

$$|H(k)| = \frac{1}{N_F} \sqrt{2 - 2 \cos \frac{2\pi k}{N_F}}$$

$$|H(k)| = \frac{1}{N_F} \sqrt{4 \sin^2 \frac{\pi k}{N_F}}$$

$$|H(k)| = \frac{2}{N_F} \operatorname{sen} \frac{\pi k}{N_F}$$

En la Figura 2.1, la cual gráfica la magnitud de la transformada de Fourier del filtro pasa altas (2.3), se observa que el filtro atenúa o reduce a cero las frecuencias bajas y permite el paso de las frecuencias altas.

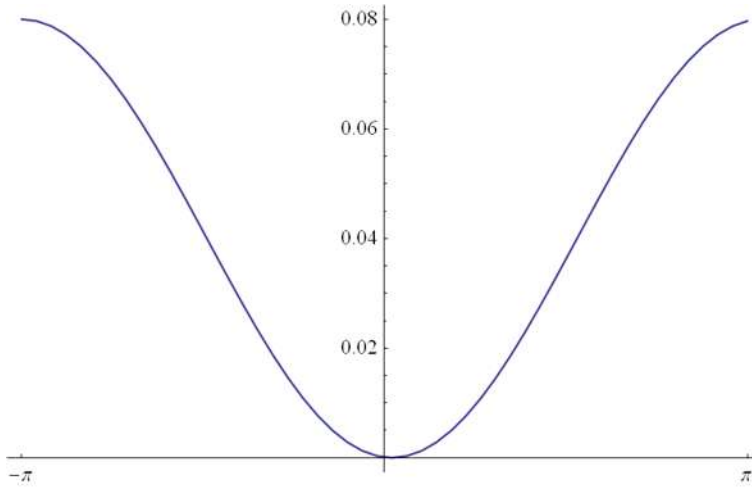


Figura 2.1: $|H(k)|$.

Entonces, la medición de frecuencia espacial definida en (2.2), determina que las regiones con frecuencia espacial alta corresponden a regiones enfocadas y regiones con frecuencia

espacial baja corresponden a regiones no enfocadas. Esto se puede comprobar con el uso de filtros de frecuencia en imágenes; usando un filtro Gaussiano como (2.4), convolucionado con una imagen de modo que sirva como filtro pasa bajas. En la Figura 2.2 se ejemplifica esto. En la Figura 2.2(a) se tiene una imagen sin alterar, y en las Figuras 2.2(b) a 2.2(d) se tienen las convoluciones de la imagen original con la Gaussiana (2.4). Incrementando progresivamente el valor de σ se aumenta el tamaño de la Gaussiana, por lo que en cada imagen se dejan pasar menos frecuencias altas y la imagen empieza a perder nitidez.

$$g(i, j, \sigma) = \frac{1}{2\pi\sigma^2} e^{\left[\frac{-i^2 + j^2}{2\sigma^2} \right]} \quad (2.4)$$



Figura 2.2: Cambio en nivel de nitidez o enfoque de la imagen de Lena, al aumentar el valor de σ en el filtro Gaussiano.

El valor de frecuencia espacial (2.2) como medición del enfoque de las imágenes de la Figura 2.2, se puede verificar en la Tabla 2.1. Conforme disminuye el nivel de nitidez de la imagen aumentando el valor de σ , disminuye el valor de frecuencia espacial.

Imagen	Frecuencia espacial
Original	20.27164649181851
$\sigma = 1.0$	8.86104014891999
$\sigma = 2.0$	5.795931008533641
$\sigma = 3.0$	4.455165079634691

Tabla 2.1: Valores de frecuencia espacial de la imagen Lena, convolucionada con una Gaussiana de diferentes valores de σ .

Si se tienen dos imágenes I_a e I_b , se denota cada bloque en la división con un índice d , cómo $I_{a,d}$ e $I_{b,d}$. Después, con una función $A(I)$ que mide el nivel de actividad de cada bloque, por ejemplo la frecuencia espacial (2.2), la fusión consiste entonces en comparar cada bloque en el mismo índice d correspondiente y tomar el bloque con el mejor nivel de actividad para formar la imagen resultante I_c . Este procedimiento se resume de la siguiente forma:

$$I_{c,d} = \begin{cases} I_{a,d} & \text{si } A(I_{a,d}) > A(I_{b,d}) \\ I_{b,d} & \text{si } A(I_{a,d}) < A(I_{b,d}) \\ (I_{a,d} + I_{b,d})/2 & \text{en caso contrario.} \end{cases}$$

En el diagrama de la Figura 2.3 se tiene el procedimiento de fusión de imágenes usando división de bloques.

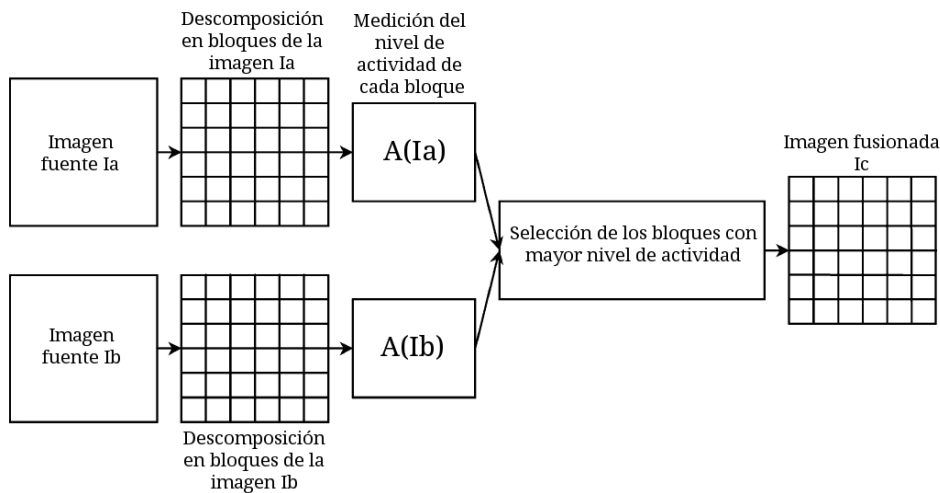


Figura 2.3: Proceso de fusión de imágenes usando división en bloques.

Uno de los parámetros que afecta a este método es el tamaño del bloque, Goshtasby menciona que al disminuir el tamaño de los bloques, se obtienen más bloques por imagen por lo que el proceso se vuelve lento. Por el contrario, al aumentar el tamaño de los bloques se obtienen menos de estos y el proceso se vuelve más rápido, pero puede que el tamaño del bloque no sea lo suficientemente grande para incluir detalles pequeños pero de alto enfoque

en la imagen [Goshtasby, 2006]. Dentro de los diferentes tipos de nivel de actividad que se pueden encontrar está por ejemplo el propuesto por Maruthi, que en el mismo procedimiento de división de bloques en [Maruthi, 2011], selecciona los bloques más enfocados en base a un nivel de actividad llamado nivel de difusión, mientras que Qingping usa un nivel de actividad basado en la medición del nivel de dimensión fractal, en [Li et al., 2013].

2.2. Métodos basados en aprendizaje

Algunos de los métodos que se encuentran dentro de esta categoría es el que usa una red neuronal para clasificar imágenes como enfocadas o borrosas, Li et al., en [Li et al., 2002] así como Madhavi y Ashok en [Madhavi y K Ashok, 2011] presentan la fusión de imágenes usando una red neuronal junto con el método de división en bloques ya mencionado, de modo que se dividen las imágenes en bloques de tamaños iguales y se forma la imagen fusionada tomando los bloques con mejor enfoque. La diferencia consiste en el uso de redes neuronales para discernir cuales son los bloques con mejor enfoque. En esta sección se muestra cómo se usa una red neuronal como clasificador, y cómo adaptarla para clasificar imágenes enfocadas.

La red neuronal más simple consiste de un perceptrón, el cual está compuesto por entradas que pueden ser valores o mediciones correspondientes a los objetos a clasificar o pueden ser las mismas salidas de otros perceptrones en el caso de una red neuronal más extendida. A cada entrada del perceptrón $x_j \in \mathbb{R}, j = 1, \dots, Q$, le corresponde el valor de un peso de conexión $w_j \in \mathbb{R}$, y la salida del perceptrón es la suma ponderada de las entradas, como se hace en (2.5), w_0 es un peso correspondiente a una unidad extra x_0 conocido como el bias. El bias es un elemento que permite al modelo ser más general, usualmente tomado como $x_0 = +1$.

$$y = \sum_{j=1}^Q w_j x_j + w_0 \quad (2.5)$$

La red entonces requiere aprender los pesos \mathbf{w} , que son los parámetros de la red neuronal, tales que se obtengan salidas correctas para los valores de entrada del sistema. Se puede ver el resultado de (2.5) cómo un producto punto,

$$y = \mathbf{w}^T \mathbf{x},$$

donde $\mathbf{w} = [w_0, w_1, \dots, w_Q]^T$ y $\mathbf{x} = [1, x_1, \dots, x_Q]^T$. El entrenamiento de la red representa calcular los valores de los pesos \mathbf{w} . En el caso más simple de la red, cuando se tiene $Q = 1$ y \mathbf{x} es un valor a clasificar, se tiene:

$$y = wx + w_0.$$

La cual es la ecuación de la recta con w cómo la pendiente y w_0 cómo la distancia al origen. Por lo tanto este perceptrón de una entrada y una salida puede ser usado para implementar una clasificación lineal. Con más de una entrada, la línea se transforma en un hiperplano y puede ser usado para implementar una clasificación multivariable. Al usar al perceptrón cómo una función discriminante lineal, el perceptrón puede separar dos clases al revisar el signo de su salida. Si se define una función umbral U cómo:

$$U(a) = \begin{cases} 1 & \text{si } a > 0 \\ 0 & \text{si no.} \end{cases}$$

Y la clase resultante que corresponde a un vector de entrada se escoge de modo que

$$\text{Clase} = \begin{cases} \text{Clase 1} & \text{si } U(\mathbf{w}^T \mathbf{x}) > 0 \\ \text{Clase 2} & \text{si no.} \end{cases}$$

Cuando se tienen $P > 2$ salidas, se tienen P perceptrones, ver Figura 2.4, y cada perceptrón tiene un vector de pesos w_i . La salida de cada perceptron se calcula de forma que

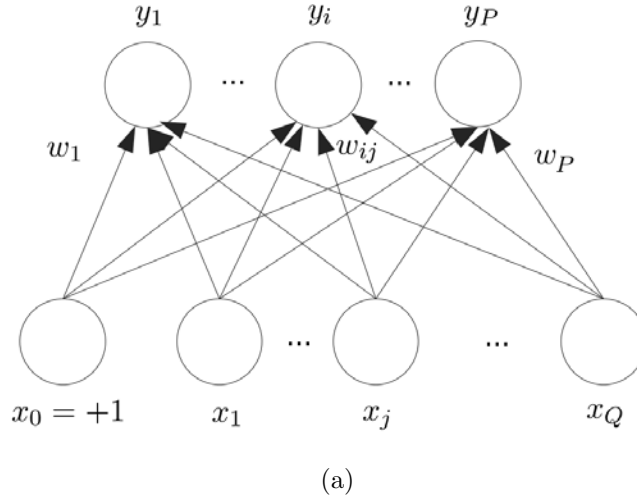


Figura 2.4: Perceptrón con P salidas, cada salida es una suma ponderada de las entradas x_j con los pesos w_{ij} .

$$y_i = \sum_{j=1}^Q w_{ij} x_j + w_{i0},$$

donde w_{ij} es el peso de la entrada x_j a la salida y_i . Cuando se clasifica un vector de entrada se escoge la Clase i , si,

$$y_i = \max_{1 < i \leq P} (y_P).$$

Los conceptos de perceptrón y red neuronal mostrados en esta sección se pueden encontrar en [Alpaydin, 2009]. Para clasificar imágenes usando la red neuronal, se requiere extraer características de las imágenes con las cuales se puedan formar los vectores de entrada para la red neuronal. Las características usadas pueden consistir de valores como el nivel de actividad en la imagen, ya mencionado en secciones anteriores, o bien por otro tipo de características como detección de bordes, o combinaciones de diversas características. Para cada imagen se determina un vector $\mathbf{v} = [c_1, c_2, \dots, c_n]^T$ donde c_i es la i -ésima característica extraída de la imagen. Si se tienen dos imágenes I_a e I_b , se tienen sus vectores de características respectivos \mathbf{v}_a y \mathbf{v}_b .

La red requiere de un conjunto de entrenamiento, compuesto de imágenes de ejemplo etiquetadas como borrosas o enfocadas. Cada imagen del conjunto tiene un vector de características correspondiente con los que se entrena la red. Una vez que la red tiene una generalización adecuada para el conjunto de imágenes de entrenamiento, es decir, puede clasificar correctamente imágenes enfocadas y no enfocadas, se puede usar el mismo método de división de bloques visto en la Sección 2.1.1 seleccionando cada bloque con la red neuronal. Ejemplos de este método son propuestos por Gao et al., en donde se entrena una red neuronal para clasificar bloques de imágenes [Gao et al., 2005], y el vector de características es formado con el promedio de gradiente de cada bloque.

2.3. Métodos basados en descomposición multi escala

Cómo su nombre indica, en este tipo de métodos primero se transforma las imágenes originales en alguna descomposición y se realiza la fusión de imágenes sobre las versiones transformadas de estas imágenes. Entre los métodos basados en transformaciones se encuentran la transformada discreta wavelet y la pirámide Laplaciana, los cuales descomponen las imágenes en múltiples resoluciones y diferentes bandas de frecuencias, lo que permite tener un análisis más detallado de la imagen.

2.3.1. Pirámide Laplaciana

La pirámide Laplaciana es una descomposición una imagen en copias de diferentes escalas y resoluciones, la cual fue propuesta por Burt y Adelson en [Burt y Adelson, 1983]. Se denomina piramidal por la forma en la que se realiza un submuestreo en cada escala de la pirámide, formando una imagen más pequeña en cada nivel. La formación de la pirámide Laplaciana inicia generando una pirámide Gaussiana. Tomando G_z cómo el z -ésimo nivel de la pirámide Gaussiana de una imagen I , el primer nivel o la base de la pirámide Gaussiana consiste en la imagen original, es decir,

$$G_0 = I,$$

y para niveles superiores en la pirámide Gaussiana, $z > 0$, se realiza una convolución del

nivel inferior en la pirámide con un filtro Gaussiano g , definido en (2.4),

$$G_z = [g * G_{z-1}]_{\downarrow 2}.$$

La notación $[...]_{\downarrow 2}$ indica un submuestreo de la imagen por un factor de 2, eliminando columnas y renglones pares en la imagen. Una vez que se tiene la pirámide Gaussiana se puede calcular la pirámide Laplaciana, denotando como L_z el nivel z de la pirámide Laplaciana. Esta se define como la diferencia entre niveles sucesivos de la pirámide Gaussiana, donde el último nivel Z de la pirámide Laplaciana es igual al último nivel de la pirámide Gaussiana:

$$L_z = G_z - g * [G_{z+1}]_{\uparrow 2}$$

$$L_Z = G_Z$$

Aquí, $[...]_{\uparrow 2}$ representa un escalamiento de la imagen por un factor de 2, insertando ceros entre las columnas y renglones de la imagen. La convolución de esta imagen escalada con el kernel g tiene el efecto de interpolar los valores de las columnas faltantes. La imagen original es recuperada revirtiendo los pasos anteriores, si se toma la imagen recuperada como \hat{G} y haciendo uso de todos los niveles de la pirámide Laplaciana, así como el nivel superior de la pirámide Gaussiana de Z niveles

$$I = \hat{G}_0$$

$$\hat{G}_Z = L_Z$$

$$\hat{G}_z = L_z + g * [\hat{G}_{z+1}]_{\uparrow 2} \quad (2.6)$$

Aplicaciones consecutivas de (2.6) producen la imagen \hat{G}_0 , que es la versión reconstruida de la imagen original I . La formación de la pirámide Laplaciana se ejemplifica en la Figura 2.5. A la izquierda se tiene una pirámide Gaussiana y a partir de diferencias entre los niveles de la pirámide Gaussiana se forma la pirámide Laplaciana mostrada a la derecha.

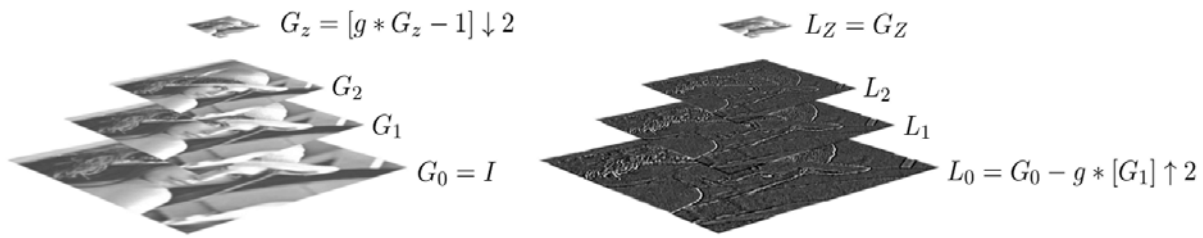


Figura 2.5: Formación de la pirámide de Laplace a partir de una pirámide Gaussiana.

Originalmente Burt y Adelson presentaron esta pirámide como un método de compresión de imágenes, y fueron Burt y Kolczynski, quienes la usaron como un método de fusión de imágenes en [Burt y Kolczynski, 1993]. Otro trabajo de fusión de imágenes que usa la pirámide Laplaciana es el de Wang y Chang en [Wang y Chang, 2011]. La fusión usando pirámides Laplacianas como se muestra en la Figura 2.6 se realiza en tres pasos, el primero consiste en descomponer las imágenes originales a fusionar en sus versiones multiescala piramidales. Después, usando como criterio el nivel de actividad de las pirámides obtenidas en el primer paso, se forma una pirámide compuesta, de forma similar al método de división de bloques. Por último se obtiene la imagen fusionada tomando la pirámide compuesta resultante del paso anterior y reconstruyéndola con la Ecuación (2.6).

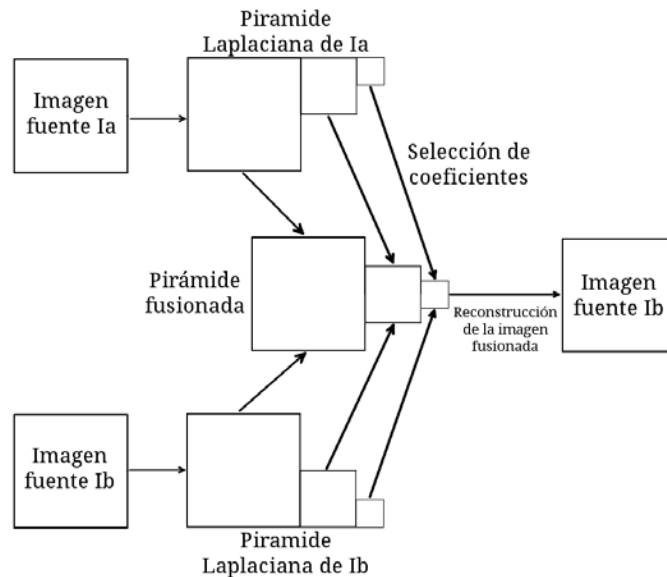


Figura 2.6: Proceso de fusión de imágenes usando la pirámide Laplaciana.

La operación realizada al obtener un nivel L_z en la pirámide de Laplace, es decir, el sustraer de un nivel G_z la imagen en el nivel superior G_{z+1} corresponde a un filtrado de frecuencias usando una diferencia de Gaussianas. En consecuencia, los diferentes niveles en la pirámide Laplaciana permiten presentar los detalles de frecuencia en diferentes resoluciones. Toet hace una modificación de esta pirámide y la presenta en su método de fusión de imágenes en [Toet, 1989], en el cual se forma la pirámide usando la división de niveles consecutivos en la pirámide Gaussiana, en lugar de la sustracción.

2.3.2. Transformada Discreta Wavelet

Por último, con la descomposición wavelet, se obtiene una representación de la imagen en diferentes resoluciones y en cada resolución se tiene información de la imagen en el dominio espacial, así como en el dominio de frecuencia. La formulación que se presenta a continuación, se puede revisar en el método de fusión de imágenes usando la misma transformada discreta wavelet propuesto por Pajares y De la Cruz en [Pajares y De la Cruz, 2004]. La descomposición consiste en la convolución de los wavelets Haar que se presentan como kernels de convolución: pasa bajas, en (2.7), y pasa altas, en (2.8), sobre la imagen que se desea transformar.

$$l = \frac{1}{\sqrt{2}}[1, 1] \quad (2.7)$$

$$h = \frac{1}{\sqrt{2}}[1, -1] \quad (2.8)$$

Los wavelets Haar son ortogonales, lo que les da una propiedad importante, la cual es la conservación de la información al realizar el proceso de descomposición inverso o síntesis, lo que quiere decir que después de realizar la transformada wavelet se puede regresar o reconstruir la imagen original sin pérdida de información. Stollnitz y Derose usan esta misma propiedad para usar la transformada discreta wavelet como un método de compresión de imágenes en [Stollnitz y Derose, 1995].

La descomposición o transformación discreta wavelet consiste en la convolución

por renglones de una imagen I con los wavelets en (2.8) y (2.7), seguido de un submuestreo o eliminación de las columnas par de las convoluciones resultantes:

$$I_L = [I * l]_{\downarrow 2col},$$

$$I_H = [I * h]_{\downarrow 2col},$$

donde $[\dots]_{\downarrow 2col}$ es un submuestreo de la imagen en el que se descartan la mitad de las columnas. Las dos imágenes resultantes son las respuestas a filtros pasa altas y pasa bajas de la imagen original en una dimensión. Al terminar este paso se vuelve a realizar una convolución con los wavelets ahora por columnas de las imágenes obtenidas en el último paso, I_L e I_H , seguido de un submuestreo de los resultados ahora eliminando los renglones par.

$$I_{LL} = [I_L * l]_{\downarrow 2ren}$$

$$I_{LH} = [I_L * h]_{\downarrow 2ren}$$

$$I_{HL} = [I_H * l]_{\downarrow 2ren}$$

$$I_{HH} = [I_H * h]_{\downarrow 2ren}$$

La Figura 2.7 contiene un diagrama que describe el proceso de un nivel descomposición de la transformada discreta wavelet. Al finalizar este proceso se obtienen cuatro imágenes que corresponden a las respuestas de los wavelets Haar. En la Figura 2.8(a), se presenta una imagen inicial y la imagen de la Figura 2.8(b) presenta su transformada discreta indicando las diferentes bandas de frecuencia que se obtienen en la descomposición: I_{LL} corresponde a las frecuencias bajas de la imagen original, la imagen I_{LH} corresponde a los detalles horizontales de la imagen y las imágenes I_{HL} e I_{HH} corresponden a los detalles verticales y diagonales respectivamente.

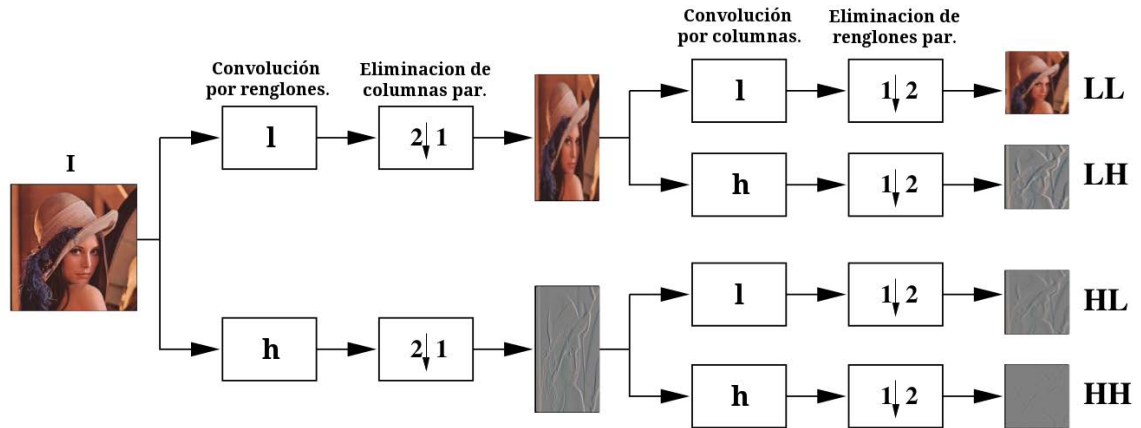


Figura 2.7: Proceso de descomposición de una imagen con los wavelets Haar.

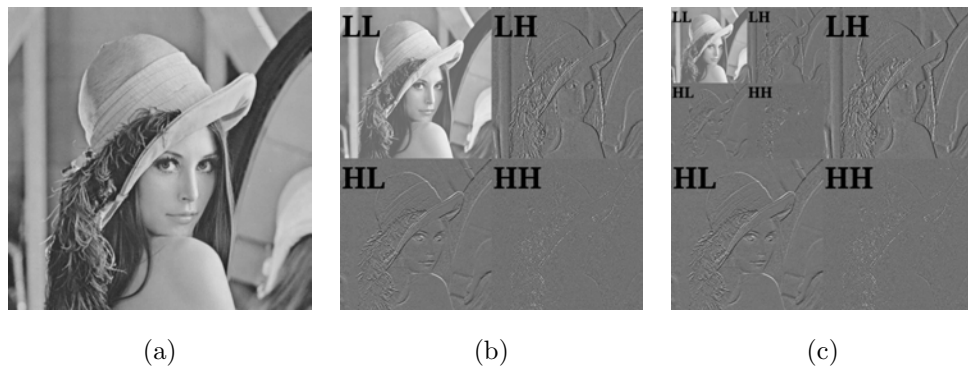


Figura 2.8: Múltiples bandas de frecuencia y resoluciones obtenidas con la transformada discreta wavelet. a) imagen original, b) descomposición wavelet de un nivel y c) descomposición wavelet de dos niveles.

Descomposiciones subsecuentes se realizan usando solamente la imagen I_{LL} por lo que la descomposición wavelet produce una descomposición piramidal en múltiples resoluciones Figura 2.8(c). La presencia de un análisis de frecuencia en diferentes direcciones hace a la transformada wavelet una alternativa a la transformada de Fourier. El proceso inverso a la transformada wavelet o la síntesis consiste en tomar las imágenes obtenidas en el proceso de descomposición y realizar primero, un escalamiento insertando ceros en renglones alternados, seguido de la convolución con los wavelets Haar:

$$\begin{aligned}
I'_{LL} &= [[I_{LL}]_{\uparrow 2ren}] * l \\
I'_{LH} &= [[I_{LH}]_{\uparrow 2ren}] * h \\
I'_{HL} &= [[I_{HL}]_{\uparrow 2ren}] * l \\
I'_{HH} &= [[I_{HH}]_{\uparrow 2ren}] * h
\end{aligned}$$

subsecuentemente se realiza un escalamiento insertando ceros en columnas alternadas y de nuevo la convolución con los wavelets Haar para finalmente sumar las imágenes resultantes:

$$\begin{aligned}
I''_{LL} &= [[I'_{LL}]_{\uparrow 2col}] * l \\
I''_{LH} &= [[I'_{LH}]_{\uparrow 2col}] * l \\
I''_{HL} &= [[I'_{HL}]_{\uparrow 2col}] * h \\
I''_{HH} &= [[I'_{HH}]_{\uparrow 2col}] * h \\
I &= I''_{LL} + I''_{LH} + I''_{HL} + I''_{HH}
\end{aligned}$$

De forma simplificada, el proceso de síntesis o reconstrucción de la imagen, consiste en tomar las cuatro imágenes obtenidas en la descomposición, y realizar un escalamiento y convolución con los wavelets Haar en el orden inverso a la descomposición. Si la descomposición original consistió en múltiples niveles, se realiza la síntesis iniciando en el nivel de descomposición más alto obteniendo la imagen I_{LL} del nivel correspondiente y usando ésta para la síntesis del nivel inferior.

La fusión de imágenes usando este esquema de descomposición consiste, de la misma forma que con la pirámide Laplaciana, en obtener la transformada wavelet discreta de cada imagen a fusionar y seleccionar los coeficientes con nivel de actividad más alto en cada pirámide para formar una pirámide compuesta, en este paso Nava et al., seleccionan estos coeficientes haciendo uso del detector de bordes Canny [Nava et al., 2008] cómo el nivel de actividad, mientras que Zhang y Blum presentan en [Zhong y Blum, 1999], una comparación de diferentes métodos de selección de coeficientes para las fusiones usando la transformada discreta wavelet. La selección de coeficientes se realiza tomando cómo nivel

de actividad el píxel con mayor valor absoluto o con mayor valor en el promedio de una ventana de píxeles. Una vez que se seleccionaron los coeficientes, se realiza la síntesis de esta pirámide compuesta y se obtiene la imagen fusionada resultante. En [Li et al., 2010] se presenta un método similar de fusión usando la transformada wavelet. Sin embargo, en este esquema no se utiliza submuestreo, ya que los autores argumentan que éste representa una pérdida de detalles en la imagen. Por tal motivo llaman a la descomposición wavelet sin submuestreo, la transformada wavelet redundante. Indhumadhi y Padmavathi hacen uso del concepto de pirámide Laplaciana junto con la transformada discreta wavelet en [Indhumadhi y Padmavathi, 2011], donde en un procedimiento similar construyen una pirámide híbrida entre los métodos de descomposición wavelet y pirámide Laplaciana, con la cual hacen la fusión de imágenes.

2.4. Conclusiones

Dentro de los métodos vistos en este capítulo se pueden mencionar aspectos comunes entre ellos, en especial el uso de una medición del nivel de actividad como un valor que determina el nivel de enfoque en la imagen. También se presentan en estos métodos el uso de algún tipo de análisis de frecuencia para conseguir este nivel de actividad. Esto se observó en el nivel de actividad por frecuencia espacial en los métodos de selección de regiones y de aprendizaje. Otra forma en la que se presentan los filtros de frecuencia, se puede observar en el caso de los métodos de descomposición multiescala en la convolución con un filtro Gaussiano en la pirámide Laplaciana, o los filtros pasa altas y pasa bajas en la descomposición wavelet. Por esta razón el método de fusión de imágenes propuesto en el Capítulo 3, se basa en la selección de regiones de alto enfoque, las cuales se determinan por medio del filtrado de frecuencias en la imagen.

Capítulo 3

Selección de regiones de alto enfoque, por medio de frecuencia.

En este capítulo se realiza una comparación de funciones que son usadas como filtros de frecuencia y que están asociadas a los métodos de fusión de imágenes revisados en el capítulo 2. También se propone un filtro pasa altas para fusión de imágenes, el cual por medio de las pruebas realizadas, se demuestra que tiene una mayor precisión y mejor capacidad para fusionar imágenes multifoco, en comparación con otros filtros que se encuentran asociados con diferentes métodos de fusión. Recordando los métodos revisados en el capítulo 2, se identifica en éstos, alguna forma de medición del nivel de actividad en las imágenes originales que serán fusionadas. Este nivel de actividad pretende determinar, en el caso de fusión de imágenes multifoco, áreas enfocadas en la imagen y lo hace por medio de una medición de las frecuencias altas en las imágenes a fusionar.

Así, por ejemplo, revisando el método propuesto por Maruthi, en [Maruthi, 2011], donde subdivide las imágenes en bloques del mismo tamaño y midiendo el nivel de actividad en cada bloque, hace una selección de las regiones con mayor enfoque para formar la imagen final, este nivel de actividad en realidad es una medición de frecuencias altas en dichos bloques. Para el caso de redes neuronales, mostrando como ejemplo al trabajo de Li et al., en [Li et al., 2002], donde los autores forman vectores de características extraídas de las imágenes, para entrenar a una red neuronal que permita clasificar imágenes borrosas o

enfocadas, estas características estaban compuestas también de una medición de frecuencias en la imagen.

Analizando el proceso de generación de la pirámide Laplaciana, se vuelve evidente el uso de la diferencia de Gaussianas en ella, es decir cada nivel en la pirámide está dado por la convolución de una imagen con la diferencia de dos Gaussianas de diferentes tamaños, seguido de un submuestreo del resultado, la diferencia de Gaussianas se usa entonces cómo en los métodos anteriores, para presentar respuestas de alta frecuencia de las imágenes a fusionar, y con estas respuestas obtener un nivel de actividad para determinar las regiones de alto enfoque. Ya que la pirámide Laplaciana usa la diferencia de Gaussianas, la cual también es usada para aproximar el Laplaciano de una Gaussiana, es de interés hacer una comparación de ambos cuando son usados cómo filtros de frecuencia.

Por último, los wavelets Haar usados en la transformada discreta wavelet, también hacen uso de un filtro pasa altas en la forma de uno de los wavelets Haar cómo se mencionó en el estado del arte, se presenta una comparación con estos Wavelets debido a que la transformada wavelet discreta es usada con frecuencia en la literatura de fusión de imágenes. Por lo tanto, esta transformada nos interesa cómo referencia para medir la eficacia del filtro pasa altas que se propone en esta tesis para fusión de imágenes. Después de revisar todos estos métodos se llega a la conclusión de que la detección de altas frecuencias forma una parte crucial en los métodos de fusión de imágenes, y que todos estos usan algún filtro pasa altas.

3.1. Fusión de imágenes usando un filtro pasa altas

Dadas dos imágenes multifoco, I_a e I_b , se pretende realizar la fusión de la siguiente forma. Primero se calcula la convolución de las imágenes originales con un filtro pasa altas h (3.1) para obtener las respuestas de frecuencias de las imágenes originales con las que se medirá su nivel de actividad, el cual mide la calidad y el enfoque de la imagen. El nivel de actividad se determina con el valor absoluto de la respuesta de altas frecuencias, cómo se propone en [Zhong y Blum, 1999].

$$\begin{aligned}\hat{I}_a &= I_a * h \\ \hat{I}_b &= I_b * h\end{aligned}\tag{3.1}$$

Las imágenes resultantes, \hat{I}_a e \hat{I}_b , son comparadas entre si, píxel a píxel para seleccionar los píxeles con el nivel de actividad más alto para formar la imagen fusionada I_c , esto mediante,

$$I_c(i, j) = \begin{cases} I_a(i, j) & \text{si } |\hat{I}_a(i, j)| > |\hat{I}_b(i, j)| \\ I_b(i, j) & \text{en caso contrario} \end{cases}\tag{3.2}$$

Este método de fusión se resume en el Algoritmo 1, y está basado en el método propuesto por Li et al., [Li et al., 2001], así cómo por Maruthi [Maruthi, 2011] que consiste en división de las imágenes en bloques y selección de los mismos en base a un valor de nivel de actividad.

Algoritmo 1: Fusión de imágenes por medio de un filtro h .

Entrada: I_a, I_b

Salida: I_c

/* Convolucionar imágenes a fusionar I_a e I_b , con un filtro pasa altas h , Ecuación (3.1). */

1 $\hat{I}_a \leftarrow I_a * h$

2 $\hat{I}_b \leftarrow I_b * h$

/* Hacer la composición de la imagen fusionada, comparando el nivel de actividad, Ecuación (3.2). */

3 **para** $i \leftarrow 0$ **hasta** $N - 1$ **hacer:**

4 **para** $j \leftarrow 0$ **hasta** $M - 1$ **hacer:**

5 **si** $|\hat{I}_a(i, j)| > |\hat{I}_b(i, j)|$ **entonces**

6 $I_c(i, j) \leftarrow I_a(i, j)$

7 **si no**

8 $I_c(i, j) \leftarrow I_b(i, j)$

9 **fin de si**

10 **fin de para**

11 **fin de para**

Una forma alternativa de implementar la operación (3.1), es haciendo uso del teorema de convolución, con el cual se presenta la equivalencia entre las operaciones de convolución y la multiplicación en los dominios espacial y de frecuencia. El teorema establece que una convolución en el dominio espacial es equivalente a una multiplicación en el dominio de frecuencia, así como una multiplicación en dominio espacial equivale a una convolución en el dominio de frecuencia. Dada una imagen I , un filtro h , y sus respectivas transformadas de Fourier:

$$I(i, j) \xleftrightarrow{\mathcal{F}} F(u, v), \quad (3.3)$$

$$h(i, j) \xleftrightarrow{\mathcal{F}} H(u, v) \quad (3.4)$$

el teorema de convolución se describe de manera formal cómo:

$$I(i, j) * h(i, j) \xleftrightarrow{\mathcal{F}} F(u, v)H(u, v)$$

$$I(i, j)h(i, j) \xleftrightarrow{\mathcal{F}} F(u, v) * H(u, v)$$

Se puede demostrar este teorema, tomando la convolución de dos señales $x(n) * y(n)$ cómo:

$$g(n) = \sum_{m=0}^{N_F-1} x(m)y(n-m)$$

y la transformada de Fourier de $g(n)$ cómo:

$$G(k) = \frac{1}{N_F} \sum_{n=0}^{N_F-1} [g(n)]e^{-j(\frac{2\pi}{N_F})nk}$$

$$G(k) = \frac{1}{N_F} \sum_{n=0}^{N_F-1} \left[\sum_{m=0}^{N_F-1} x(m)y(n-m) \right] e^{-j(\frac{2\pi}{N_F})nk}$$

cambiando el orden de la sumatoria

$$G(k) = \frac{1}{N_F} \sum_{m=0}^{N_F-1} x(m) \left[\sum_{n=0}^{N_F-1} y(n-m) e^{-j(\frac{2\pi}{N_F})nk} \right]$$

haciendo $\hat{n} = n - m$

$$G(k) = \frac{1}{N_F} \sum_{m=0}^{N_F-1} x(m) \left[\sum_{\hat{n}=0}^{N_F-1} y(\hat{n}) e^{-j(\frac{2\pi}{N_F})(\hat{n}+m)k} \right]$$

$$G(k) = \frac{1}{N_F} \sum_{m=0}^{N_F-1} x(m) e^{-j(\frac{2\pi}{N_F})mk} \left[\sum_{\hat{n}=0}^{N_F-1} y(\hat{n}) e^{-j(\frac{2\pi}{N_F})\hat{n}k} \right]$$

$$G(k) = N_F \left[\frac{1}{N_F} \sum_{m=0}^{N_F-1} x(m) e^{-j(\frac{2\pi}{N_F})mk} \right] \left[\frac{1}{N_F} \sum_{\hat{n}=0}^{N_F-1} y(\hat{n}) e^{-j(\frac{2\pi}{N_F})\hat{n}k} \right]$$

$$G(k) = N_F X(k) Y(k)$$

donde la multiplicación se realiza elemento a elemento, esta relación nos permite entonces tomar una señal I y un filtro h , obtener sus transformadas de Fourier respectivas (3.3) y (3.4), después realizar una multiplicación elemento a elemento

$$\hat{F}(u, v) = F(u, v) H(u, v)$$

y después usar la transformada de Fourier inversa sobre el resultado:

$$\hat{F}(u, v) \xrightarrow{\mathcal{F}^{-1}} \hat{I}(i, j)$$

donde \hat{I} es equivalente a la imagen I filtrada con h , utilizando una convolución como (3.1). Por último, este teorema permite usar la transformada de Fourier para observar de manera más intuitiva el comportamiento que tiene un filtro de frecuencia, ya que se puede apreciar

a simple vista la forma que tiene, y cuales valores de frecuencia atenúa o reduce a 0 y cuales permite pasar, así cómo definir si se trata de un filtro pasa altas, pasa bajas o pasa banda. Aun cuando el filtrado se realice por medio de convolución en el dominio espacial, se puede observar la magnitud de la transformada de Fourier del filtro, lo que nos permite determinar su comportamiento.

3.2. Filtros de frecuencia asociados a los métodos de fusión de imágenes

La clasificación de un filtro cómo pasa altas, pasa bajas o pasa banda, se asigna por el efecto que tiene sobre los valores de frecuencia en una señal, en esta sección se observa la magnitud de la transformada de Fourier de las funciones usadas cómo filtros pasa altas, de modo que pueda servir cómo indicador de las frecuencias que afecta, un ejemplo de esto se muestra a continuación.

En la Figura 3.1(b) se presenta un ejemplo de un filtro de frecuencia correspondiente a una Gaussiana g , representado por la magnitud de su transformada discreta de Fourier. En esta representación de frecuencias los valores en el centro de la imagen corresponden a valores de baja frecuencia mientras que alejándose del centro se encuentran las altas frecuencias. Además, las áreas negras corresponden a un valor numérico de 0, por lo que áreas con este color corresponden a una atenuación de las frecuencias por el filtro. En este filtro solo los coeficientes de alta frecuencia en los bordes del espectro de frecuencias se reducen a 0, por lo que el filtro se denomina cómo un pasa bajas. Calculando la convolución entre g y la imagen I de la Figura 3.1(a), se obtiene la respuesta \hat{I} mostrada en la Figura 3.1(c), en la que, por la pérdida de altas frecuencias, se tiene una imagen borrosa. En la imagen 3.1(d) se tiene otro filtro en el cual únicamente los coeficientes de bajas frecuencias se hacen 0, en este caso se tiene un filtro pasa altas. Este último es el filtro que se propone en este trabajo para determinar las regiones con alto enfoque y del cual se darán más detalles la Sección 3.3.

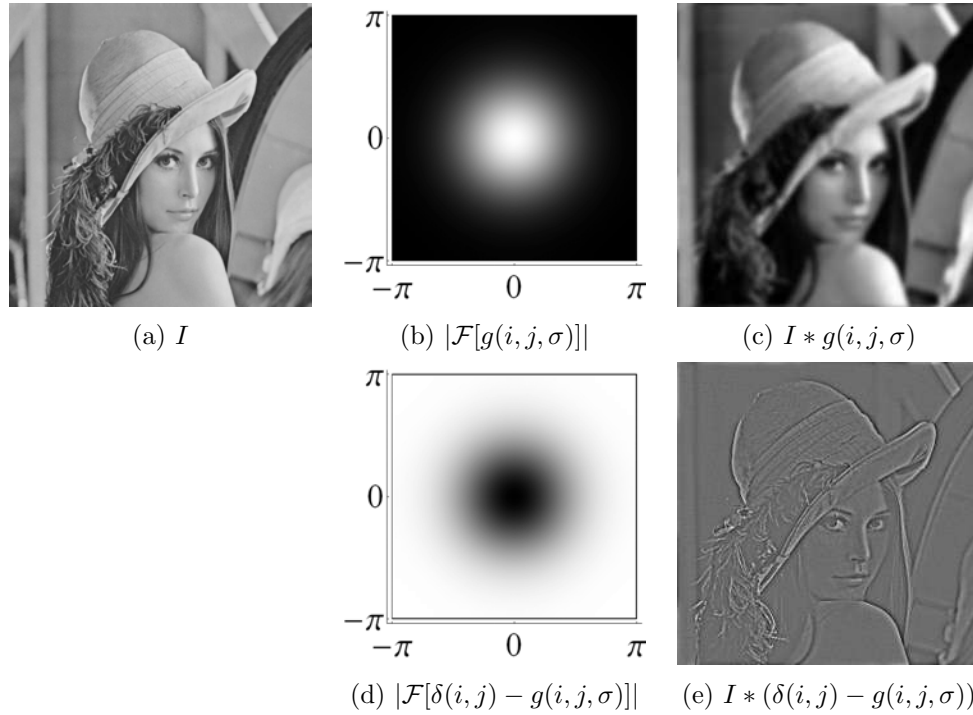


Figura 3.1: a) Imagen original a filtrar, b) y d) magnitud de la transformada de fourier de los filtros, c) y e) sus respectivas respuestas.

3.2.1. Laplaciano de una Gaussiana

El Laplaciano de una Gaussiana (LoG), se obtiene por medio del operador Laplaciano, definido cómo la divergencia del gradiente. Generalmente se denota de la forma: $(\nabla \cdot \nabla)\phi = \nabla^2\phi$ cuando ϕ se trata de una función escalar. El operador ∇ está definido para un sistema de coordenadas cartesiano cómo las primeras derivadas parciales de una función y se conoce cómo el gradiente, entonces para una Gaussiana $g(i, j, \sigma)$, el gradiente corresponde a:

$$\nabla g(i, j, \sigma) = \left[\frac{\partial g(i, j, \sigma)}{\partial i}, \frac{\partial g(i, j, \sigma)}{\partial j} \right].$$

Usando el producto punto sobre el operador ∇ , se obtiene el operador Laplaciano, que consiste en la suma de las segundas derivadas parciales.

De nuevo con una Gaussiana $g(x, y, \sigma)$:

$$\begin{aligned}\nabla^2 g(i, j, \sigma) &= (\nabla \cdot \nabla)g(i, j, \sigma) = \left[\frac{\partial}{\partial i}, \frac{\partial}{\partial j} \right] \cdot \left[\frac{\partial}{\partial i}, \frac{\partial}{\partial j} \right] g(i, j, \sigma), \\ \nabla^2 g(i, j, \sigma) &= (\nabla \cdot \nabla)g(i, j, \sigma) = \left[\frac{\partial}{\partial i} \left(\frac{\partial g(i, j, \sigma)}{\partial i} \right) + \frac{\partial}{\partial j} \left(\frac{\partial g(i, j, \sigma)}{\partial j} \right) \right], \\ \nabla^2 g(i, j, \sigma) &= \frac{\partial^2 g(i, j, \sigma)}{\partial i^2} + \frac{\partial^2 g(i, j, \sigma)}{\partial j^2},\end{aligned}$$

donde las primeras y segundas derivadas parciales de g , Ecuación (2.4), corresponden a:

$$\begin{aligned}\frac{\partial g(i, j, \sigma)}{\partial i} &= \frac{1}{2\pi\sigma^2} e^{\left[\frac{-(i^2+j^2)}{2\sigma^2} \right]} \left(-\frac{i}{\sigma^2} \right), \\ \frac{\partial g(i, j, \sigma)}{\partial j} &= \frac{1}{2\pi\sigma^2} e^{\left[\frac{-(i^2+j^2)}{2\sigma^2} \right]} \left(-\frac{j}{\sigma^2} \right), \\ \frac{\partial^2 g(i, j, \sigma)}{\partial i^2} &= \frac{1}{2\pi\sigma^2} e^{\left[\frac{-(i^2+j^2)}{2\sigma^2} \right]} \left[-\frac{1}{\sigma^2} + \frac{i^2}{\sigma^4} \right], \\ \frac{\partial^2 g(i, j, \sigma)}{\partial j^2} &= \frac{1}{2\pi\sigma^2} e^{\left[\frac{-(i^2+j^2)}{2\sigma^2} \right]} \left[-\frac{1}{\sigma^2} + \frac{j^2}{\sigma^4} \right].\end{aligned}$$

Finalmente, el Laplaciano de una Gaussiana queda definida por

$$LoG(i, j, \sigma) = g(i, j, \sigma) \left[-\frac{2}{\sigma^2} + \frac{(i^2 + j^2)}{\sigma^4} \right]. \quad (3.5)$$

En las Figuras 3.2(a) y 3.2(b) se muestra una gráfica formada por el Laplaciano de una Gaussiana en dos dimensiones, vista desde arriba y de forma lateral, mientras que en las Figuras 3.2(c) y 3.2(d) se muestra la magnitud de su transformada discreta de Fourier. Se puede ver cómo el Laplaciano de una Gaussiana reduce a cero los componentes de bajas frecuencias en el centro del espectro de frecuencia, así como los componentes de altas frecuencias en los extremos del espectro, dejando pasar solo una banda de frecuencias entre las altas y bajas, por lo que se tiene un filtro pasa banda.

El valor de σ en (3.5) determina el tamaño de la banda de frecuencias que el filtro deja pasar. Aumentar el tamaño de σ disminuye el tamaño de la banda y disminuir σ produce el efecto opuesto. Usando una σ demasiado pequeña proporciona un Laplaciano

de una Gaussiana con la forma de un impulso unitario, por lo que en su transformada de Fourier es un valor constante que deja pasar todo el espectro de frecuencias. Un valor de σ demasiado grande en el dominio espacial da la forma de una constante, que en dominio de frecuencia se da cómo un impulso unitario no permitirá el paso de frecuencias.

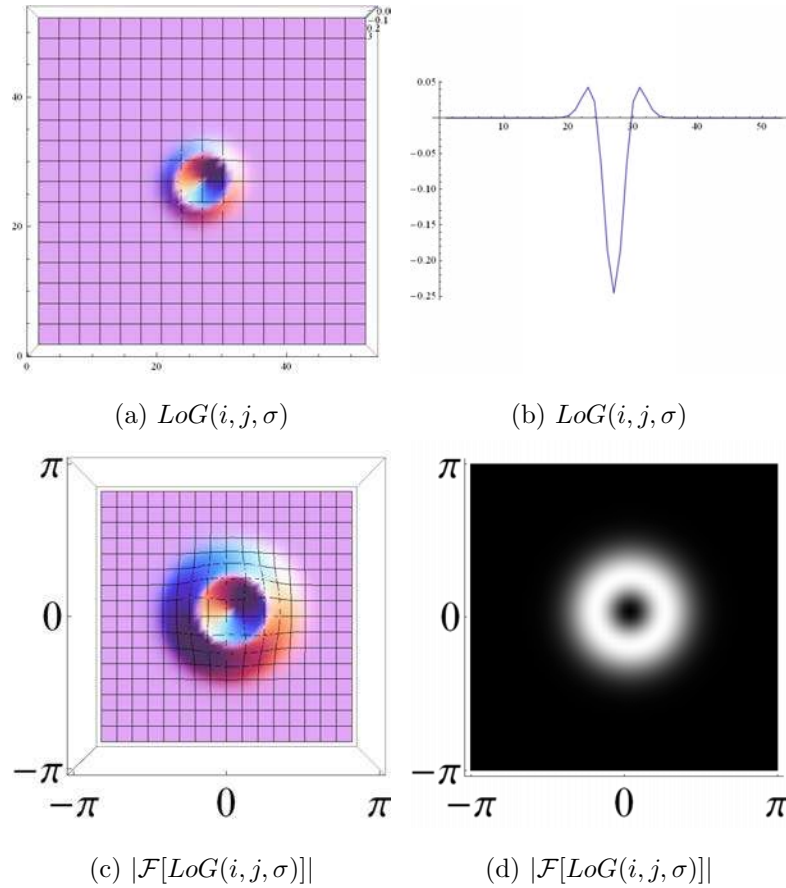


Figura 3.2: Laplaciano de una Gaussiana con $\sigma = 1.0$, y la magnitud de su transformada de Fourier.

3.2.2. Diferencia de Gaussianas

Una diferencia de Gaussianas se utiliza para aproximar el Laplaciano de una Gaussiana. Esta relación se puede demostrar por medio de la ecuación de difusión de calor (3.6) y la definición de un espacio escala Gaussiano. La siguiente demostración puede ser consultada en [Lowe, 2004].

$$\frac{\partial f}{\partial t} = \xi \nabla^2 f \quad (3.6)$$

Para una función $f(i, j, t)$, de dos variables espaciales (i, j) , una variable de tiempo t y una constante de difusión de calor ξ , la cual depende del material en el cual se disipa la temperatura en el sistema, la ecuación de difusión de calor es:

$$\frac{\partial f}{\partial t} = \xi \left(\frac{\partial^2 f}{\partial i^2} + \frac{\partial^2 f}{\partial j^2} \right),$$

donde los términos en el paréntesis corresponden al Laplaciano de f tomando solo las variables espaciales (i, j) . Originalmente la ecuación de difusión de calor es usada como un modelo para describir la distribución de calor en una región por el paso del tiempo, y de forma general, es usada para determinar cómo cambia una función f a través del tiempo t .

El espacio escala Gaussiano para una imagen I , es una familia de señales $L(i, j; \sigma)$ definida por la convolución de I con la Gaussiana $g(i, j, \sigma)$. En este caso σ determina el nivel del espacio escala. De acuerdo con esta definición un espacio escala permite representar una imagen con diferentes niveles de detalles de una forma similar a cómo se hace en la pirámide laplaciana [Burt y Adelson, 1983]. Para el caso del espacio escala Gaussiano, el nivel $\sigma = 0$ corresponde a la imagen original, mientras que niveles superiores o con valores de $\sigma > 0$ corresponden a una convolución de la imagen original con una Gaussiana de tamaño σ , estos niveles superiores representan a la imagen original I en versiones de bajo detalle. El espacio-escala Gaussiano tiene un comportamiento que puede ser descrito por la ecuación de difusión, por lo que reemplazando la función f con la Gaussiana $g(i, j, \sigma)$ y parametrizando con $t = \sigma$ en (3.6) se tiene:

$$\frac{\partial g}{\partial \sigma} = \sigma \nabla^2 g.$$

Esta ecuación de difusión puede ser aproximada por medio de diferencias finitas como se muestra:

$$\sigma \nabla^2 g = \frac{\partial g}{\partial \sigma} \approx \frac{g(i, j, \tau\sigma) - g(i, j, \sigma)}{\tau\sigma - \sigma},$$

al reacomodar los términos se tiene:

$$\begin{aligned} (\tau\sigma - \sigma)\sigma \nabla^2 g &\approx g(i, j, \tau\sigma) - g(i, j, \sigma), \\ (\tau - 1)\sigma^2 \nabla^2 g &\approx g(i, j, \tau\sigma) - g(i, j, \sigma), \\ (\tau - 1)\sigma^2 LoG(i, j, \sigma) &\approx DoG(i, j, \sigma, \tau). \end{aligned}$$

Donde,

$$DoG(i, j, \sigma, \tau) = g(i, j, \tau\sigma) - g(i, j, \sigma). \quad (3.7)$$

El factor de escalamiento $(\tau - 1)$, determina la relación entre las sigmas en la diferencia de Gaussianas así como el nivel de aproximación al Laplaciano de una Gaussiana, el error de aproximación disminuye a 0 conforme τ se acerca a 1, por lo que, de acuerdo a Marr y Hildreth [Marr y Hildreth, 1980], el Laplaciano de una Gaussiana es el límite de la diferencia de Gaussianas. La aproximación del Laplaciano de una Gaussiana usando la diferencia de Gaussianas con una constante $\tau = 1.1$ se muestra en las Figuras 3.3(a) y 3.3(b).

Observando la magnitud de la transformada discreta de Fourier en las Figuras 3.3(c) y 3.3(d) se concluye, de la misma forma que con el Laplaciano de una Gaussiana, que la diferencia de Gaussianas usada como un filtro de frecuencia produce un pasa bandas de frecuencia. Por último, en la Figura 3.4 se presentan el Laplaciano de una Gaussiana y la Diferencia de Gaussianas superpuestas una sobre la otra para diferentes valores de τ , donde se aprecia la similitud entre el Laplaciano de una Gaussiana y su aproximación usando diferencia de Gaussianas.

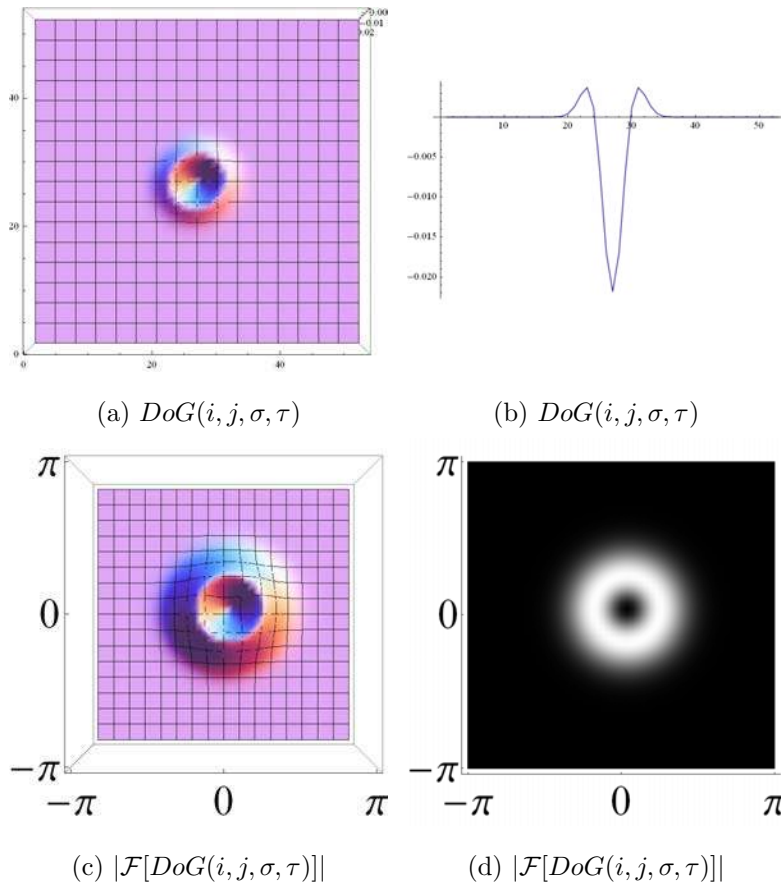


Figura 3.3: Diferencia de Gaussiana con $\sigma = 1.0$, $\tau = 1.1$ y la magnitud de su transformada de Fourier

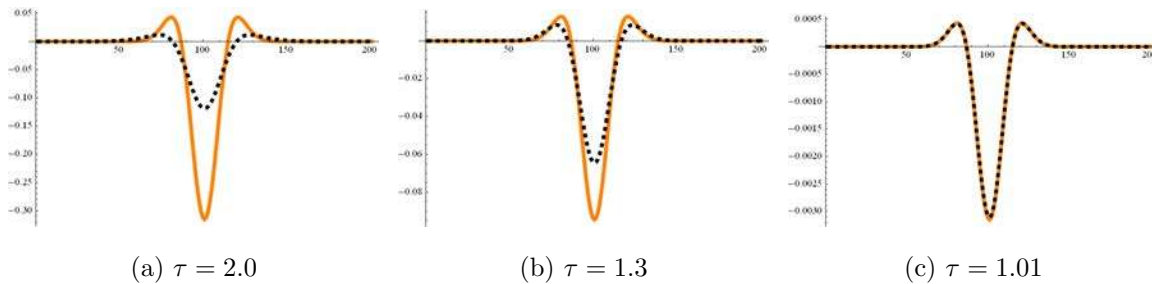


Figura 3.4: Comparación de las gráficas del Laplaciano de una Gaussiana, y diferentes niveles de aproximación a este, usando diferencia de Gaussianas. La línea punteada corresponde a la diferencia de Gaussianas, en todas las gráficas se usa un valor de $\sigma = 1.0$

3.2.3. Wavelets Haar

Los wavelets Haar son usados como filtros de frecuencia en la transformada discreta wavelet. Con los kernels de filtros pasa bajas $l(n) = \frac{1}{\sqrt{2}}[1, 1]$ y pasa altas $h(n) = \frac{1}{\sqrt{2}}[-1, 1]$ y usando la convolución de estos kernels en diferentes secuencias producen los cuatro wavelets distintos (3.8) a (3.11), cada uno de ellos se usa como un filtro de frecuencia.

$$LL = l * l = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad (3.8)$$

$$LH = l * h = \frac{1}{2} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \quad (3.9)$$

$$HL = h * l = \frac{1}{2} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \quad (3.10)$$

$$HH = h * h = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (3.11)$$

La Figura 3.5 presenta las magnitudes de la transformada discreta de Fourier de los diferentes wavelets Haar. Ya que el wavelet HH presenta el filtrado que más frecuencias altas deja pasar, se compara con las otras funciones para determinar si tiene un mejor funcionamiento al distinguir regiones enfocadas.

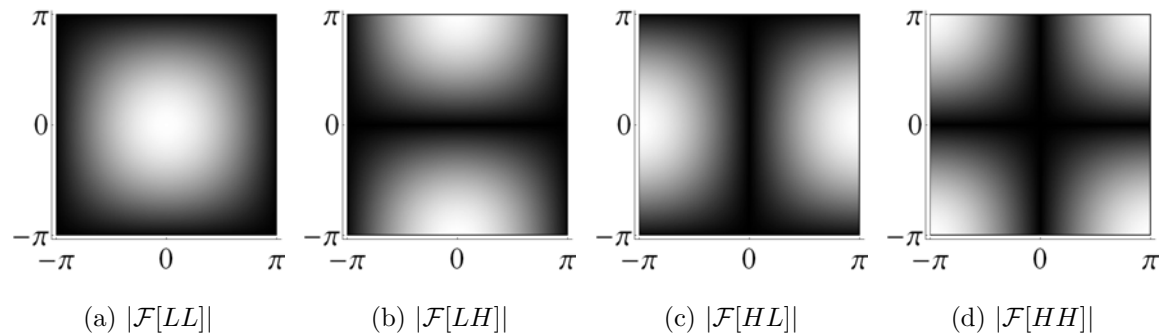


Figura 3.5: Magnitud de las transformadas de Fourier de los wavelets Haar.

3.3. Filtro pasa altas propuesto

El filtro propuesto cómo pasa altas en este trabajo, se basa en las observaciones hechas al filtrado de frecuencias presentado en las secciones anteriores usando el Laplaciano de una Gaussiana, la diferencia de Gaussianas y los wavelets Haar. En todos estos casos las respuestas de frecuencia obtenidas presentan una pérdida de altas frecuencias, observable de manera sencilla en los extremos del espectro en las magnitudes de las transformadas de Fourier de cada uno, por lo que el método de filtrado propuesto usa un filtro pasa altas que no presenta esta pérdida de frecuencias altas.

El filtro propuesto entonces, consiste en la convolución de una imagen I con una Gaussiana $g(i, j, \sigma)$, la cual funciona cómo un pasa bajas, y después sustraer esta convolución de la imagen original I , el resultado de esta sustracción es una imagen donde las frecuencias bajas son extraídas de la imagen original dejando sólo las frecuencias altas:

$$I - (I * g(i, j, \sigma)) = I * (\delta(i, j) - g(i, j, \sigma)) = I * G(i, j, \sigma), \quad (3.12)$$

$$G(i, j, \sigma) = \delta(i, j) - g(i, j, \sigma), \quad (3.13)$$

donde $\delta(i, j)$ es la delta de Dirac, con la cual se puede realizar la factorización de I en (3.12) dado que la convolución de una imagen I con $\delta(i, j)$ produce la misma imagen I . Haciendo un análisis de la respuesta de frecuencia con la magnitud de la transformada discreta de Fourier del filtro pasa altas $G(i, j, \sigma)$ propuesto, se observa que éste no presenta la misma pérdida de las altas frecuencias en los extremos del espectro de frecuencias. Esta diferencia es lo que le da una mayor precisión al usarlo para distinguir regiones de alto enfoque en imágenes.

En la Figura 3.6(a) se muestra la forma que tiene el filtro propuesto, y en las Figuras 3.6(b) y 3.6(c) se presenta la magnitud de su transformada de Fourier, en donde se observa cómo se deja pasar los valores en las áreas de alta frecuencia en los extremos del espectro, mientras que se hace una reducción a 0 en las áreas de coeficientes de bajas frecuencias.

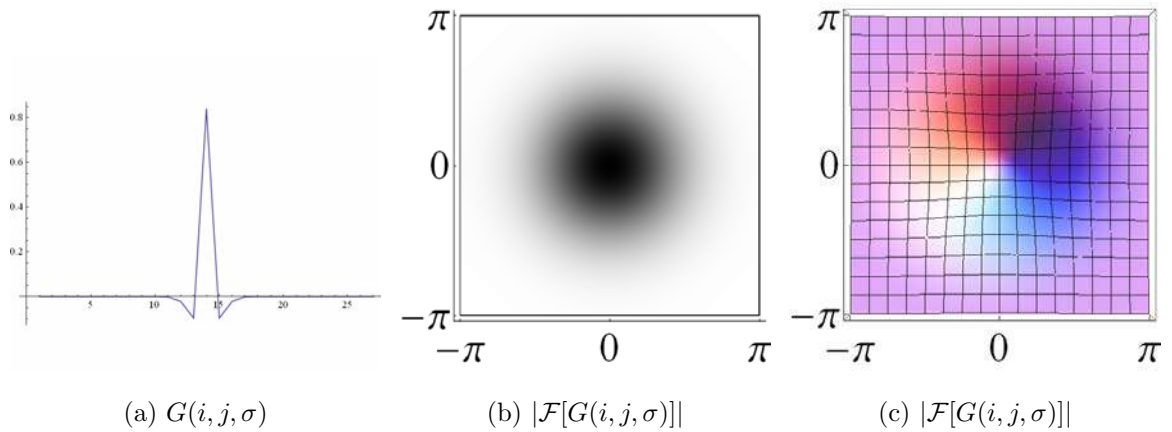


Figura 3.6: Filtro pasa altas propuesto, con $\sigma = 1.0$ y la magnitud de su TDF.

3.4. Medición de la calidad de la fusión

Se desea calificar el resultado de la fusión, de modo que se pueda determinar qué algoritmo de fusión tiene mejor desempeño. Una medida de calidad debería determinar la habilidad del proceso de fusión para transferir, de la manera más precisa posible toda la información visual en las imágenes originales a la imagen resultante. En el presente caso de investigación, la información que se desea transmitir se trata de las áreas enfocadas en las imágenes originales.

Existe una multitud de mediciones de calidad distintas, las cuales caen dentro de dos categorías dependiendo del tipo de imágenes con la que comparan la fusión. La primera se trata de la medición sobre imágenes sintéticas, o con imagen de referencia. Estas imágenes multifoco son generadas de forma artificial, por ejemplo realizando una deformación en la imagen, y después realizando el mismo procedimiento en una copia de la imagen original pero de forma complementaria. En este trabajo se realiza un emborronado de la imagen original o referencia, para producir dos imágenes sintéticas. En este tipo de imágenes se puede usar la imagen original sin deformar como comparación o la imagen verdadera que se desea obtener. En este trabajo se generan imágenes de esta manera, donde la deformación usada es un emborronado de regiones de la imagen. Después de fusionar las imágenes, el resultado es comparado con la imagen referencia. Una de las razones por la cual se opta por

este tipo de imágenes de prueba es la falta de un método de medición de calidad de fusión de imágenes validado en forma general.

Idealmente, el método de fusión debería presentar una imagen resultante idéntica a la imagen de referencia. Cuando no se tiene la imagen de referencia, cómo es en el caso de imágenes de ejemplos reales obtenidos con cambios de enfoque en la cámara, se trata con mediciones que cuantifican el nivel de información en la imagen fusionada. A continuación se presenta una revisión de las diferentes medidas de calidad, usadas en diferentes trabajos de fusión de imágenes, para comparar las fusiones resultantes.

3.5. Medición de la calidad en imágenes sintéticas

3.5.1. Índice universal de calidad

Propuesto por Bovik y Wang, en [Bovik y Zhou, 2002], pretende modelar distorsiones en imágenes cómo una combinación de tres factores: pérdida de correlación, distorsiones de iluminación y distorsiones de contraste. La medición de similitud, teniendo dos imágenes, I_a e I_b , y está dada por:

$$Q(i, j) = \frac{\sigma_{I_a I_b}}{\sigma_{I_a} \sigma_{I_b}} \cdot \frac{2\bar{I}_a \bar{I}_b}{\bar{I}_a^2 + \bar{I}_b^2} \cdot \frac{2\sigma_{I_a} \sigma_{I_b}}{\sigma_{I_a}^2 + \sigma_{I_b}^2}, \quad (3.14)$$

donde

$$\begin{aligned} \bar{I}_a &= \frac{1}{NM} \sum_{i=0}^N \sum_{j=0}^M I_a(i, j), \\ \bar{I}_b &= \frac{1}{NM} \sum_{i=0}^N \sum_{j=0}^M I_b(i, j), \\ \sigma_{I_a}^2 &= \frac{1}{NM} \sum_{i=0}^N \sum_{j=0}^M (I_a(i, j) - \bar{I}_a)^2, \\ \sigma_{I_b}^2 &= \frac{1}{NM} \sum_{i=0}^N \sum_{j=0}^M (I_b(i, j) - \bar{I}_b)^2, \end{aligned}$$

$$\sigma_{I_a I_b} = \frac{1}{NM} \sum_{i=0}^N \sum_{j=0}^M (I_a(i, j) - \bar{I}_a)(I_b(i, j) - \bar{I}_b).$$

El valor de similitud con esta medición está entre -1 y 1 obteniendo 1 únicamente cuando las imágenes x y y son idénticas. El primer factor de (3.14) mide el valor de correlación lineal entre la imagen original y la fusionada. El segundo factor mide que tan cercanas son las medias del valor de intensidad, por lo que compara el nivel del brillo. Por último, el tercer factor mide la diferencia entre los contrastes de las imágenes. Esta medida es usada como referencia en trabajos como [Shnayderman et al., 2003, Blasch y Boulevard, 2008, Hossny y Nahavandi, 2009] que buscan una mejor medida de calidad de imágenes.

3.5.2. Suma de diferencias cuadradas

Esta medida realiza una comparación sobre la imagen original I , que se toma como los valores reales y la fusión resultante I_c de un par de imágenes sintéticas, calculando la raíz cuadrada del promedio del cuadrado los errores entre los valores de intensidad, esto definido en (3.15). Se desea que estos errores sean mínimos indicando un mayor parecido entre las imágenes.

$$S(I, I_c) = \sqrt{\frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} [I(i, j) - I_c(i, j)]^2}{NM}}. \quad (3.15)$$

Dado que esta medida define la similitud entre las imágenes basándose en las diferencias en las intensidades de cada píxel de la imagen fusionada y la referencia, es más sensible a los errores de iluminación que a distorciones estructurales como los cambios de enfoque. Esto quiere decir, que esta medida puede medir un error más grande con dos imágenes de diferentes intensidades de píxel, que la que asigna a dos imágenes con diferencias de nitidez.

3.6. Medición de la calidad en imágenes sin referencia

3.6.1. Entropía

Esta medida determina el nivel de información contenida en una señal. La entropía H de una imagen se define cómo

$$H(I_c) = - \sum_{l=0}^{D-1} P_r(l) \log P_r(l), \quad (3.16)$$

donde D es el número total de intensidades de píxel diferentes en la imagen fusionada I_c , y $P_r(l)$ es la probabilidad de ocurrencia del valor de intensidad l para una imagen I . Un valor alto de entropía corresponde a un mejor enfoque en la imagen fusionada. El valor máximo de entropía es obtenido cuando cada nivel de intensidad de grises en la imagen tiene el mismo número de ocurrencias. La entropía, de acuerdo con Leung et al., [Leung et al., 2001], no puede distinguir entre información y ruido, por lo que la medición puede estar siendo influida por el nivel del ruido más que por el nivel de información útil en la imagen.

3.6.2. Desviación estandar

La desviación estandar, definida en (3.17), mide el nivel nitidez en la imagen por medio de la media de los valores de intensidad de los píxeles en la imagen. Esta medida se usa con la idea de que las diferencias entre los píxeles y la media, corresponden a información contenida en la imagen. Las regiones con alto nivel de información se pueden considerar cómo regiones de alto enfoque. Entre más grande sea la diferencia entre cada píxel y el valor de la media, mayor será el valor de la desviación estandar y, por lo tanto, se tiene una mayor dispersión en la distribución de los valores de intensidad de los píxeles de la imagen, consecuentemente es posible que la imagen contenga más información y mejor enfoque. Sin embargo cómo en el caso de la entropía definida en (3.16), la medida no puede diferenciar entre distorsiones o ruido en la imagen.

$$\sigma_{I_c}^2 = \frac{1}{NM} \sum_{i=0}^N \sum_{j=0}^M (I_c(i, j) - \bar{I}_c)^2. \quad (3.17)$$

Esta medida, en comparación con la suma de diferencias cuadradas, se puede usar sobre una imagen fusionada sin necesidad de tener una imagen original para compararla.

3.6.3. Promedio de gradiente

Mide el nivel de contraste entre las variaciones de las texturas así como en los detalles en la imagen, entre más grande sea el valor de promedio de gradiente, mayor enfoque tendrá la imagen. Un ejemplo del uso de esta medición para determinar la calidad de fusión de imágenes se puede encontrar en el trabajo de Xuelong en [Hu et al., 2010].

$$\bar{G}(I_c) = \frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \sqrt{\frac{\left(\frac{\partial}{\partial i} I_c(i, j)\right)^2 + \left(\frac{\partial}{\partial j} I_c(i, j)\right)^2}{2}} \quad (3.18)$$

3.7. Desempeño de mediciones de calidad en la fusión

Finalmente, en la Figura 3.7 se muestra el comportamiento de las mediciones de calidad presentadas en esta sección, sobre una imagen I cómo la de Lena en la Figura 3.1(a). La imagen I es emborronada usando una Gaussiana $g(i, j, \sigma)$ con diferentes valores de σ y se presentan en las gráficas las mediciones obtenidas. En las Figuras 3.7(a), 3.7(b) y 3.7(c) conforme aumenta el nivel de σ de la Gaussiana, disminuye el nivel de nitidez de la imagen. En consecuencia, las medidas presentan una disminución en su valor y por lo tanto una disminución en la calidad de la imagen. Por otro lado, en 3.7(d), el error o diferencia entre la imagen original y las versiones convolucionadas aumentan conforme σ crece.

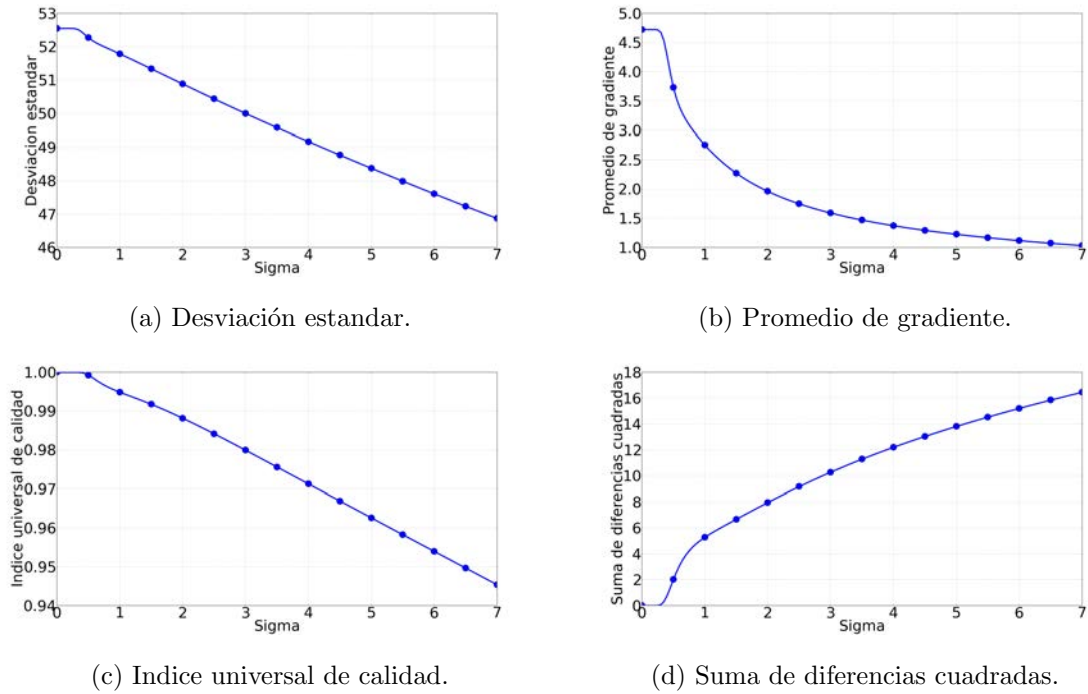


Figura 3.7: Aquí se muestra cómo el emborronado de una imagen disminuye la medición de calidad de la misma.

3.8. Generación de imágenes de prueba

Para realizar una comparación del funcionamiento de la fusión de imágenes del Algoritmo 1 para cada uno de los filtros pasa altas revisados en la Sección 3.2, se usan imágenes sintéticas generadas a partir de una imagen de referencia I . La obtención de las imágenes usadas en las pruebas realizadas se describe de la siguiente forma. Primero, se simula un desenfoque de cámara haciendo uso de una Gaussiana $g(i, j, \sigma)$, convolucionandola con una imagen I , produciendo una imagen emborronada o de baja nitidez I_g . Luego se define un mapa de decisión $M_o(i, j) \in \{0, 1\}$ con (3.19) que determina en cada coordenada o píxel, la regla a seguir para la formación de las imágenes sintéticas.

$$M_o(i, j) = \begin{cases} 1 & \text{Escoger píxel de la imagen enfocada } I \\ 0 & \text{Escoger píxel de la imagen emborronada } I_g \end{cases} \quad (3.19)$$

Considere una imagen I , cómo la mostrada en la Figura 3.8(a), y la imagen I_g mostrada en la Figura 3.8(b), que es la convolución de la imagen I con una Gaussiana $g(i, j, \sigma_g)$ con desviación estandar $\sigma_g = 6.0$. Para este ejemplo se toma el valor de σ_g únicamente para facilitar la observación del emborronado de la imagen. Después, usando el Algoritmo 2 con el mapa M_o que es definido a priori (por ejemplo con el mostrado en la Figura 3.8(c)), se obtienen las dos imágenes sintéticas. La imagen I_a de la Figura 3.8(d) se obtiene con el mapa M_o usando (3.20) y la imagen I_b de la Figura 3.8(e) se obtiene con el complemento de M_o con (3.21). El mapa M_o tiene la forma o contorno de uno de los lobos, con el propósito de simular un enfoque distinto o de separar a cada lobo en dos planos de enfoque.

$$I_a(i, j) = I(i, j)M_o(i, j) + I_g(i, j)(1 - M_o(i, j)) \quad (3.20)$$

$$I_b(i, j) = I(i, j)(1 - M_o(i, j)) + I_g(i, j)M_o(i, j) \quad (3.21)$$

Algoritmo 2: Obtención de imágenes sintéticas.

Entrada: $I, M_o \in \{0, 1\}, \sigma$

Salida: I_a, I_b

/* Convolucionar I con un filtro Gaussiano $g(i, j, \sigma)$. */

1 $I_g \leftarrow I * g(i, j, \sigma)$

/* Con I, I_g y el mapa M_o de (3.19), obtener */

/* las imágenes sintéticas. */

2 **para** $i \leftarrow 0$ **hasta** $N - 1$ **hacer:**

3 **para** $j \leftarrow 0$ **hasta** $M - 1$ **hacer:**

4 Calcular:

5 $I_a(i, j)$ con (3.20)

6 $I_b(i, j)$ con (3.21)

7 **fin de para**

8 **fin de para**

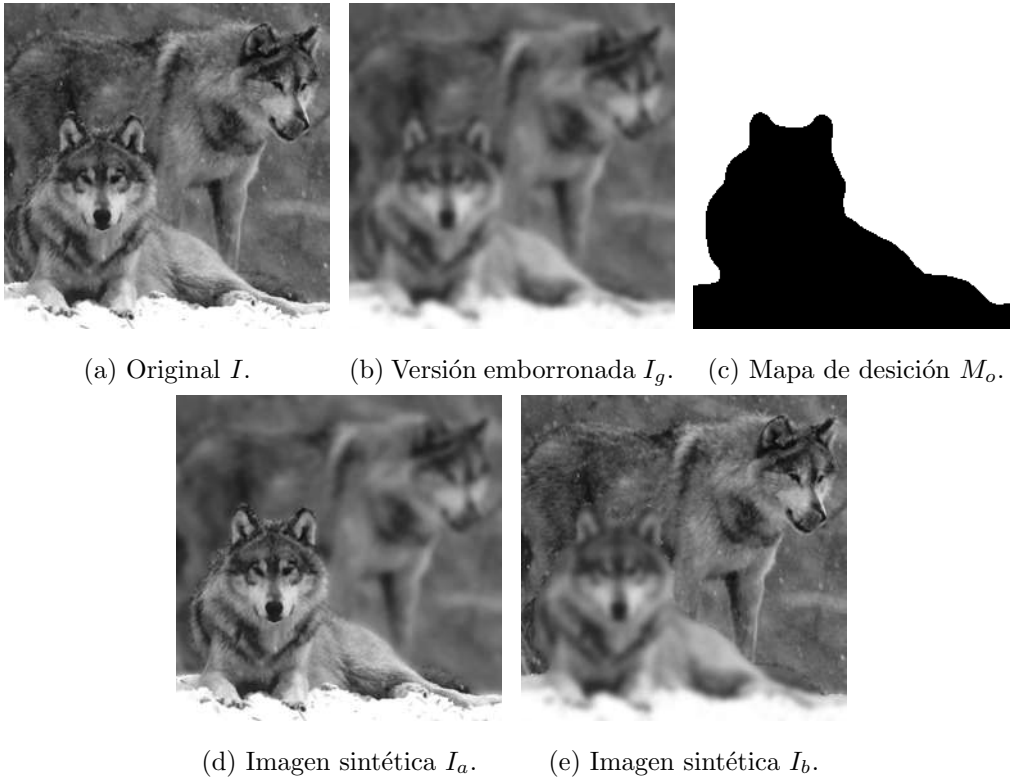


Figura 3.8: Generación de imágenes multi-foco sintéticas, a) imagen inicial I , b) I_g resultante de la convolución de la imagen I con el filtro Gaussiano g , c) mapa de decisión M_o , con el que se eligen píxeles de las imágenes I e I_g , d) y e) imágenes sintéticas resultantes.

3.8.1. Mapa de decisión resultante

Con el método de fusión del Algoritmo 1 visto en la Sección 3.1, se puede además, producir un mapa de decisión M_r a partir de las dos imágenes de respuesta de frecuencias \hat{I}_a e \hat{I}_b obtenidas con la convolución de las imágenes a fusionar con un filtro de frecuencia que determina las regiones de alto enfoque. El mapa de decisión M_r es creado indicando sobre un arreglo binario, en cual de las dos respuestas se tiene la región o píxel con nivel de actividad más alto cómo se hace con (3.22) y el Algoritmo 3 de la sección 3.8.1. Esta es la misma comparación de niveles de actividad de (3.2) con la que se obtiene la imagen fusionada I_c en el Algoritmo 1 con el cual se fusionan imágenes, pero devolviendo además

un mapa binario.

$$M_r(i, j) = \begin{cases} 1 & \text{si } |\hat{I}_a(i, j)| > |\hat{I}_b(i, j)| \\ 0 & \text{en caso contrario} \end{cases} \quad (3.22)$$

Algoritmo 3: Obtención de la imagen fusionada I_c y el mapa M_r

Entrada: I_a, I_b

Salida: $I_c, M_r \in \{0, 1\}$

/* Convolucionar imágenes a fusionar con */

/* un filtro pasa altas h . */

1 $\hat{I}_a \leftarrow I_a * h$

2 $\hat{I}_b \leftarrow I_b * h$

/* Se obtienen la imagen compuesta I_c con (3.2) y */

/* el mapa de decisión resultante M_r con (3.22). */

3 **para** $i \leftarrow 0$ **hasta** $N - 1$ **hacer:**

4 **para** $j \leftarrow 0$ **hasta** $M - 1$ **hacer:**

5 **si** $|\hat{I}_a(i, j)| > |\hat{I}_b(i, j)|$ **entonces**

6 $I_c(i, j) \leftarrow I_a(i, j)$

7 $M_r(i, j) \leftarrow 1$

8 **si no**

9 $I_c(i, j) \leftarrow I_b(i, j)$

10 $M_r(i, j) \leftarrow 0$

11 **fin de si**

12 **fin de para**

13 **fin de para**

La construcción de este mapa de decisión, permite introducir una medida de comparación entre el mapa decisión original M_o con el que se crearon las imágenes sintéticas y el mapa de decisión resultante M_r construido por el método de fusión. Dados los mapas M_o y M_r , éstos se comparan a nivel píxel basándose en criterios de segmentación. El propósito de esto es que la medición no involucre errores de medición cómo el que se tiene en la suma de diferencias cuadradas por los niveles de intensidad, o interferencia por ruido en la imagen

cómo en el caso de la entropía. Además de que al tratar con el mapa de decisión se tiene una comparación de los resultados enfocada en coherencia espacial. Para lograr esto se utiliza el coeficiente de Tanimoto, del que se dan más detalles en la subsección siguiente.

3.8.2. Coeficiente Tanimoto

El coeficiente de Tanimoto, también conocido cómo la medida de similitud de Jaccard para conjuntos binarios presentado por Jaccard, en [Jaccard, 1901], sirve cómo una medida de similitud entre dos conjuntos. El coeficiente de Tanimoto se define cómo el valor de cardinalidad de la intersección de dos conjuntos, dividido entre la cardinalidad de su unión. En este caso, cada conjunto consiste en un arreglo binario con valores 0 o 1, la medición del coeficiente de Tanimoto obtenida dados el mapa original M_o y el mapa resultante M_r es entonces:

$$T(M_o, M_r) = \frac{|M_o \cap M_r|}{|M_o \cup M_r|}$$

La cual se traduce a la siguiente fórmula al tratar con arreglos binarios.

$$T(M_o, M_r) = \frac{\sum_{i=0}^N \sum_{j=0}^M M_o(i, j) \wedge M_r(i, j)}{\sum_{i=0}^N \sum_{j=0}^M M_o(i, j) \vee M_r(i, j)} \quad (3.23)$$

Recordando que el arreglo M_o corresponde al mapa de decisión que genera las imágenes sintéticas de prueba y M_r corresponde al mapa de decisión en relación a los niveles de actividad dadas las imágenes de respuestas de frecuencia. El valor de similitud obtenido con el coeficiente Tanimoto se encuentra entre 0 y 1. Se obtiene un 0 cuando los dos conjuntos son disjuntos o no tienen miembros en común y 1 cuando se tienen mapas idénticos:

$$T(M_o, M_r) = \frac{|M_o \cap M_r|}{|M_o \cup M_r|} = \frac{|M_o|}{|M_o|} = 1$$

$$T(M_o, M_r) = \frac{|M_o \cap M_r|}{|M_o \cup M_r|} = \frac{|\emptyset|}{|M_o|} = 0$$

3.9. Pruebas experimentales con imágenes sintéticas

Esta prueba representa una simplificación del problema real, ya que se tienen elementos inexistentes en el caso de las imágenes reales, específicamente, el mapa de decisión M_o con el que se generan las imágenes sintéticas y la imagen de referencia. Estos elementos representan aspectos de la solución que se desea encontrar al resolver el problema de fusión, es decir, en el caso del mapa de decisión, nos permite saber de antemano, la correspondencia de los píxeles en las imágenes a fusionar en la imagen resultante fusionada. Y en el caso de la imagen de referencia, se presenta la misma solución que se busca.

Las pruebas fueron efectuadas sobre un par de imágenes sintéticas a fusionar cómo las mostradas en las Figuras 3.9(d) y 3.9(e). Estas son generadas de acuerdo al Algoritmo 2. La imagen original usada es la de Lena con un tamaño de 512 x 512 píxeles tal cómo se muestra en la Figura 3.9(a), Esta se convoluciona con una Gaussiana $g(i, j, \sigma_g)$ para la creación de la imagen I_g de la Figura 3.9(b), el mapa M_o se muestra en la Figura 3.9(c).

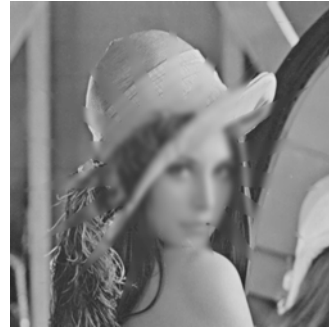
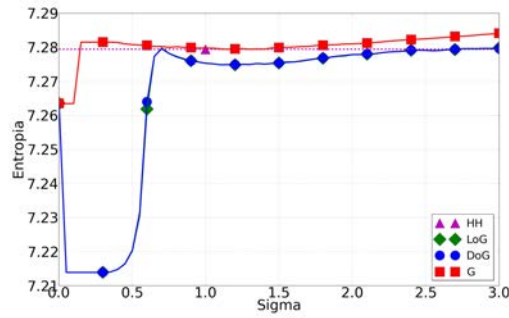
(a) Original I .(b) Versión emborronada I_g .(c) Mapa de decisión M_o .(d) Imagen sintética I_a .(e) Imagen sintética I_b .

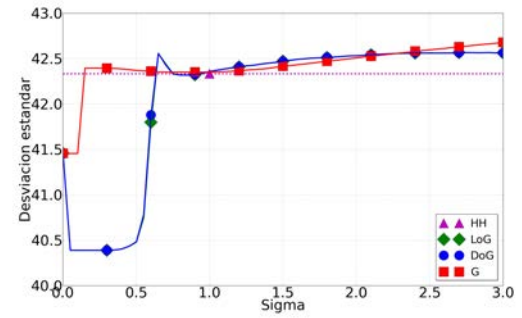
Figura 3.9: Conjunto de imágenes sintéticas usadas en las pruebas.

La forma del mapa M_o fue escogida específicamente por la complejidad de su contorno, el cual tiene relación con el algoritmo de segmentación que será usado para mejorar la fusión y que es explicado con más detalle en el Capítulo 4. Siguiendo el procedimiento descrito por el Algoritmo 3, se hace la convolución de estas imágenes sintéticas con cada una de las funciones usadas como filtros mostradas en la Sección 3.2: $LoG(i, j, \sigma)$ (3.5), $DoG(i, j, \sigma, k)$ (3.7), el wavelet Haar HH (3.11) y el filtro pasa altas $G(i, j, \sigma)$ (3.13) que se propone en la Sección 3.3. Cada una produce un resultado de fusión y un mapa de decisión distinto. Después, se determina la calidad de las imágenes fusionadas usando las medidas de calidad de imagen de la sección 3.4: desviación estandar (3.17), entropía (3.16), promedio de gradiente (3.18), suma de diferencias cuadradas (3.15) y el índice de universal de calidad (3.14). Por último, usando el coeficiente de Tanimoto, definido en (3.23), como medida de similitud, se compara el mapa M_o con el mapa M_r producido por cada fusión, si los algoritmos de fusión son correctos M_o y M_r deben ser idénticos y el valor del coeficiente de Tanimoto tenderá a 1.

En las pruebas que se realizaron, las imágenes sintéticas son generadas con el Algoritmo 2 usando un valor de σ constante, mientras que para filtrar las frecuencias en estas imágenes se varia el valor de σ con el que se entonan los filtros usados, con excepción del wavelet Haar el cual no cambia de tamaño. El valor de σ con el que se entonan los filtros, determina el ancho de banda de frecuencias que dejan pasar, y por lo tanto, afecta la calidad de la fusión. Variando el valor de σ para cada uno de estos filtros se obtienen diferentes resultados de fusión haciendo un filtrado diferentes anchos de banda con lo que se determina cual tiene mejor funcionamiento. Las fusiones obtenidas de las imágenes de Lena de La Figura 3.9 son comparadas en la Figura 3.10. Para esta comparación se grafica el ancho de banda filtrado por cada función, contra la calidad de la fusión que el mismo filtrado produce. Las mediciones de entropía, en la Figura 3.10(a), y desviación estandar, Figura 3.10(b), presentan resultados similares. Ambos indican mejor calidad conforme el valor de la medición incrementa. Los mapas de decisión correspondientes a las fusiones resultantes que estas dos mediciones presentan como las mejores, se muestran en las Figuras 3.13(a) a 3.13(d) y 3.13(e) a 3.13(h), respectivamente.



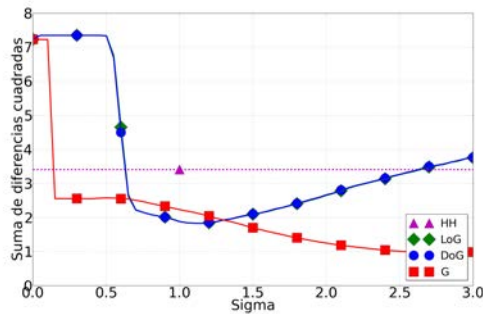
(a) Entropía.



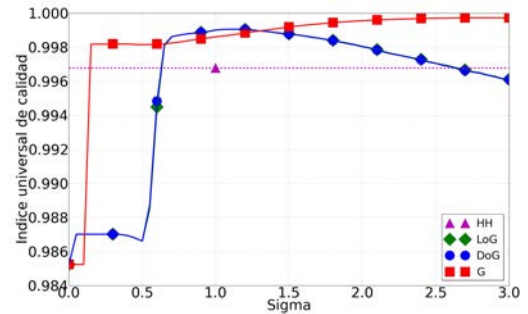
(b) Desviación estándar.

Figura 3.10: Resultados de la comparación de fusiones, se comparan los filtros variando el valor del ancho de banda filtrado contra la calidad de la fusión.

En las mediciones de suma de diferencias cuadradas 3.11(a) la cual mide el nivel de similitud de la imagen fusionada con la imagen de referencia I , por lo que el mejor valor de calidad medido corresponde a 0, cuando ambas imágenes son idénticas, y el índice universal de calidad 3.11(b) que también mide la similitud de la imagen fusionada con la imagen de referencia, en este caso el valor de calidad más alto corresponde a 1 cuando las imágenes son idénticas. Los mapas de decisión de estos resultados se muestran en las Figuras 3.13(i) a 3.13(l) para la suma de diferencias cuadradas y 3.14(a) a 3.14(d) para el índice universal de calidad.



(a) Suma de diferencias cuadradas.



(b) Índice de calidad universal.

Figura 3.11: Calidad contra frecuencia filtrada.

El promedio de gradiente indica una mejor calidad de la fusión conforme éste

aumenta, en 3.12(a) se tiene los resultados de la prueba para el promedio de gradiente. Los mapas de decisión correspondientes a las mejores fusiones según esta medición son mostrados en las Figuras 3.14(e) a 3.14(h). La Figura 3.12(b) muestra la comparación de los mapas de decisión M_o y M_r , con el coeficiente de Tanimoto, de acuerdo al valor de este coeficiente, las mejores fusiones son las que presentan un coeficiente con valor cercano a 1, recordando que un coeficiente de Tanimoto de 1 indica dos mapas idénticos. La diferencia de Gaussianas (3.7) con un valor de $\tau = 1.01$ proporcionó resultados muy similares a los del Laplaciano de una Gaussiana (3.5) en todas las pruebas realizadas. El wavelet Haar HH (3.11) presenta un resultado constante, ya que el ancho de banda de frecuencias que filtra no cambia.

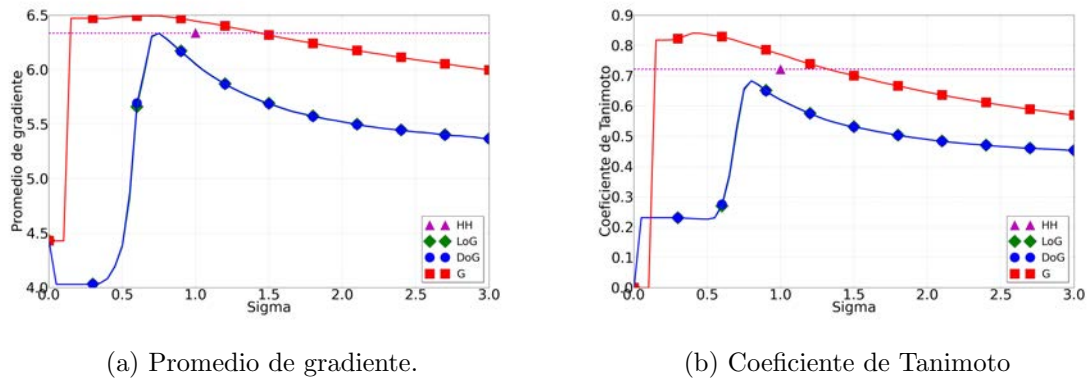


Figura 3.12: Calidad contra frecuencia filtrada.

Al comparar las fusiones usando las medidas de calidad y después de observar los mapas resultantes que corresponden a los mejores resultados según estas medidas, observe que las medidas de calidad con excepción del promedio de gradiente y el coeficiente de Tanimoto, presentan una incongruencia en los resultados. Por ejemplo, en los resultados de las Figuras 3.13 y 3.14 las medidas presentan cómo mejores resultados, a mapas que tienen muchos errores, siendo que otra medida asigna cómo mejores resultados a mapas que tienen menos errores, e.g., en las Figuras 3.14(i) a 3.14(l) correspondientes al coeficiente de Tanimoto.

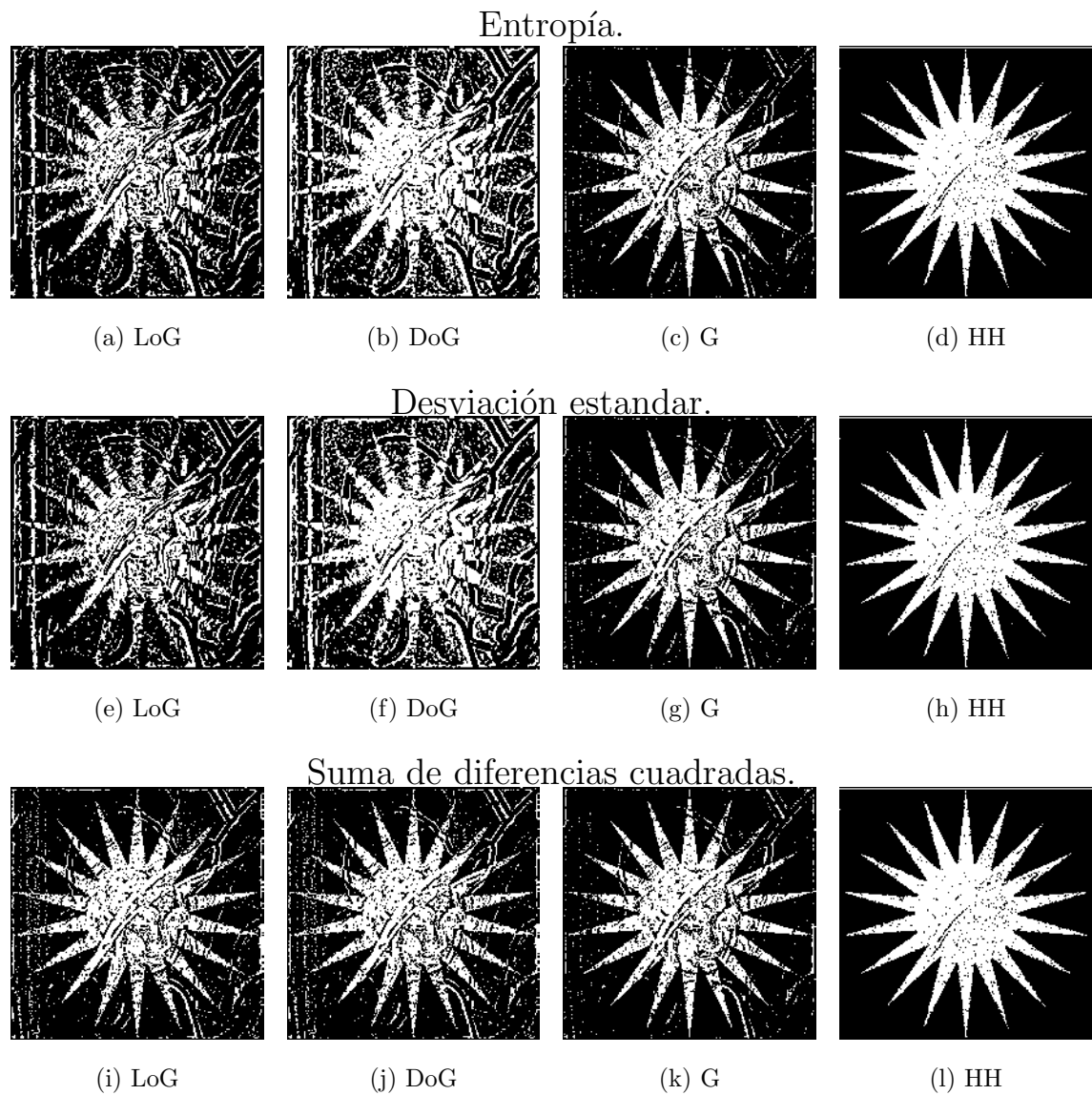


Figura 3.13: Mapas de decisión binarios M_r , de los mejores resultados de fusión según las medidas de calidad: (a) a (d) Entropía, (e) a (h) Desviación estandar, (i) a (l) Suma de diferencias cuadradas.

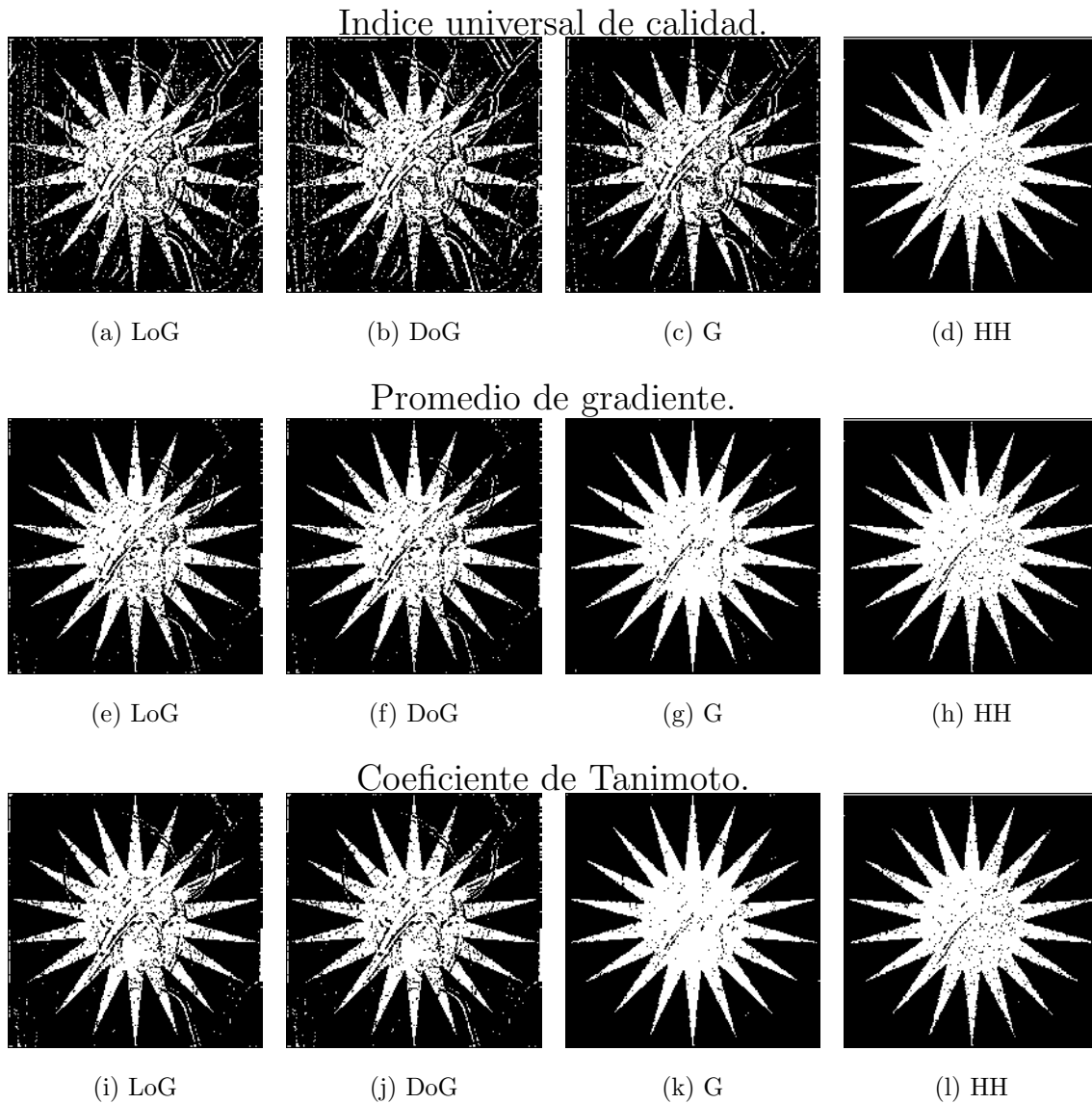


Figura 3.14: Mapas de decisión binarios M_r , de los mejores resultados de fusión según las medidas de calidad: (a) a (d) Índice universal de calidad, (e) a (h) Promedio de gradiente, (i) a (l) Coeficiente de Tanimoto.

Con esto se concluye que el coeficiente de Tanimoto presenta una medida de calidad de la fusión congruente con el resultado obtenido, y tiene el criterio de calidad de fusión más acertado en el caso de imágenes sintéticas, ya que puede compararse de forma directa el mapa obtenido con la fusión y el mapa M_o que se usa para generar las imágenes sintéticas.

Por el contrario, las otras medidas presentan un resultado inadecuado, ya que como se mencionó en el Capítulo 3 algunas medidas, son propensas a ser afectadas por ruido o cambios en la intensidad de píxel. Este efecto no está presente en el coeficiente de Tanimoto ya que hace una comparación de forma directa entre el mapa de decisión que genera las imágenes sintéticas y el obtenido en la fusión. Tomando esto en cuenta, se puede decir, finalmente, que el filtro propuesto presenta el mejor filtrado de frecuencias, y se usa para seleccionar las regiones de alto enfoque en la imagen en el método de fusión que se presenta en esta tesis.

En la Figura 3.15 se tiene el mapa M_r , obtenido con una fusión de las imágenes sintéticas de la Figura 3.9 usando el filtro pasa altas propuesto, definido en (3.13). Este mapa es el que presenta el mejor resultado del coeficiente de Tanimoto en la gráfica de la Figura 3.12(b). Este mapa presenta un resultado que puede ser mejorado, los píxeles en el mapa de decisión que se encuentran mal asignados pueden ser corregidos para obtener un mapa de decisión M_r mejorado, y que sea más parecido al mapa M_o de la Figura 3.9(c) con lo que se obtendría una fusión de mejor calidad. Se usará un algoritmo de segmentación basado en campos aleatorios de Markov para conseguir esta mejora, el cual se presenta en el siguiente capítulo.

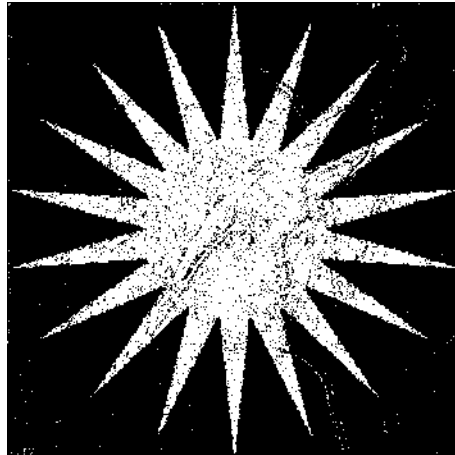


Figura 3.15: Mapa de decisión obtenido de la fusión de las imágenes sintéticas de la figura 3.9, usando el filtro pasa altas (3.13) propuesto, con un valor de $\sigma = 0.4$

3.9.1. Conclusiones del capítulo

Al terminar los experimentos se llegó a la conclusión de que el filtro pasa altas propuesto produce los mejores resultados. Y que las medidas de calidad, aunque son usadas con frecuencia en el estado del arte y presentadas en los experimentos de este capítulo, son inadecuados para medir la calidad de la fusión de imágenes multifoco. Estas no presentan una congruencia con los mapas de decisión resultantes, los cuales presentan un indicador visual preciso de cuales son las mejores funciones para el filtrado de frecuencias. Por esta razón se usará el coeficiente de Tanimoto como la medida real del funcionamiento del método de fusión para el caso de imágenes sintéticas, dejando la comparación de imágenes reales o sin referencia, en forma visual, presentando acercamientos a las áreas de interés en los resultados.

Capítulo 4

Segmentación del mapa de decisión.

El propósito de la segmentación de una imagen es dividirla en K regiones de píxeles que compartan una similitud de sus propiedades. La segmentación de una imagen I produce las subregiones R_1, R_2, \dots, R_K de forma que $\bigcup_{k=1}^K R_k = I$, y $R_k \cap R_l = \emptyset, \forall k \neq l$; es decir, que la unión de todas las regiones forma la imagen original I y que ninguna de estas regiones se traslapa. En este capítulo se presenta la teoría detrás de la formación de un campo aleatorio de Markov, así como un algoritmo de segmentación basado en estos campos, se pretende usar la segmentación como un método para mejorar el resultado de la fusión de imágenes multifoco que se obtiene por medio del Algoritmo 3 presentado en la Sección 3.2.

4.1. Campos aleatorios de Markov

Un campo aleatorio de Markov es un modelo gráfico de probabilidad que consiste de un conjunto de variables aleatorias cuyas relaciones de dependencia son descritas por medio de un grafo no dirigido. Cada nodo en el grafo corresponde a una de las variables aleatorias. Se les llama campos aleatorios de Markov por la propiedad Markoviana que presentan estas variables en el modelo, es decir, el valor que puede tomar cada variable o nodo en el grafo tiene una probabilidad que depende solo del valor que tienen los vecinos

de ese nodo y es independiente de los nodos fuera de este vecindario.

El esquema de un campo aleatorio de Markov usado en el procesamiento de imágenes presentado en este capítulo se puede encontrar en [Rangarajan y Chellappa, 1995]. Se inicia con la definición de un conjunto de variables aleatorias B que está asociado a una imagen I , de forma que cada elemento o variable de este conjunto es representado por $B(s)$ donde s se refiere a un sitio o píxel con coordenadas (i, j) en una imagen I . Sea b uno de los estados en los que se puede encontrar el modelo, es decir: $B = \{b^{(1)}, b^{(2)}, \dots, b^{(K)}\}$ y para cada uno de estos estados, $b(s)$ es una de las variables aleatorias o uno de los valores que puede tomar el campo aleatorio en el sitio s . Por último se define un sistema de vecinos $\mathcal{N}_b(s)$ que consiste en los píxeles que rodean al sitio o elemento s . Los campos Markovianos usan este sistema de vecinos como una forma de representar el conocimiento a priori del sistema en forma de distribuciones de probabilidad locales.

El sistema de vecinos usado puede tomar diferentes formas. En la Figura 4.1(a) se presentan ejemplos de diferentes órdenes de sistemas de vecinos. Conforme se incrementa el orden del sistema se toman en cuenta un mayor número de vecinos y se tiene más conocimiento del sistema, pero se aumenta la complejidad de cómputo del modelo. En este trabajo se utiliza un sistema de vecinos de primer orden o de cuatro vecinos para los campos de Markov. Si el elemento o píxel s en la imagen es representado por la posición (i, j) , tomando un sistema de vecinos de primer orden, se tiene que el conjunto de vecinos de s es: $\mathcal{N}_b(s) = \mathcal{N}_b(i, j) = \{B(i, j + 1), B(i, j - 1), B(i + 1, j), B(i - 1, j)\}$. En la Figura 4.1(b) se muestra la disposición del vecindario de cuatro píxeles.

Por último, un campo aleatorio de Markov requiere la definición de un clique. Tomando a $(B, \mathcal{N}_b(s))$ como el grafo correspondiente al campo de Markov, donde B es el conjunto de nodos o sitios s y \mathcal{N}_b determina el conjunto de vecinos o conexiones del grafo. Para un sitio s , un clique \mathcal{C} en b , es un subconjunto de B tal que \mathcal{C}_1 consiste de un solo sitio $\{s\}$ para un clique de primer orden, \mathcal{C}_2 consiste de un par de sitios vecinos $\{s, t\}$ para un clique de segundo orden, \mathcal{C}_3 consiste de tres sitios vecinos $\{s, t, u\}$ para un clique de tercer orden, y así sucesivamente. La colección de todos los cliques en B es $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3 \dots$

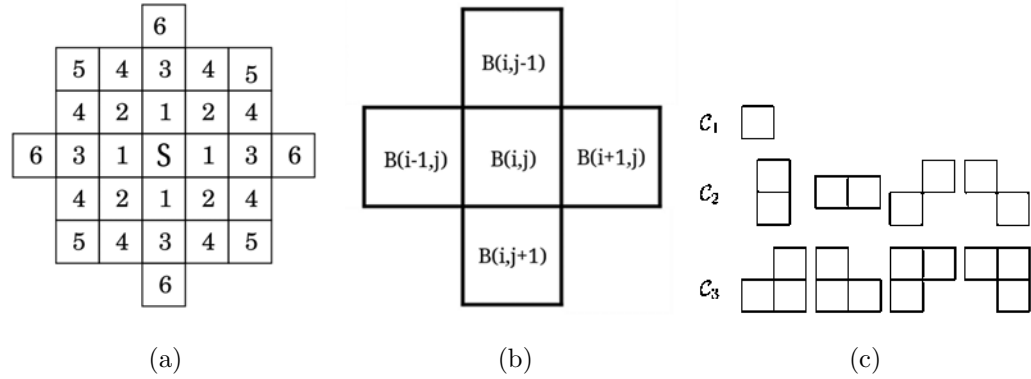


Figura 4.1: Componentes de un campo aleatorio de Markov a) Sistemas de vecinos de diferentes órdenes, b) Un sistema de vecinos de primer orden, c) Formas de los cliques de primer, segundo y tercer orden.

En la Figura 4.1(c) se presentan ejemplos de la disposición de los píxeles en los cliques de primer, segundo y tercer orden en el grafo. Para el sistema de vecinos \mathcal{N}_b de primer orden, se usan cliques de segundo orden, es decir de dos vecinos.

Considere la probabilidad condicional $Pr(B(s) = b(s)|B(t) = b(t), \forall t \in \mathcal{N}_b(s))$ que indica la probabilidad de que un sitio s en B tome un valor, que depende únicamente del valor que tienen sus vecinos. Para calcular esta probabilidad se requiere saber cual es la probabilidad conjunta de que el modelo B se encuentre en un estado b o bien, $Pr(B = b)$, ya que esta es la que contiene la representación completa de la imagen I . Para obtener la probabilidad conjunta se hace uso de la definición de una distribución de Gibbs (4.1),

$$Pr(B = b) = \frac{1}{Z_b} e^{[-E(b)]}, \quad (4.1)$$

donde $E(b)$ es una función de energía que representa la suma de todos los cliques en el estado b del campo aleatorio. Z_b en (4.1) consiste en la suma de la energía de todas las configuraciones o estados posibles de B . Puesto que los cliques representan la información que se conoce del sistema, el costo que estos aportan a la función de energía es definido dependiendo del tipo de problema que se pretende resolver con el campo aleatorio. Esta distribución corresponde originalmente a un campo aleatorio de Gibbs, y se puede utilizar

cómo la probabilidad conjunta del modelo del campo aleatorio de Markov si el campo cumple con las condiciones descritas por Hammersley y Clifford en [Hammersley y Clifford, 1971]. Específicamente, un campo de Gibbs está caracterizado por,

1. Positividad.

Establece que la probabilidad de tener cualquier estado b en el modelo B es positiva.

$$Pr(B = b) > 0$$

2. Localidad.

Esta propiedad indica la independencia entre las probabilidades del conjunto de vecinos de un sitio y el resto de sitios en el modelo, esta condición es la misma propiedad Markoviana del sistema.

$$Pr(B(s) = b(s)|B(t) = b(t), t \neq s, \forall t \in I) = Pr(B(s) = b(s)|B(t) = b(t), \forall t \in \mathcal{N}_b(s))$$

3. Homegenidad.

Establece que la probabilidad de que un sitio s en B , tome cierto valor, depende solo del valor de sus vecinos, y que esto se cumple para todos los sitios s en el campo aleatorio.

$$Pr(B(s) = b(s)|B(t) = b(t), \forall t \in \mathcal{N}_b(s)) \text{ es la misma para todos los sitios } s.$$

4.1.1. Segmentación controlada por entropía

El método de segmentación con modelos de Markov cómo los mostrados en la Sección 4.1 y que se utiliza para mejorar el método de fusión propuesto, fue desarrollado por Rivera et al., en [Rivera et al., 2007]. Si se representan los valores de intensidad de píxel de la imagen I cómo la suma de funciones determinísticas y ruido se tiene el modelo siguiente,

$$I(s) = \sum_{k=1}^K b_k(s)\Phi(s; \theta_k) + n_g(s), \quad (4.2)$$

donde $n_g(s) : s \in I$ es una variable que se comporta cómo ruido con una distribución normal y $\Phi(s; \theta_k)$ es una función determinística que depende del parámetro θ_k , correspondiente al valor de la media de b . La apariencia de I , en la Ecuación (4.2), depende de la naturaleza de las funciones Φ . Si cada función Φ es suave, es decir, está compuesta por regiones continuas de valores, I tendrá una apariencia suave excepto en las fronteras de las regiones adyacentes en la segmentación, R_k, R_l .

En el caso de segmentación de imágenes, se toma al estado b del campo aleatorio de Markov B , definido en la sección anterior, cómo una de las soluciones posibles que puede tener la segmentación de la imagen. En el problema de fusión de imágenes la segmentación propuesta consta de dos regiones correspondientes a cada valor en el mapa de decisión. Cada sitio $b(s)$ será interpretado cómo una distribución que indica la probabilidad de que un píxel en un sitio s pertenezca a una región R_k en particular, de modo que $b_k = 1$ *sii* $s \in R_k$ y $\sum_{k=1}^K b_k(s) = 1$. Para la solución del problema se trata a b cómo un campo Markoviano usando la distribución Gibbs, definida en, (4.1) cómo el conocimiento a priori del sistema. En la segmentación de la imagen se desea que b sea espacialmente suave en todas sus regiones, para evitar la proliferación de áreas pequeñas asignadas a regiones particulares debido al ruido en la imagen o, en el caso abordado, a errores en el mapa de decisión resultante. Además, se requiere que cada vector $b(s)$ cumpla con las restricciones siguientes:

$$\begin{aligned} a) \quad b_k(s) &\geq 0, \quad k = 1, \dots, K, \quad s \in I \\ b) \quad \sum_{k=1}^K b_k(s) &= 1, \quad s \in I \end{aligned} \tag{4.3}$$

Estas restricciones, junto al requerimiento de una b con regiones suaves, se presentan en la función de energía,

$$E(b) = \sum_{\langle s,t \rangle} \lambda \|b(s) - b(t)\|^2 - \mu \sum_{s \in I} \sum_{k=1}^K b_k^2(s), \tag{4.4}$$

El primer término de (4.4) se encarga de la suavidad de las regiones en b , la sumatoria involucra todos los cliques de segundo orden en b . La constante λ determina el peso que se le da a este término. Conforme se aumenta λ se tiene una segmentación con regiones

más suaves. El segundo término es conocido como el índice Gini [Gini, 1955], el cual tiene un comportamiento similar a la entropía; es máximo cuando la distribución, en este caso el vector de cada sitio $b(s)$, es uniforme, y es mínimo a medida que la distribución se aproxima a una función delta, es decir que uno de los valores en el vector $b(s)$ es 1 y el resto es 0. La constante μ en el segundo término, determina el control de la entropía sobre la segmentación. Usando la regla de Bayes sobre la distribución Gibbs, definida en (4.1), y la función de energía, definida en (4.4), así como el modelo propuesto de una imagen I definido en (4.2), como una suma de funciones determinísticas, se obtiene la función de probabilidad posterior:

$$P(b|I, \bar{\theta}) = \frac{1}{Z_b} e^{-E(b, \bar{\theta})}. \quad (4.5)$$

Con la función de energía

$$E(b, \bar{\theta}) \approx \sum_{s \in I} \sum_{k=1}^K b_k^2(s) [-\log v_k(s) - \mu] + \lambda \sum_{\langle s, t \rangle} \|b(s) - b(t)\|^2, \quad (4.6)$$

donde $v_k(s)$ es una medida de verosimilitud que indica la probabilidad de que un píxel o sitio s en la imagen I , tome un cierto valor de intensidad, esto acuerdo a la media θ y asumiendo que el sistema sigue un modelo Gaussiano,

$$v_k(s) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\left[\frac{(I(s) - \theta_k)^2}{2\sigma_k^2}\right]}.$$

Para obtener el estimador de b se desea minimizar la función de energía definida en (4.6), sujeta a las restricciones de igualdad, en (4.3). Esta minimización se realiza usando multiplicadores de Lagrange. El Lagrangiano correspondiente a la función de energía, en (4.6), es

$$\mathcal{L}(b, \bar{\theta}, \gamma) = E(b, \bar{\theta}) - 2 \sum_{s \in I} \gamma_s \left[\sum_{k=1}^K b_k(s) - 1 \right], \quad (4.7)$$

donde γ_s es el multiplicador correspondiente al Lagrangiano. Después, se iguala el gradiente de (4.7) a cero y se resuelve para b , con lo que se tiene,

$$\begin{aligned}
\frac{\partial \mathcal{L}(b, \bar{\theta}, \gamma)}{\partial b_k(s)} &= b_k(s)[- \log v_k(s) - \mu] + \lambda \sum_{t \in \mathcal{N}(s)} [b_k(s) - b_k(t)] - \gamma_s = 0 \\
&= b_k(s)[- \log v_k(s) - \mu + \lambda |\mathcal{N}(s)|] - \lambda \sum_{t \in \mathcal{N}(s)} b_k(t) - \gamma_s = 0 \quad (4.8)
\end{aligned}$$

$$\frac{\partial \mathcal{L}(b, \bar{\theta}, \gamma)}{\partial \gamma_s} = \sum_{k=1}^K b_k(s) - 1 = 0.$$

Haciendo

$$\begin{aligned}
\alpha_k(s) &= \lambda \sum_{t \in \mathcal{N}(s)} b_k(t) \\
\beta_k(s) &= - \log v_k(s) - \mu + \lambda |\mathcal{N}(s)|
\end{aligned}$$

se puede reescribir (4.8) cómo:

$$b_k(s)\beta_k(s) - \alpha_k(s) - \gamma_s = 0$$

$$b_k(s) = \frac{\alpha_k(s) + \gamma_s}{\beta_k(s)}. \quad (4.9)$$

Para encontrar el multiplicador γ_s en (4.9) haciendo:

$$\sum_{i=1}^K b_i(s) = \sum_{i=1}^K \frac{\alpha_i(s) + \gamma_s}{\beta_i(s)}$$

y por la restricción de igualdad (4.3)

$$1 = \sum_{i=1}^K \left(\frac{\alpha_i(s)}{\beta_i(s)} \right) + \gamma_s \sum_{i=1}^K \left(\frac{1}{\beta_i(s)} \right)$$

$$\gamma_s = \frac{1 - \sum_{i=1}^K \left(\frac{\alpha_i(s)}{\beta_i(s)} \right)}{\sum_{i=1}^K \left(\frac{1}{\beta_i(s)} \right)} \quad (4.10)$$

Finalmente sustituyendo γ_s , de (4.10), en (4.9) se obtiene la ecuación de actualización,

$$b_k(s) = \frac{\alpha_k(s)}{\beta_k(s)} + \frac{1 - \sum_{i=1}^K \frac{\alpha_i(s)}{\beta_i(s)}}{\sum_{i=1}^K \frac{\beta_k(s)}{\beta_i(s)}} \quad (4.11)$$

De esta manera se actualiza el campo aleatorio b , recordando que el minimizar la función de energía provoca una disminución en la entropía del sistema, con lo que el campo b converge en la segmentación de la imagen, la segmentación esta controlada por el valor de λ en (4.6) el cual determina la potencia de la segmentación al modificar el costo proporcionado por cada clique en el campo aleatorio. Por último si no se conocen los parámetros θ del sistema, se puede usar la misma función de energía, en (4.6), para obtener un estimador de θ realizando el mismo procedimiento de minimización de la función de energía ahora con respecto a θ , obteniendo la siguiente función de actualización de parámetros θ ,

$$\begin{aligned} \frac{\partial E(b, \bar{\theta})}{\partial \theta_k} &= \sum_{s \in I} b_k^2(s) [2(\theta_k - I(s))] = 0 \\ \sum_{s \in I} b_k^2(s) \theta_k &= \sum_{s \in I} b_k^2(s) I(s) \\ \theta_k &= \frac{\sum_{s \in I} b_k^2(s) I(s)}{\sum_{s \in I} b_k^2(s)} \end{aligned} \quad (4.12)$$

Este procedimiento de segmentación se puede agregar al Algoritmo 3 para obtener el Algoritmo 4, que describe el método de fusión de imágenes propuesto.

Algoritmo 4: Algoritmo de fusión de imágenes con segmentación.

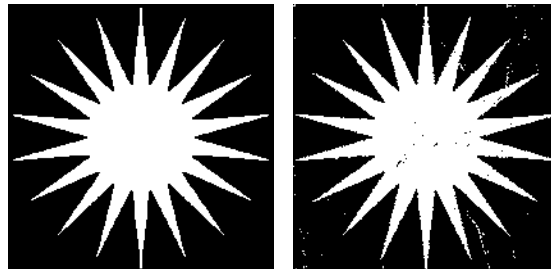
Entrada: $I_a, I_b, \sigma, \lambda$ **Salida:** I_c /* Convolucionar imágenes a fusionar con el filtro pasa altas
propuesto, G (3.13) de la sección 3.3. */1 $\hat{I}_a \leftarrow I_a * G$ 2 $\hat{I}_b \leftarrow I_b * G$ /* Obtener mapa de decisión M_r , comparando el nivel de
actividad de las respuestas de frecuencia (3.2) con el
algoritmo 3 *//* Tomar el mapa de decisión M_r como el campo aleatorio de
Markov inicial b . */3 $b \leftarrow M_r$ 4 **mientras** $iter < iter_{max}$ **hacer:**5 | Actualiza b (4.11);6 | Calcula θ (4.12);7 | $iter \leftarrow iter + 1$;8 **fin de mientras**/* Para $K = 2$, es decir dos imágenes a fusionar y por lo tanto
dos regiones en la segmentación, $b_0(i, j)$ es la probabilidad de
que el píxel $I_c(i, j)$ se tome de I_a , y $b_1(i, j) = (1 - b_0(i, j))$ */9 **para** $i \leftarrow 0$ **hasta** $N - 1$ **hacer:**10 | **para** $j \leftarrow 0$ **hasta** $M - 1$ **hacer:**11 | | **si** $b_0(i, j) > b_1(i, j)$ **entonces**12 | | | $I_c(i, j) \leftarrow I_a(i, j)$ 13 | | **si no**14 | | | $I_c(i, j) \leftarrow I_b(i, j)$ 15 | | **fin de si**16 | **fin de para**17 **fin de para**

La segmentación para mejorar el resultado de fusión, funciona de forma que al obtener el mapa de decisión M_r , éste es segmentado, proporcionando un mapa con regiones de píxeles uniformes o suaves. Se desea que las regiones en la segmentación sean suaves, por que la proliferación de píxeles o regiones de píxeles solitarios representan errores en la selección de regiones de alto o bajo enfoque, e.g., un píxel asignado como una región de alto enfoque no debería estar rodeado de regiones de píxeles de bajo enfoque. Al finalizar la segmentación, se espera obtener un mapa sin errores como regiones de píxeles solitarios. La segmentación propuesta se basa en la probabilidad que tiene cada píxel, de ser un píxel correspondiente a una región de alto o bajo enfoque, esto determinado por sus vecinos.

4.2. Resultados

4.2.1. Segmentación del mapa de decisión obtenido con la fusión de imágenes sintéticas

En esta sección se presentan los resultados de la mejora del mapa de decisión M_r , lograda con la segmentación de imágenes. En la Figura 4.2(a) se muestra el mapa de decisión original M_o usado en las pruebas del Capítulo 3, con el cual se generan las imágenes sintéticas. En la Figura 4.2(b) se presenta el mapa de decisión M_r obtenido con el Algoritmo 3. Se desea que la segmentación proporcione una mejora en el mapa de decisión M_r obtenido en la fusión de imágenes, para que este resulte lo mas parecido posible al mapa original M_o .



(a) Mapa de decisión original. (b) Mapa resultante de la fusión.

Figura 4.2: Mapas de decisión obtenidos en la fusión de imágenes usando el Algoritmo 3.

Recordando que el valor de λ en (4.6) controla el nivel de la segmentación, cómo se muestra en las Figuras 4.3(a) a 4.3(d) el usar una λ muy grande es contraproducente para la fusión. Al aumentar demasiado el valor de λ , disminuyen las áreas o píxeles incorrectos en el mapa de decisión en forma de regiones de píxeles solitarios, pero también se presenta una pérdida de precisión en el contorno correspondiente a los objetos enfocados en las imágenes. Esto se observa principalmente en los detalles del mapa cómo las puntas de la estrella, por lo que debe presentarse un balance entre la preservación de estos contornos y la mejora del mapa.

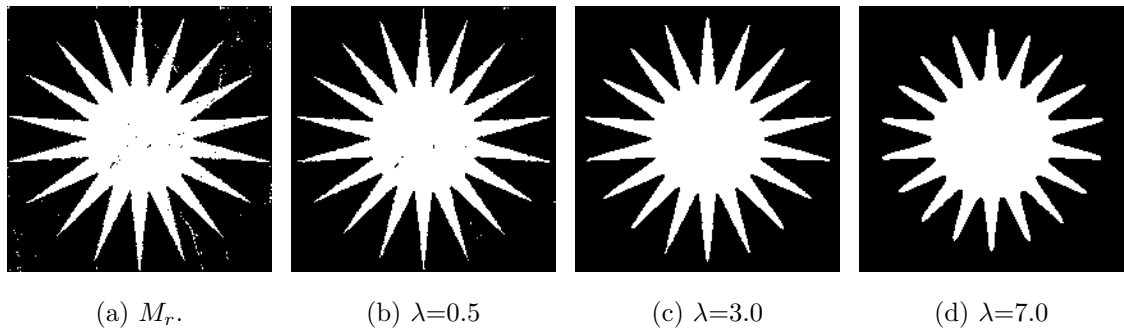


Figura 4.3: Resultados de la segmentación para diferentes valores de λ .

4.2.2. Segmentación del mapa obtenido con la fusión de imágenes reales

Para el caso de imágenes reales el experimento consistió en tomar el el mapa de decisión M_r que presentaba menos errores, seguido de una segmentación del mismo en diferentes valores de λ para la segmentación. Se consiguen los resultados de diferentes λ por que no se tiene un método para identificar el mejor mapa segmentado de forma automática. Por lo que el mejor mapa es asignado por inspección visual. A continuación se presentan los resultados de la segmentación del mapa de decisión de la Figura 4.4, correspondiente a un caso real del problema. En la Figura 4.4(a) se muestra el mapa resultante sin segmentar, obtenido con el Algoritmo 3. En las Figuras restantes 4.4(b) a 4.4(i), se presentan distintos resultados de segmentación de este mapa, correspondientes a diferentes valores de λ . De nuevo se observa que el segmentar el mapa de decisión proporciona una mejora del resultado.

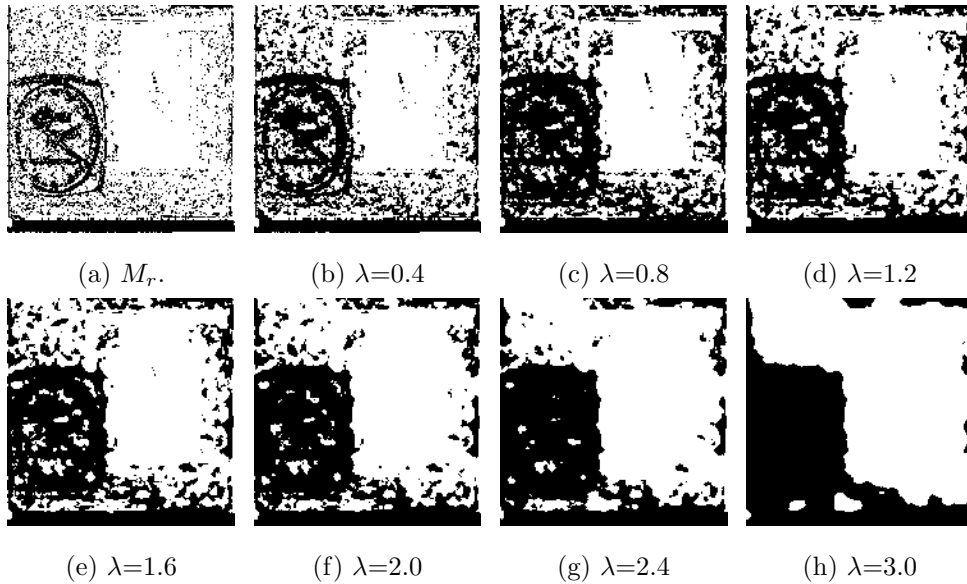


Figura 4.4: Resultados de la segmentación para diferentes valores de λ

Finalmente, observando en las Figuras 4.5(a) y 4.5(b), se tienen la selección de regiones de alto enfoque de las imágenes originales usando el mapa de decisión segmentado, que para este ejemplo se trata del mapa de la Figura 4.4(h), se observa que con el mapa segmentado se logra obtener una separación en regiones de los planos de enfoque de cada reloj.



Figura 4.5: Selección de las regiones de alta frecuencia en las imágenes originales, usando el mapa de decisión segmentado, el mapa usado corresponde a el filtro pasa altas propuesto, con un valor de $\sigma = 0.3$ y la segmentación usando un valor de $\lambda = 3.0$.

4.3. Conclusiones

En las pruebas realizadas se observó una mejora en todos los mapas de decisión para todos los valores de σ con el que se filtraron las imágenes, por lo que se muestra la utilidad de usar la segmentación sobre el mapa de decisión. El valor de λ con el cual se obtiene el mejor resultado de segmentación, para el caso de la imagen de Lena, se consigue con el coeficiente de Tanimoto. Para el caso de las imágenes de los relojes, el mejor resultado se determina de forma subjetiva por observación de los resultados ya que no se cuenta con un mapa original M_o de referencia. Entre los efectos notados al usar la segmentación sobre los mapas de decisión, fue que, en los casos donde se tiene un mapa de decisión con una forma simple, por ejemplo un círculo, se puede elevar aun más el valor de λ sin perder la forma del contorno del mapa de decisión, cómo en el caso del mapa que tiene un contorno con forma de estrella mostrado en la Figura 4.2(a).

Capítulo 5

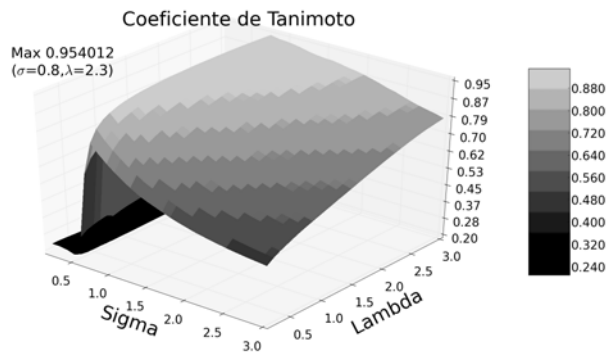
Resultados.

Usando el método de fusión de imágenes del Algoritmo 4 y probándolo con las distintas funciones vistas en el Capítulo 3, se consiguieron los siguientes resultados. En las Figuras 5.1 a 5.4 se comparan las fusiones resultantes de las imágenes sintéticas de Lena en la Figura 3.9, de la misma forma en que se fusionaron en el Capítulo 3 pero agregando el algoritmo de segmentación para mejorar el resultado. La comparación se realiza en base al coeficiente de Tanimoto sobre los mapas de decisión resultantes de cada fusión, en la comparación se varía el ancho de banda filtrado cambiando el valor de σ , de la misma forma que en las pruebas del Capítulo 3 y para esta comparación, también se varía el valor de λ que controla la segmentación del mapa de decisión. En todas las gráficas se presenta el mejor resultado, así como los parámetros correspondientes para su obtención.

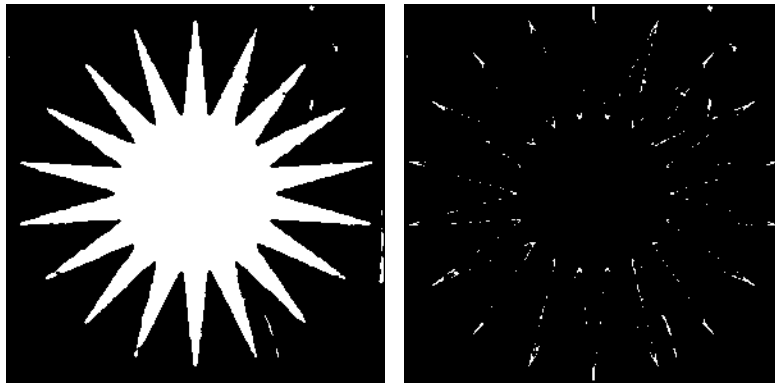
Empezando con el Laplaciano de una Gaussiana, definido en (3.5), se presenta en la gráfica de la Figura 5.1(a) la gráfica correspondiente al coeficiente de Tanimoto, variando el nivel de segmentación λ contra el valor de σ que determina el ancho de banda filtrado. El resultado indica que el coeficiente de Tanimoto más alto se da con los parámetros; $\sigma = 0.8$ y $\lambda = 2.3$. Recordando que este coeficiente indica una similitud entre el mapa de decisión original, M_o , y el obtenido con el algoritmo de fusión 4, M_r , y que esta similitud es más alta conforme el coeficiente se acerca a 1. El mapa de decisión correspondiente a estos parámetros es mostrado en la Figura 5.1(b). Observando el mapa resultante es evidente la mejora lograda con el algoritmo de segmentación, el cual eliminó la mayor parte de los

errores en el mapa de decisión.

La Figura 5.1(c) presenta el error entre el mapa de decisión M_o y el resultante de la fusión M_r segmentado. Este mapa se calcula como la diferencia entre los mapas y se puede ver cómo las puntas en el contorno de la estrella presentan el mayor error en la segmentación mientras que las líneas rectas presentan menos errores.



(a) Coeficiente de Tanimoto para las fusiones con el Laplaciano de una Gaussiana (3.5).



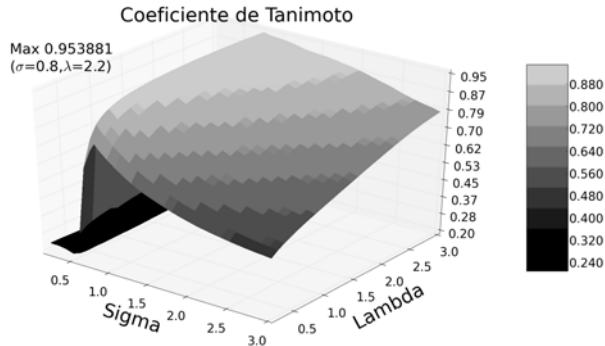
(b) Mapa M_r segmentado correspondiente al mejor coeficiente de Tanimoto.

(c) Error entre los mapas M_o y M_r segmentado con mejor coeficiente de Tanimoto.

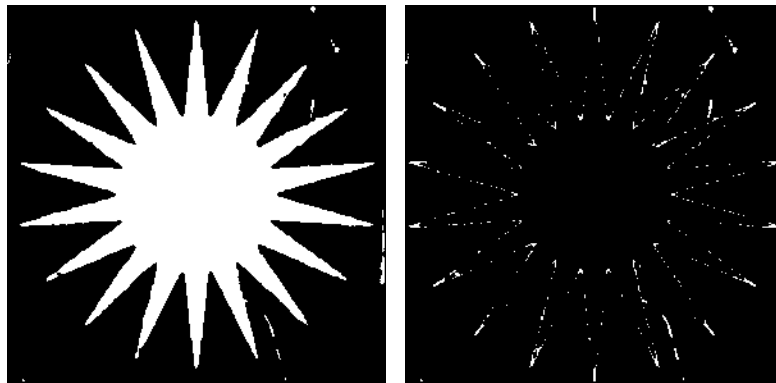
Figura 5.1: Fusión de imágenes con el Laplaciano de una Gaussiana (3.5) y segmentación.

Aunque el Laplaciano de una Gaussiana y la diferencia de Gaussiana presentan resultados similares en las pruebas del capítulo 3 y las presentadas en esta sección, obser-

vando el error en los incisos (c) de las Figuras 5.1 y 5.2, se puede ver cómo el Laplaciano de una Gaussiana funciona mejor que la aproximación de este con la diferencia de Gaussianas, ya que presenta menos áreas blancas o errores.



(a) Coeficiente de Tanimoto para las fusiones con una diferencia de Gaussianas (3.7)

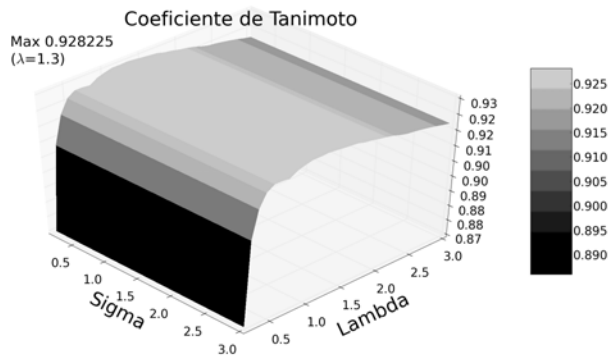


(b) Mapa M_r segmentado correspondiente al mejor coeficiente de Tanimoto. (c) Error entre los mapas M_o y M_r segmentado con mejor coeficiente de Tanimoto.

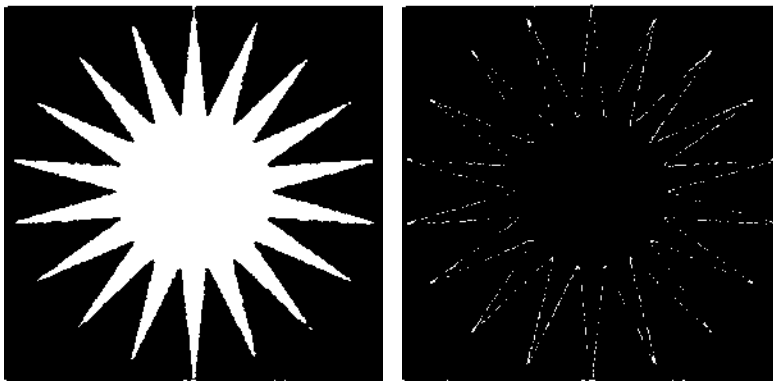
Figura 5.2: Fusión de imágenes con una diferencia de Gaussianas (3.7) y segmentación.

Finalmente comparando el resultado obtenido con el wavelet Haar HH (3.11) Figura 5.3 y el pasa altas propuesto (3.13) Figura 5.4, los cuales presentan los mejores resultados, se puede ver que con el wavelet Haar, se eliminaron los errores en cuanto a regiones de píxeles solitarios se refiere, pero tiene un menor desempeño que el filtro pasa altas propuesto si

se observa el error en los mapas de los incisos (c) de cada figura.

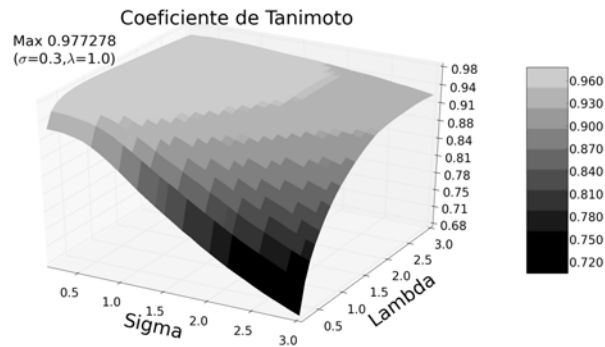


(a) Coeficiente de Tanimoto para las fusiones con el wavelet Haar HH
(3.11)

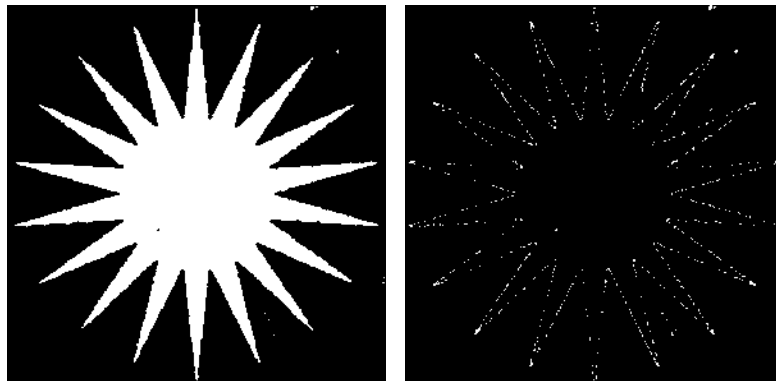


(b) Mapa M_r segmentado correspondiente al mejor coeficiente de Tanimoto.
(c) Error entre los mapas M_o y M_r segmentado con mejor coeficiente de Tanimoto.

Figura 5.3: Fusión de imágenes con el wavelet Haar HH (3.11) y segmentación.



(a) Coeficiente de Tanimoto para las fusiones con el filtro propuesto $G(i, j, \sigma)$. (3.13)



(b) Mapa M_r segmentado correspondiente al mejor coeficiente de Tanimoto. (c) Error entre los mapas M_o y M_r segmentado con mejor coeficiente de Tanimoto.

Figura 5.4: Fusión de imágenes con el filtro propuesto (3.13) y segmentación.

Por último, se muestran los resultados finales obtenidos el método de fusión de imágenes propuesto, definido por el algoritmo 4, usando el filtro (3.13) propuesto en la Sección 3.3, así como los resultados que se obtienen con otros métodos de fusión como la transformada discreta wavelet. La pirámide Laplaciana y el promedio de las imágenes, se realiza una comparación visual de las fusiones resultantes de imágenes sin referencia como las mostradas en las Figuras 5.5 y 5.6.

Dado que no se cuenta con un mapa de decisión de referencia como en el caso

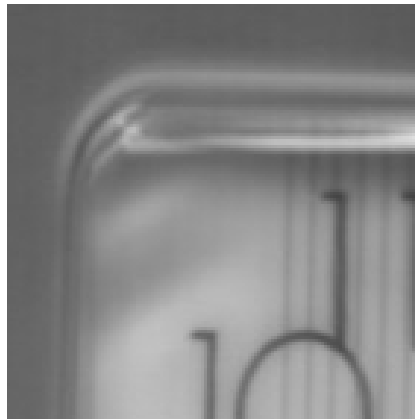
de las imágenes sintéticas y que cómo se mostró en el tercer capítulo, las mediciones usadas en otros trabajos no reflejan la calidad de los resultados de manera confiable. En el caso de el método propuesto en este trabajo, se escoge el mapa de decisión de forma subjetiva observando los mapas que logran separar las regiones de alta frecuencia de forma correcta. La implementación de la transformada wavelet se realizó de acuerdo al tutorial presentado en [Pajares y De la Cruz, 2004] mientras que la pirámide Laplaciana se toma de [Burt y Kolczynski, 1993], el método de promedio de las imágenes usa la Ecuación (2.1) del segundo capítulo.



(a)



(b)



(c)

Figura 5.5: Resultados de fusión usando el promedio de las imágenes, a) resultado de la fusión, b) acercamiento al reloj izquierdo, c) acercamiento al reloj derecho.

En la fusión producida por promedio (2.1) en la Figura 5.5(a) se tiene el resultado con el enfoque más bajo, en los acercamientos de este resultado en las Figuras 5.5(b) y 5.5(c)

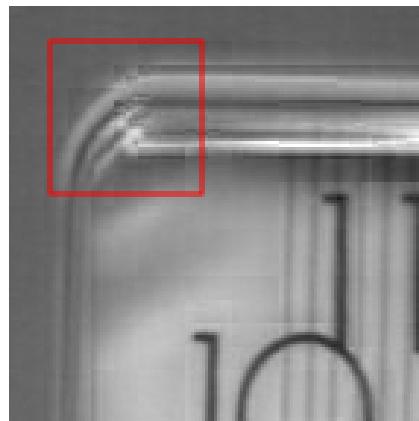
se observa fácilmente el bajo rendimiento de este método. Para la transformada discreta wavelet se tiene el resultado de la Figura 5.6(a) que en general tiene un mejor enfoque que el resultado obtenido con el promedio, pero realizando un acercamiento al reloj izquierdo en del resultado en la Figura 5.6(b), se observa uno de los problemas que presenta el método, que son los artefactos de bloques. Se pueden ver por ejemplo en las manecillas del reloj en forma de líneas verticales u horizontales o en forma de pequeños bloques. El mismo efecto se puede observar en la Figura 5.6(c) que presenta un acercamiento a la esquina del reloj y que presenta más artefactos de bloques.



(a)



(b)



(c)

Figura 5.6: Resultados de fusión usando la transformada discreta wavelet, a) resultado de la fusión, b) acercamiento al reloj izquierdo, c) acercamiento al reloj derecho.

La fusión con la pirámide Laplaciana en la Figura 5.7(a) presenta un mejor resultado al promedio y la transformada discreta wavelet, pero aun presenta pequeñas distorsiones

en la imagen resultante, que son visibles en los acercamientos a las letras del reloj en la Figura 5.7(b) y a la esquina del reloj izquierdo en la Figura 5.7(c).

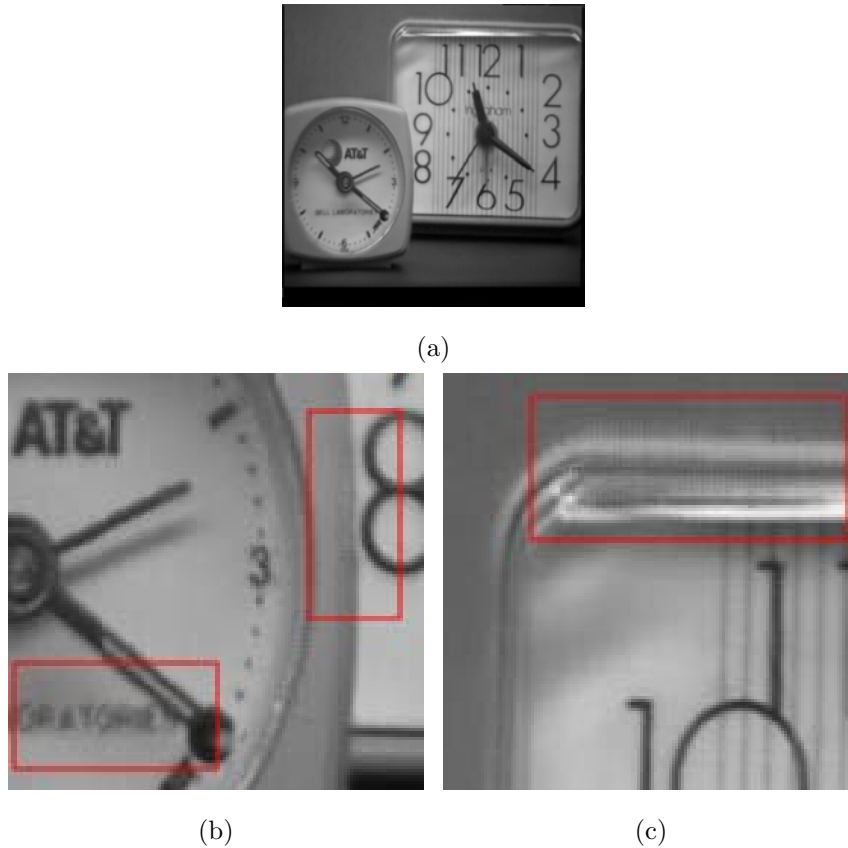


Figura 5.7: Resultados de fusión usando la pirámide Laplaciana, a) resultado de la fusión, b) acercamiento al reloj izquierdo, c) acercamiento al reloj derecho.

Finalmente se presentan los resultados obtenidos usando el método propuesto, que es descrito por el Algoritmo 4. Dado que el método produce un valor distinto para cada valor de λ y que no se tiene una imagen de referencia para las imágenes con las que se prueba, se escoge el mapa de decisión que con un criterio cualitativo presenta el mejor resultado. En la Figura 5.8(a) se tiene el mapa de decisión seleccionado sobre la imagen resultante para indicar cuales fueron las regiones tomadas de las imágenes originales para formar la imagen fusionada presentada en la Figura 5.8(b), en los acercamientos del resultado en las Figuras 5.8(c) y 5.8(d) se muestra cómo el método produce el resultado con mejor enfoque,

aunque el mismo proceso de selección de regiones que da este resultado, produce las líneas con cambio de contraste cercanas al ocho del reloj en la imagen 5.8(c) y sobre la esquina del reloj derecho en la Figura 5.8(d).

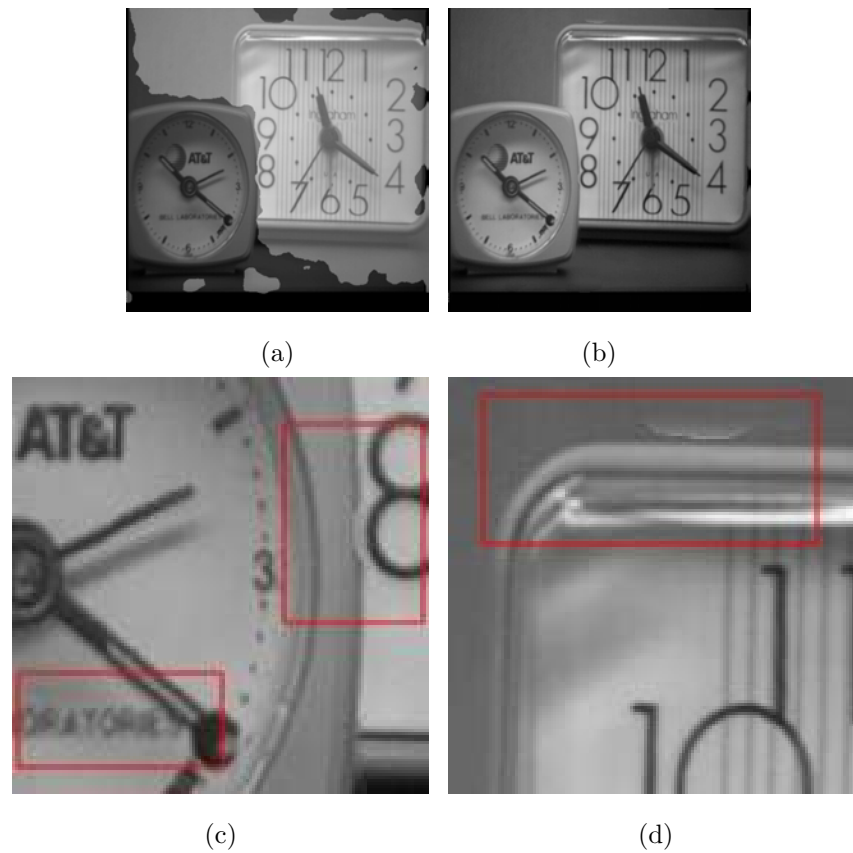


Figura 5.8: Resultados de fusión usando el método propuesto, a) resultado de la fusión, b) acercamiento al reloj izquierdo, c) acercamiento al reloj derecho.

En las pruebas siguientes, se presenta la fusión de dos conjuntos de imágenes multifoco. En ambos casos las imágenes fueron tomadas realizando un cambio de enfoque en tres objetos a distancias distintas. En el primer caso en la Figura 5.9 se presentan las imágenes originales a fusionar en 5.9(a) a 5.9(c), seguidas de los resultados obtenidos de los métodos de fusión. En la Figura 5.9(h) se muestra el mapa de decisión M_r obtenido por el filtro pasa altas propuesto. En esta prueba, dado que se tienen tres distancias de enfoque, se usa un mapa de decisión ternario.

Al formar el mapa de decisión, el filtro puede encontrarse con regiones continuas, es decir, regiones sin textura, bordes u otra forma de diferenciarse de otras imágenes en cuanto a enfoque se refiere. Un ejemplo de esto sucede en la prueba de la Figura 5.9, en las imágenes esta región se trata de la pared en el fondo de la escena. Ya que el filtro no puede diferenciar entre las regiones de la pared, este asigna estas regiones a un plano de enfoque u otro, de forma alternada o discontinua para formar el mapa de decisión M_r . A pesar de no poder diferenciar estas regiones el filtro presenta un buen resultado ya que estas regiones no presentan información útil en forma de regiones de alto enfoque, por lo que pueden ser asignadas a uno u otro plano de enfoque sin disminuir la calidad de la fusión.

En la Figura 5.9(i) se muestra el mapa segmentado, se puede observar como el método de segmentación mejora la calidad del mapa, presentándolo con regiones suaves o continuas. Se observan resultados de fusión similares, entre el método propuesto y el método de pirámide de Laplace, mientras que en la transformada discreta wavelet el resultado presenta distorsiones de artefactos de bloque como en las pruebas anteriores. En la última prueba presentada en la Figura 5.11 se obtienen resultados similares. Con los métodos de promedio y transformada wavelet teniendo los resultados inferiores y los métodos de pirámide Laplaciana y el propuesto similares entre si y mejores que el resto.

En la Figura 5.10 se muestra un acercamiento a las fusiones resultantes del método propuesto y el de la pirámide Laplaciana. En este ejemplo se puede observar mas fácilmente las ventajas y desventajas de cada método. En el caso del método propuesto, este presenta líneas de segmentación cerca de las regiones que presentan un cambio de enfoque en las imágenes originales, estas se pueden observar cerca del asa. Para el método de la pirámide Laplaciana, se presentan también distorsiones en las mismas regiones donde cambia el enfoque de las imágenes, además se puede observar en este ejemplo, como las letras en la taza se encuentran mejor definidas en el método propuesto.



(a)

(b)

(c)



(d) Promedio.



(e) Transformada wavelet.



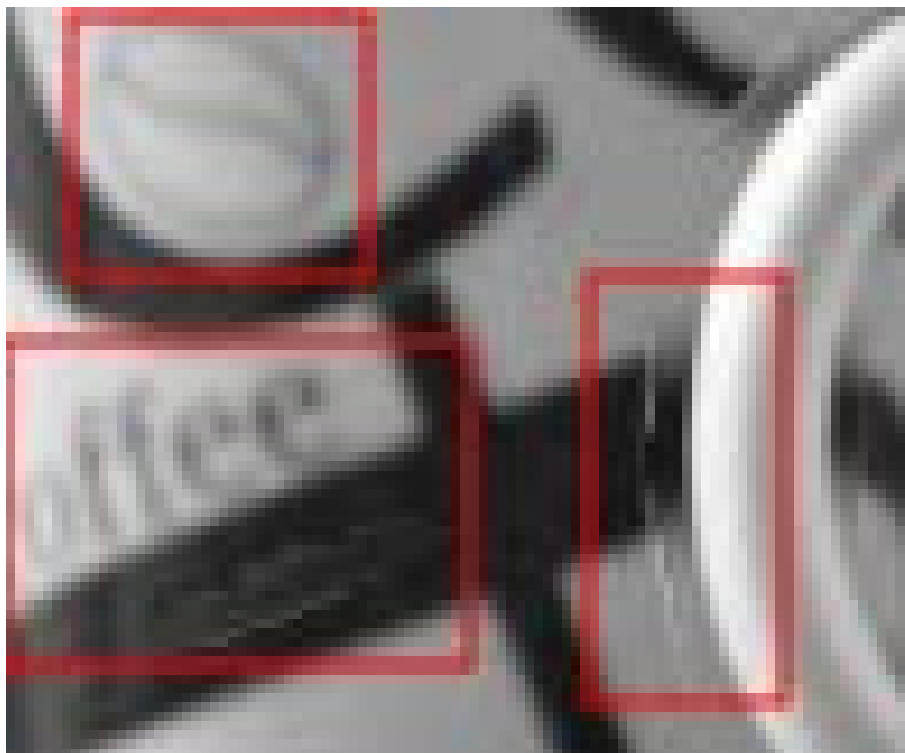
(f) Pirámide Laplaciana.



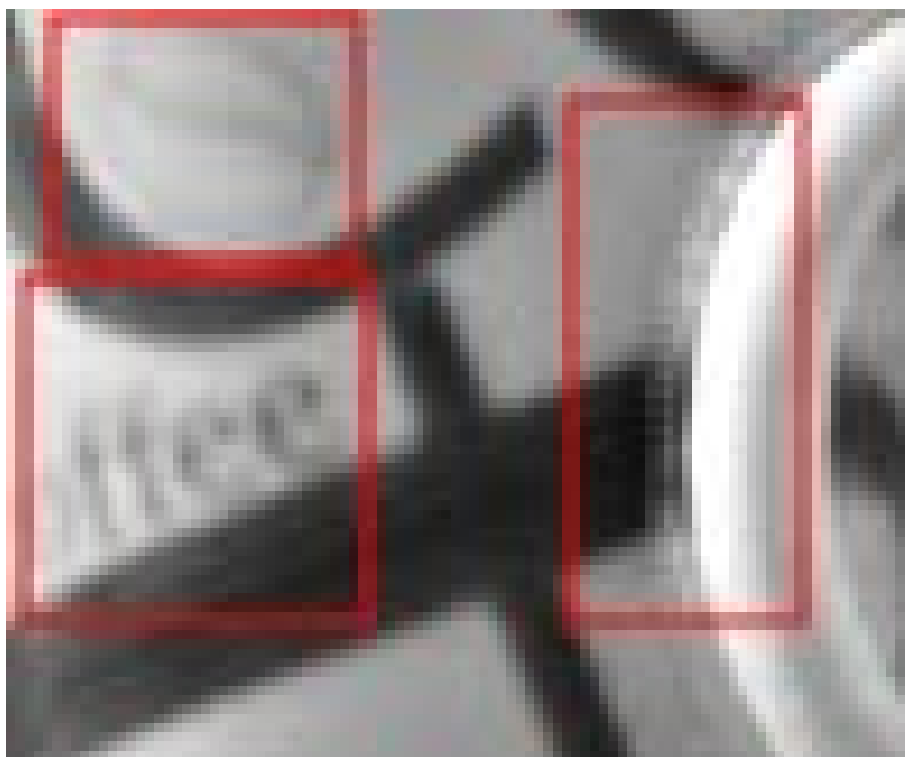
(g) Método propuesto.

(h) Mapa M_r .(i) Mapa M_r segmentado.

Figura 5.9: Resultados de los diferentes métodos de fusión.



(a) Método propuesto.



(b) Pirámide Laplaciana.

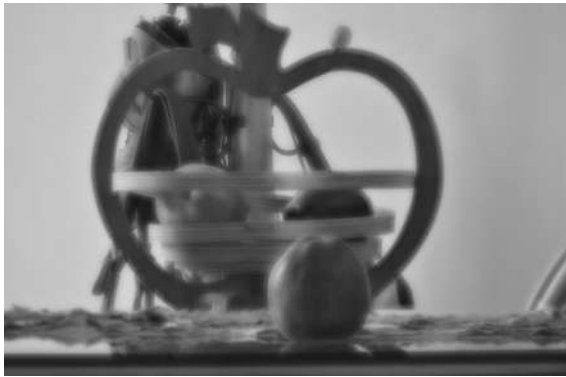
Figura 5.10: Comparación del método propuesto y el método de pirámide Laplaciana.



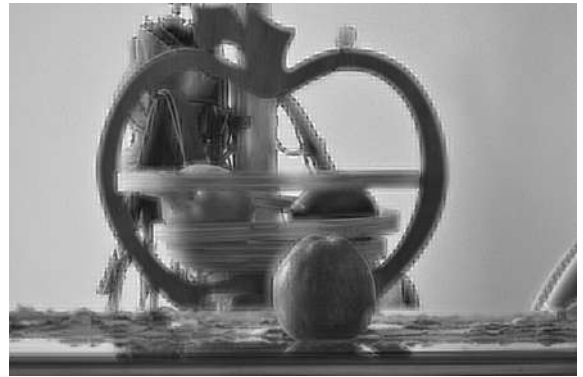
(a)

(b)

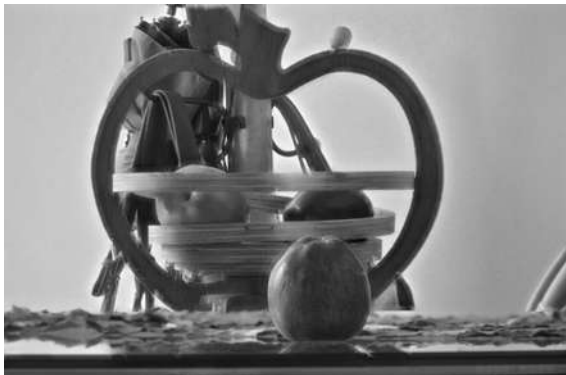
(c)



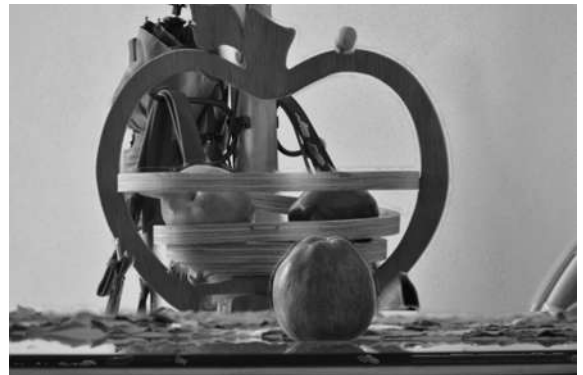
(d) Promedio.



(e) Transformada wavelet.



(f) Pirámide Laplaciana.



(g) Método propuesto.

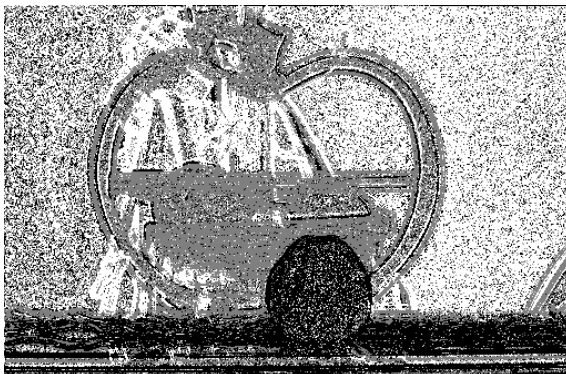
(h) Mapa M_r .(i) Mapa M_r segmentado.

Figura 5.11: Resultados de los diferentes métodos de fusión.

5.1. Conclusiones del capítulo

Al terminar las pruebas se concluye que el método propuesto presenta los mejores resultados. El método de la transformada discreta wavelet, aunque presenta una mejora en el enfoque en comparación con el método de promedio de imágenes, también presenta errores en forma de artefactos de bloques. De la misma forma, en el método de pirámide Laplaciana se presentan distorsiones en el resultado, aunque no son tan visibles como en el caso de la transformada wavelet. El método propuesto presenta un buen resultado ya que hace una selección de áreas de alto enfoque de forma directa de los píxeles en las imágenes originales, produciendo menos efectos de distorsiones en comparación con los otros métodos.

Capítulo 6

Conclusiones y trabajos futuros

6.1. Conclusiones Generales

En este trabajo se realizó una introducción a un problema común en cámaras y dispositivos ópticos. También se presentaron los métodos más usados en la solución del mismo, los cuales representan la base de muchos otros trabajos de fusión de imágenes o que al menos son usados como comparación. Se mostró la relación entre las frecuencias en la imagen y el enfoque que tienen.

Se introdujo un método de fusión basado en la selección de regiones de alto enfoque usando como criterio los valores de frecuencia en las imágenes. En base a los experimentos realizados, comparando diferentes funciones como filtros de frecuencia, se determina la mejor forma de comparar las imágenes a fusionar, produciendo un método de fusión de imágenes multifoco. Por último, se presentó además una mejora del resultado de este método por medio de un algoritmo de segmentación basado en probabilidad, incluyendo experimentos que demuestran esta mejora.

Dentro de las complicaciones encontradas en la realización del trabajo, la mayor de éstas, fue el tener una medida de comparación objetiva de los resultados, ya que muchas de las encontradas en la literatura presentaban resultados incongruentes con los experimentos realizados. Estas medidas en realidad si presentan una medición del enfoque de la imagen, sin embargo todas presentaban el mismo problema que es el error en la medición por distorsiones

en los resultados de los métodos por ejemplo en forma de artefactos de bloques. El ejemplo más sencillo de este error de medición de calidad, se puede ver en el caso de la medición de promedio de gradiente, del resultado de la fusión de imágenes usando la transformada discreta wavelet.

Recordando que la transformada discreta wavelet presentaba artefactos de bloques, las mediciones obtenidas con el promedio de gradientes, presentaban medidas más altas o de mejor calidad en imágenes con artefactos de bloques. Aun cuando se utiliza un ejemplo con imágenes sintéticas, el promedio de gradiente mide una mejor calidad en la imagen fusionada con artefactos de bloques, en comparación con la imagen de referencia con la que se generan las sintéticas. Lo que quiere decir que la medición no es apropiada para el método, este mismo tipo de situaciones son exploradas en [Hossny y Nahavandi, 2009] quien determina que cada una de estas medidas es adecuada para métodos de fusión distintos.

6.2. Trabajos Futuros

- El aspecto que más importancia tiene en cuanto investigaciones futuras sería el de obtener una medición de calidad de fusión de imágenes adecuada para diferentes métodos de fusión. La medida más adecuada sería una que no sea afectada por distorsiones inherentes a los mismos métodos, cómo artefactos de anillos o de bloques, o en el caso del método propuesto y de métodos de selección de regiones, los límites de las regiones segmentadas presentaban un cambio de contraste que afectaba el valor obtenido de estas mediciones. Esta es una de las razones por la cual se usó el coeficiente de Tanimoto para comparar resultados.
- Uno de los problemas encontrados en la fusión de imágenes multifoco es la necesidad de realizar un registro de imágenes. Al capturar imágenes con diferentes enfoques, ya sea por cambiar la distancia del lente a los objetos, o por cambios en el nivel de acercamiento del lente entre la primera imagen y la segunda, se presentan diferencias entre las imágenes, no solo de enfoque sino también espaciales. Esto quiere decir que algunos objetos en la escena en una imagen parecerán de diferentes tamaños o con ángulos distintos en la otra. Además, al capturar las fotografías en casos reales y prácticos es

muy improbable que la escena se mantenga fija por los movimientos de la cámara al tomar las fotografías. Estas diferencias espaciales presentan un error considerable en el resultado de la fusión. El registro de imágenes proporciona una solución a este problema al realizar transformaciones geométricas e.g., traslación, rotación, escalamiento, etc... a la imagen lo que le permite alinearlas espacialmente una de otra, de forma que todas las imágenes parezcan ser obtenidas desde la misma posición. Dada la facilidad con la que se da un problema de registro, agregar un método de registro de imágenes al método de fusión propuesto en este trabajo sería una mejora considerable.

- Entre los detalles que se pueden explorar para trabajos futuros, se encuentra una investigación más detallada de los diferentes niveles de actividad usados para la selección de regiones de alto enfoque. En este trabajo solo se exploró el método de selección del píxel con el mayor valor absoluto en la respuesta de frecuencia de la imagen a algún filtro. Se puede también investigar el uso de la segmentación en esquemas de fusión de métodos multiescala como la pirámide Laplaciana o la transformada discreta wavelet para mejorar el resultado de la fusión con estos métodos.
- Por último sería de interés agregar al método propuesto en este trabajo, una manera de determinar los parámetros más adecuados para realizar la fusión es decir una λ dinámica o adecuada para cualquier conjunto de imágenes multifoco. También existe la posibilidad de usar algún otro criterio de convergencia a la solución de la segmentación, en lugar de usar un número determinado de iteraciones.

Glosario

CCD Los Dispositivos de Carga Acoplada son sensores capturadores de imágenes donde cada pixel es representado por un capacitor. Cuando la luz hace contacto con el capacitor, esta es almacenada como una carga eléctrica.

CMOS Semiconductor Oxido Metalico Complementario Dispositivos capturadores de imágenes, donde cada pixel usa un semiconductor CMOS como un sensor de luz.

Material fotosensible Material compuesto por una capa que reacciona al contacto con la luz formando una imagen latente. En las cámaras analógicas se trata del rollo de film, que ha sido reemplazado por sensores digitales.

Microscopía Es el campo técnico del uso de microscopios para la examinación de muestras u objetos, que no pueden ser vistos o no están en el rango de visualización humana.

Multiplicadores de Lagrange Método de optimización para buscar el máximo o el mínimo de una función, sujeta a restricciones de igualdad.

TDF La transformada discreta de Fourier es una representación de una señal discreta en el dominio de tiempo como una suma de exponenciales complejas en el dominio de frecuencia.

Teledetección Observación remota de la superficie terrestre, via imagenes satelitales o aereas, la traducción al inglés es remote sensing.

Referencias

- [Allen y Triantaphillidou, 2010] Allen, E. y Triantaphillidou, S. (2010). *The Manual of Photography and Digital Imaging*, volumen 3. Taylor & Francis.
- [Alpaydin, 2009] Alpaydin, E. (2009). *Introduction to Machine Learning (Adaptive Computation and Machine Learning series)*. The MIT Press.
- [Blasch y Boulevard, 2008] Blasch, E. y Boulevard, C. (2008). Image Quality Assessment for Performance Evaluation of Image Fusion. *Society*, paginas 583–588.
- [Bovik y Zhou, 2002] Bovik, A. C. y Zhou, W. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84.
- [Burt y Adelson, 1983] Burt, P. J. y Adelson, E. H. (1983). The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4):532–540.
- [Burt y Kolczynski, 1993] Burt, P. J. y Kolczynski, R. (1993). Enhanced image capture through fusion. In *1993 (4th) International Conference on Computer Vision*, paginas 173–182. IEEE Computer Society Press.
- [Eskicioglu y Fisher, 1995] Eskicioglu, A. y Fisher, P. (1995). Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965.
- [Gao et al., 2005] Gao, L.-Q., Wang, R., Yang, S., y Chai, Y.-H. (2005). An image fusion algorithm based on RBF neural networks. *2005 International Conference on Machine Learning and Cybernetics*, 8:5194–5199.
- [Gini, 1955] Gini, C. (1955). *Variability and Mutability*. Memorie di metodologica statistica.

- [Goshtasby, 2006] Goshtasby, A. A. (2006). Fusion of multifocus images to maximize image information. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volumen 6229, paginas 17–21.
- [Hammersley y Clifford, 1971] Hammersley, J. M. y Clifford, P. (1971). Markov fields on finite graphs and lattices.
- [Hariharan et al., 2007] Hariharan, H., Koschan, A., y Abidi, M. (2007). Multifocus Image Fusion by Establishing Focal Connectivity. In *2007 IEEE International Conference on Image Processing*, paginas III – 321–III – 324.
- [Hossny y Nahavandi, 2009] Hossny, M. y Nahavandi, S. (2009). Image fusion algorithms and metrics duality index. In *2009 16th IEEE International Conference on Image Processing ICIP*, paginas 2193–2196.
- [Hu et al., 2010] Hu, X., Lu, H., Zhang, L., y Serikawa, S. (2010). A New Type of Multifocus Image Fusion Method Based on Curvelet Transforms. *2010 International Conference on Electrical and Control Engineering (ICECE)*., paginas 172–175.
- [Indhumadhi y Padmavathi, 2011] Indhumadhi, N. y Padmavathi, G. (2011). Enhanced Image Fusion Algorithm Using Laplacian Pyramid and Spatial frequency Based Wavelet Algorithm. *International Journal on Computer Science and Engineering*, 1(5):298–303.
- [Jaccard, 1901] Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:241–272.
- [Jenkins y White, 1957] Jenkins, F. A. y White, H. E. (1957). *Fundamentals of optics*. McGraw-Hill.
- [Langford et al., 2012] Langford, M., Fox, A., y Smith, R. S. (2012). *Langford's Basic Photography: The Guide for Serious Photographers*. CRC Press.
- [Leung et al., 2001] Leung, L. W., King, B., y Vohora, V. (2001). Comparison of image data fusion techniques using Entropy and INI. *22nd Asian Conference on Remote Sensing...*, 5(9):152–157.

- [Li et al., 2013] Li, Q., Du, J., y Xu, L. (2013). Multi-Focus Image Fusion Using the Local Fractal Dimension. *International Journal of Advanced Robotic Systems*, 10(251):260–271.
- [Li et al., 2001] Li, S., Kwok, J. T., y Wang, Y. (2001). Combination of images with diverse focuses using the spatial frequency. *Information Fusion*, 2(3):169–176.
- [Li et al., 2002] Li, S., Kwok, J. T., y Wang, Y. (2002). Multifocus Image Fusion using Artificial Neural Networks. *Pattern Recognition Letters*, 23(8):985–997.
- [Li y Yang, 2008] Li, S. y Yang, B. (2008). Multifocus image fusion using region segmentation and spatial frequency. *Image and Vision Computing*, 26(7):971–979.
- [Li et al., 2010] Li, X., He, M., y Roux, M. (2010). Multifocus image fusion based on redundant wavelet transform. *IET Image Processing*, 4(4):283–293.
- [Lowe, 2004] Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110.
- [Madhavi y K Ashok, 2011] Madhavi, R. y K Ashok, B. (2011). An all Approach for Multi-Focus Image Fusion Using Neural Network. *International Journal of Computer Science and Telecommunications*, 2(8):23–29.
- [Marr y Hildreth, 1980] Marr, D. y Hildreth, E. C. (1980). Theory of Edge Detection. *Proceedings of the Royal Society B: Biological Sciences*, 207(1167):187–217.
- [Maruthi, 2011] Maruthi, R. (2011). Spatial Domain Method for Fusing MultiFocus Images using Measure of Fuzziness. *International Journal of Computer Applications*, 20(7):48–51.
- [Nava et al., 2008] Nava, R., Escalante Ramírez, B., y Cristóbal, G. (2008). A novel multi-focus image fusion algorithm based on feature extraction and wavelets. In *Photonics Europe*, paginas 700028–700028–10. International Society for Optics and Photonics.
- [Pajares y De la Cruz, 2004] Pajares, G. y De la Cruz, J. M. (2004). A wavelet-based image fusion tutorial. *Pattern Recognition*, 37(9):1855–1872.
- [Rangarajan y Chellappa, 1995] Rangarajan, A. y Chellappa, R. (1995). *Markov Random Field Models in Image Processing*. MIT Press.

- [Rivera et al., 2007] Rivera, M., Ocegueda, O., y Marroquin, J. L. (2007). Entropy-controlled quadratic markov measure field models for efficient image segmentation. *IEEE Transactions on Image Processing*, 16(12):3047–3057.
- [Shnayderman et al., 2003] Shnayderman, A., Gusev, A., y Eskicioglu, A. M. (2003). Multidimensional image quality measure using singular value decomposition. *Proceedings of SPIE image quality and system performance*, 5294(1):82–92.
- [Stollnitz y Deroose, 1995] Stollnitz, E. J. y Deroose, T. D. (1995). Wavelets for Computer Graphics: A Primer. *Computer Graphics and Applications, IEEE*, 15(3):76 – 84.
- [Toet, 1989] Toet, A. (1989). Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters*, 9(4):245–253.
- [Wang y Chang, 2011] Wang, W. y Chang, F. (2011). A multi-focus image fusion method based on laplacian pyramid. *Journal of Computers*, 6(12):2559–2566.
- [Woods, 2008] Woods, G. (2008). *Digital Image Processing 3rd edition*. Prentice Hall.
- [Zaveri et al., 2009] Zaveri, T., Zaveri, M., Shah, V., y Patel, N. (2009). A Novel Region Based Multifocus Image Fusion Method. In *2009 International Conference on Digital Image Processing*, paginas 50–54. IEEE.
- [Zhong y Blum, 1999] Zhong, Z. y Blum, R. S. (1999). A Categorization of Multiscale-decomposition-based Image Fusion Schemes with a Performance Study for a Digital Camera Application. *Proceedings of the IEEE*, 87(8):1315–1326.