

# Fotografía computacional: aperturas codificadas para estimación de profundidad y corrección de desenfoque

PROYECTO FIN DE CARRERA

Autora: Lara Presa Irazábal

Directora: Belén Masiá Corcoy

Ponente: Dr. Diego Gutiérrez Pérez

Ingeniería Superior de Telecomunicación

Curso 2010-2011

Departamento de Informática e Ingeniería de Sistemas

Centro Politécnico Superior

Universidad de Zaragoza



Departamento de  
Informática e Ingeniería  
de Sistemas  
**Universidad** Zaragoza



Centro  
Politécnico Superior  
**Universidad** Zaragoza



Julio de 2011



*You don't take a photograph, you make it*  
Ansel Adams (1902-1984)

*No photographer is as good as the simplest camera*  
Edward Steichen (1879-1973)

*I eagerly await new concepts and processes. I believe that the electronic image will be  
the next major advance. Such systems will have their own inherent and inescapable  
structural characteristics, and the artist and functional practitioner will again strive to  
comprehend and control them*  
Ansel Adams (1902-1984)



## Resumen

La expansión e implantación de la fotografía digital en los últimos años ha hecho surgir el concepto de Fotografía Computacional, un nuevo campo multidisciplinar que abarca temas de óptica, procesamiento de imágenes, informática gráfica, percepción o visión por computador.

Una cámara convencional captura una proyección 2D de escenas reales 3D. Uno de los problemas inherentes a este proceso de captura es la limitada profundidad de campo, que causa la aparición de zonas borrosas en partes de la escena que se encuentran alejadas del plano focal. Otro problema, objeto de muchos trabajos de investigación, es la recuperación de las profundidades de la escena, perdidas en una captura convencional. El objetivo de este proyecto es la obtención tanto de la información de profundidad de la escena, como de una imagen nítida recuperando el enfoque en las zonas borrosas (*"deblurring"*).

En este proyecto se propone como solución el uso de aperturas codificadas, ya usadas durante décadas en campos como la óptica o la astronomía, insertando una máscara en el objetivo de la cámara que modifique la apertura circular de la cámara; obteniendo una respuesta en el dominio frecuencial más favorable que la de una apertura circular convencional. Lo que hacen, esencialmente, estos patrones, es codificar ópticamente la luz que llega al sensor de una forma en que, tras la posterior decodificación, se pueda recuperar más información de la escena que con una apertura convencional. Esto nos permite estimar, mediante el análisis del grado de desenfoque de la zona de la escena, la profundidad a la que se encuentra respecto del plano de enfoque, lo que se conoce como "profundidad a partir del desenfoque" (DFD, *depth from defocus*). Asimismo permite una recuperación de la imagen enfocada mediante técnicas de deconvolución más o menos complejas. La imagen nítida resultante y el mapa de profundidades pueden ser combinados para múltiples e interesantes aplicaciones fotográficas, que incluyen segmentación de la escena automática, selección de distintos objetos o profundidades de la escena y reenfoque como postproceso.

Se explora en este proyecto tanto el uso de una única apertura codificada como el uso de pares de aperturas codificadas. En primer lugar se selecciona bajo ciertos criterios una apertura codificada diseñada para DFD. Para esta apertura se estudia la obtención de mapas de profundidad y además se analiza el problema de re-enfoque (*"deblurring"*) mediante el estudio y comparativa de distintos métodos de deconvolución. En segundo lugar se implementa un método de optimización que permite obtener pares de aperturas codificadas óptimas para DFD. Finalmente se realizan validaciones tanto por simulación como sobre soporte físico de las aperturas codificadas seleccionadas; mostrando y realizando una comparativa de los resultados obtenidos.

Además se explora la introducción de valores no binarios en los patrones de aperturas codificadas. En la mayoría de estudios previos se indica que se limita el espacio de búsqueda a píxeles opacos y transparentes principalmente por costes computacionales, sin existir otras razones de peso, por lo que se decide diseñar aperturas codificadas no binarias (semitransparentes). Aunque los resultados obtenidos no son en ningún caso concluyentes, los primeros indicios muestran que no debe limitarse el espacio de búsqueda de la apertura codificada a espacios binarios, ya que las aperturas no binarias presentan resultados mejores. Este estudio de investigación se incluyó en una publicación aceptada en el Congreso Ibero-americano de Informática Gráfica (SIAGC 2011), resultando entre las tres mejores publicaciones del congreso.



# Índice general

<b>I Memoria</b>	<b>9</b>
<b>1. Introducción</b>	<b>11</b>
1.1. Objetivo, alcance y contexto del proyecto . . . . .	11
1.2. Estructura del trabajo . . . . .	12
1.3. Planificación . . . . .	12
<b>2. Las aperturas codificadas en fotografía computacional</b>	<b>13</b>
2.1. Introducción . . . . .	13
2.2. Desenfoque y profundidad de campo . . . . .	14
2.3. Codificando el desenfoque: análisis en frecuencia . . . . .	15
2.4. Soluciones propuestas . . . . .	18
<b>3. Apertura codificada óptima</b>	<b>19</b>
3.1. Introducción . . . . .	19
3.2. Diseño y obtención de la apertura codificada . . . . .	20
3.2.1. Criterio de selección del filtro . . . . .	21
3.3. Deconvolución sparse . . . . .	23
3.4. Recuperación de profundidad . . . . .	24
3.4.1. Obtención del vector de pesos . . . . .	25
3.5. Resultados . . . . .	26
3.5.1. Calibración . . . . .	26
3.5.2. Captura de imágenes . . . . .	28
3.5.3. Resultados del proceso de <i>deblurring</i> . . . . .	30
3.5.4. Resultados de la obtención del mapa de profundidades . . . . .	37
<b>4. Pares de aperturas codificadas</b>	<b>39</b>
4.1. Introducción . . . . .	39
4.2. Función objetivo . . . . .	39
4.3. Algoritmo genético para obtención de pares de aperturas codificadas . . . . .	42
4.4. Recuperación de profundidad . . . . .	45
4.5. Resultados . . . . .	45
4.5.1. Método de validación empleado . . . . .	45
4.5.2. Resultados de la obtención de mapas de profundidades . . . . .	47
4.5.3. Resultados del proceso de <i>deblurring</i> . . . . .	49
<b>5. Aperturas no binarias para <i>deblurring</i></b>	<b>53</b>
5.1. Introducción . . . . .	53
5.2. Obtención de aperturas no binarias optimizadas para <i>deblurring</i> . . . . .	53
5.3. Análisis de aperturas no binarias . . . . .	54
5.4. Resultados y comparativa sobre imágenes reales . . . . .	55

<b>6. Conclusiones</b>	<b>57</b>
6.1. Conclusiones del trabajo realizado . . . . .	57
6.2. Trabajo futuro . . . . .	58
6.3. Conclusiones personales . . . . .	58
<b>II Bibliografía</b>	<b>61</b>
<b>III Apéndices</b>	<b>65</b>
<b>A. Deconvolución basada en priors de imágenes naturales</b>	<b>67</b>
A.1. Introducción . . . . .	67
A.2. Deconvolución basada en prior de derivadas gaussianas . . . . .	68
A.2.1. Deconvolución dominio frecuencial . . . . .	68
A.2.2. Deconvolución dominio espacial . . . . .	69
A.3. Deconvolución basada en prior de derivadas sparse . . . . .	69
<b>B. Modelo estadístico de imágenes</b>	<b>71</b>
<b>C. Criterio de evaluación par de aperturas</b>	<b>73</b>
C.1. Formulación de DFD ( <i>Depth from defocus</i> ) . . . . .	73
C.2. Criterio de selección . . . . .	74
<b>D. Cómo insertar una apertura codificada en una lente Canon EF 50mm f1.8 II</b>	<b>77</b>
<b>E. Artículo: Coded Apertures for Defocus Deblurring</b>	<b>81</b>

# Índice de figuras

1.1.	Diagrama de Gantt de las actividades realizadas.	12
2.1.	PSF a distintas profundidades para una lente Canon 50mm f1.8 II con apertura convencional parcialmente cerrada.	14
2.2.	Modelo 2D de captura de imagen de una cámara con lente convergente delgada	15
2.3.	Corte 1D de las transformadas de Fourier de las imágenes capturas $Y(v,w)$ para diferentes escalas del filtro.	16
2.4.	Identificación de la escala correcta a partir de la imagen observada.	17
2.5.	Logaritmo de la potencia espectral para la apertura circular, apertura optimizada para DFD y apertura optimizada para <i>deblurring</i> .	17
3.1.	Diagrama de flujo a seguir para la obtención del mapa de profundidades y la imagen recuperada.	19
3.2.	Diagrama de flujo a seguir para la obtención del mapa de profundidades y la imagen recuperada.	20
3.3.	Corte transformadas de Fourier en 1D para tres escalas distintas de la apertura convencional y la apertura codificada escogida como óptima para DFD.	23
3.4.	Representación gráfica del almacenamiento de los errores de reconstrucción	25
3.5.	Cuerpo de cámara reflex empleada y apertura codificada colocada sobre la lente del objetivo.	27
3.6.	Array de diodos LED empleado para la calibración de las PSFs	27
3.7.	PSFs obtenidas sin umbralizar para distintas profundidades	28
3.8.	Ejemplo escena capturada con objetos distribuidos a lo largo de nueve profundidades.	28
3.9.	Geometría del diámetro de blur para lente delgada	29
3.10.	Ringing debido al uso de kernel o PSF de tamaño elevado.	31
3.11.	Comparación de distintos algoritmos de deconvolución de un objeto situado a 45 cm del plano focal y capturado con la apertura codificada.	32
3.12.	Comparación de distintos algoritmos de deconvolución para un objeto situado a 85 cm del plano focal y capturado con la apertura codificada.	32
3.13.	Comparación de distintos algoritmos de deconvolución para un objeto situado a 105 cm del plano focal y capturado con la apertura codificada. (Captura 3, valor de apertura $f = 4.0$ , $d = 1.40m$ ).	33
3.14.	Detalles a 95 cm recuperados con <i>ringing</i> debido al elevado grado de desenfoque.	33
3.16.	Aperturas empleadas para analizar la deconvolución sparse.	33
3.15.	Aparición de artefactos en la deconvolución al usar un kernel de tamaño menor o mayor del adecuado.	34
3.17.	Logaritmo de la potencia espectral para la apertura circular y la apertura codificada diseñada para <i>deblurring</i>	34
3.18.	Apertura circular. Recuperación de objetos desenfocados situados a 40 cm del plano focal con deconvolución sparse y con deconvolución de Wiener.	35

3.19. Apertura circular. Recuperación de objetos desenfocados situados a 70 cm del plano focal con deconvolución sparse y con deconvolución de Wiener. . . . .	36
3.20. Apertura codificada diseñada para deblurring. Recuperación de objetos desenfocados situados a 40 cm del plano focal con deconvolución sparse y con deconvolución de wiener. . . . .	36
3.21. Apertura codificada diseñada para <i>deblurring</i> . Recuperación de objetos desenfocados situados a 70 cm del plano focal con deconvolución sparse y con deconvolución de Wiener. . . . .	37
3.22. Mapas de profundidades considerando y sin considerar vector de pesos. . . . .	38
 4.2. Potencia espectral y fase del par de aperturas convencionales y del par optimizado de aperturas codificadas . . . . .	41
4.3. Función de coste E(d) para el par óptimo de Zhou y el par de aperturas circulares	42
4.4. Esquema del algoritmo genético implementado . . . . .	43
4.5. Pares de aperturas obtenidas para distintos parámetros iniciales del algoritmo genético. . . . .	44
4.6. Convergencia del algoritmo genético de pares de aperturas codificadas para distintos parámetros iniciales. . . . .	44
4.7. Diagrama de flujo del procedimiento de validación sobre soporte físico: obtención del mapa de profundidades $U$ y la imagen enfocada $I$ a partir de las dos imágenes capturadas con el par de aperturas. . . . .	46
4.8. Diagrama de flujo del procedimiento de validación por simulación. . . . .	47
4.9. Imagen usada para la validación del par de aperturas. . . . .	47
4.10. Pares de aperturas incluidas en la validación por simulación junto con las obtenidas en el algoritmo genético. . . . .	47
4.11. Porcentajes de error en el cálculo del mapa de profundidades para cada par de aperturas. . . . .	48
4.12. Logaritmo de la potencia espectral para el par óptimo de Zhou y nuestro par con mejor evaluación del genético. . . . .	48
4.13. Estimación del mapa de profundidades para distintos pares de aperturas. . . . .	49
4.14. Simulación de desenfoque y <i>deblurring</i> para distintas aperturas y pares de aperturas. . . . .	51
4.15. Porcentaje de error según la métrica L2 en la recuperación de imagen enfocada para cada tipo de apertura. . . . .	52
4.16. Resultados métrica SSIM para cada tipo de aperturas . . . . .	52
 5.1. Valor promedio de la métrica R(k) y tiempo promedio hasta la convergencia del algorimo para aperturas binarias y no binarias. . . . .	54
5.2. Valor promedio de la norma $L_2$ para aperturas binarias y no binarias y para distintos parámetros del algoritmo genético. . . . .	55
5.3. Apertura binaria y no binaria con mejor resultado del criterio de evaluación del algoritmo genético. . . . .	56
 D.1. Canon EOS 500D con objetivo Canon 50mm f1.8 II. . . . .	77
D.2. Primeros pasos para desmontar el objetivo. . . . .	78
D.3. Componentes de la lente Canon EF 50mm f1.8 . . . . .	78
D.4. Canon EF 50mm f1.8 II con apertura codificada. . . . .	78

# Índice de cuadros

3.1. Parámetros para cada captura de imágenes . . . . .	30
3.2. Parámetros para cada captura de imágenes . . . . .	35
4.1. Resultados norma L2 para cada tipo de apertura. . . . .	50



# **Parte I**

# **Memoria**



# Capítulo 1

## Introducción

### 1.1. Objetivo, alcance y contexto del proyecto

El objetivo de este trabajo es estudiar y analizar del uso de aperturas codificadas en la fotografía digital y en especial su contribución para la estimación de profundidad y corrección de desenfoque. El trabajo se ha realizado dentro del grupo de investigación GIGA (Grupo de Informática Gráfica Avanzada) en la línea de investigación de un nuevo campo multidisciplinar, la Fotografía Computacional.

En los últimos años, el campo de la Fotografía Computacional ha mostrado importantes avances en el procesado de imágenes. La idea principal es codificar la luz de distintos modos antes de que ésta alcance el sensor, de modo que su posterior decodificación permita obtener más información de la escena. Para tal fin existen varios métodos, desde el uso de luz estructurada pasando por nuevos dispositivos ópticos hasta aperturas moduladas o obturador. En este trabajo se optará por hacer uso de aperturas codificadas.

El uso de aperturas codificadas consiste en insertar una máscara de píxeles opacos y transparentes sobre la lente de la cámara, codificando la luz antes de alcanzar el sensor de la cámara, de modo que permita, al decodificar la imagen posteriormente, obtener más información que con el uso de la apertura convencional.

El objetivo de este proyecto es doble:

1. En primer lugar se desea obtener información de profundidad de una escena capturada.
2. En segundo lugar se quiere recuperar el enfoque en zonas de la escena que aparecen borrosas <sup>1</sup>.

El grado de desenfoque de la escena, como veremos, variará en función de la profundidad; por lo tanto a partir del análisis del grado de desenfoque o cantidad de *blur* presente en esa zona de la escena se podrá estimar su profundidad. Esto es lo que se conoce como profundidad a partir de desenfoque (DFD, *depth from defocus*).

En primer lugar se hace uso de **una única apertura codificada diseñada para DFD** partiendo del trabajo previo de Levin *et al* [10]. Así, partiendo de una única imagen se requiere estimar su profundidad y recuperar el enfoque de objetos borrosos. En el proceso de *deblurring* se usará un método de deconvolución novedoso, basado en priors de derivadas sparse. Se realiza una serie de capturas de imágenes reales sobre las que se compara y analiza los resultados de *deblurring* con otros métodos de deconvolución. Además, procesando las imágenes de las capturas, se verá que la estimación del mapa de profundidades de la escena a partir de una única imagen presenta una fuerte dependencia con el sistema y la imagen, de modo que se decide explorar otro método más eficiente.

---

<sup>1</sup>Hay principalmente dos causas para la aparición de zonas borrosas: el movimiento de la cámara o del objeto a capturar, o el desenfoque. Este proyecto trata estas últimas.

El segundo método optará por el uso de **parejas de aperturas codificadas diseñadas para DFD** (Zhou *et al.*[23]). En primer lugar se obtienen pares de aperturas codificadas óptimas para DFD mediante un algoritmo genético. Posteriormente para los pares de aperturas seleccionados, se realiza una comparativa, mediante validación por simulación, tanto de los resultados de *deblurring* como de la estimación del mapa de profundidades.

Adicionalmente, se decide explorar la **introducción de valores no binarios (semi-transparentes) en el diseño de aperturas codificadas para deblurring**. Este estudio se incluyó en un artículo de investigación aceptado en el Congreso Ibero-americano de Informática Gráfica (SIAGC 2011).

## 1.2. Estructura del trabajo

El trabajo realizado consta principalmente de tres partes que comprenden el Capítulo 3, Capítulo 4 y Capítulo 5. De esta manera, el Capítulo 2 introduce los conceptos teóricos tratados a lo largo del trabajo, introduciendo y justificando el uso de aperturas codificadas en Fotografía Computacional. En el Capítulo 3 se explora el uso de una única apertura codificada óptima [10], realizando además un estudio comparativo entre distintos métodos de deconvolución y sus resultados. Posteriormente, en el Capítulo 4 se considera uno de los métodos de DFD que captura dos imágenes mediante el uso de un par de aperturas codificadas [23]. Finalmente, en el Capítulo 5, se estudia la ampliación del espacio de búsqueda en el diseño de aperturas codificadas para *deblurring* a valores no binarios. El Capítulo 6 sintetizará las conclusiones derivadas de los resultados anteriores.

## 1.3. Planificación

Durante los seis meses aproximados de duración del proyecto, se ha dividido el trabajo en tareas de documentación, estudio de aperturas no binarias, implementación de métodos para la obtención de aperturas para DFD, simulación y comparativa de distintos métodos de deconvolución o *deblurring*, implementación de la obtención de mapas de profundidad, validación tanto por simulación como en soporte físico de las aperturas codificadas y redacción de la memoria. Tanto para las tareas de implementación como de simulación se ha empleado el software matemático Matlab. La distribución a lo largo del tiempo se puede ver en la Figura 1.1.

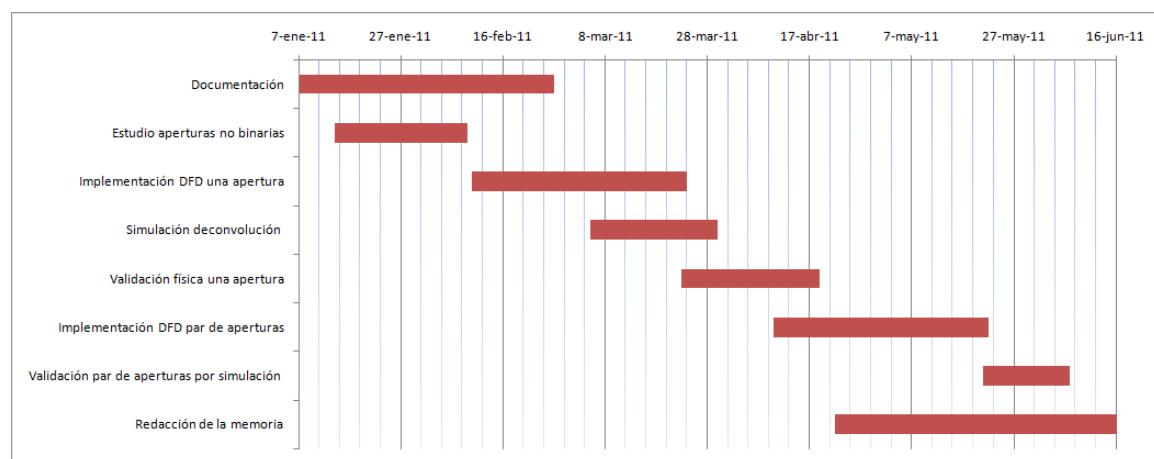


Figura 1.1: Diagrama de Gantt de las actividades realizadas.

## Capítulo 2

# Las aperturas codificadas en fotografía computacional

### 2.1. Introducción

El uso inicial de aperturas codificadas surge en el contexto del campo de la astronomía y de la medicina, donde permitieron mejorar la relación señal a ruido captando mayor información de rayos gamma y rayos X, ya que el uso de lentes tradicionales no podían ser usadas a esas longitudes de onda [2, 5]. Mientras, en el campo de la óptica, el uso de aperturas codificadas ha permitido el incremento de la profundidad de campo y mejorar la respuesta en frecuencia de la apertura convencional.

En los últimos años, las aperturas codificadas han ido tomando mayor importancia en visión por computador y fotografía computacional [12, 20] Una de las principales aplicaciones en este campo es la mejora del proceso de *deblurring*, recuperación de imágenes nítidas a partir de imágenes desenfocadas [20, 24]. Para conseguir esto, se diseñan aperturas codificadas cuya respuesta en frecuencia elimine las limitaciones de la apertura circular convencional. Al insertar la apertura codificada en el sistema óptico se consigue codificar el desenfoque de modo que sea óptimo preservando sus componentes frecuenciales.

Otra aplicación de las aperturas codificadas es la recuperación de profundidad de una escena. Uno de los problemas inherentes en la fotografía tradicional es la limitada profundidad de campo, que causa la aparición de zonas borrosas en partes de la escena que se encuentran alejadas del plano focal. El desenfoque de estas zonas puede emplearse para estimar la profundidad a la que se encuentran. De aquí surge la idea de DFD (*depth from defocus*) una de las aproximaciones para recuperar información 3D de la escena. Para una imagen capturada, puntos de la escena que se sitúen sobre el plano focal aparecerán enfocados, mientras que puntos de la escena alejados del plano focal aparecerán borrosos. Por lo tanto, mediante la estimación del tamaño de blur de un punto de la escena, se puede estimar su profundidad.

Existen varias técnicas empleadas para la estimación de profundidad; algunas consisten en capturar múltiples imágenes de una escena usando distintos parámetros de enfoque [14, 6, 8]. Mientras, por otro lado están los métodos que emplean máscaras ópticas para mejorar la discriminación de profundidad [9, 3].

A continuación se introducen los conceptos de desenfoque y profundidad de campo, y más concretamente la relación entre el grado de desenfoque y distancia. Además, se introducen conceptos teóricos de cómo se codifica el desenfoque y su análisis en el dominio frecuencial. Finalmente se presentan las soluciones propuestas.

## 2.2. Desenfoque y profundidad de campo

Para un objeto plano situado a una distancia  $D_k$  el proceso de captura de una imagen puede ser modelado como una convolución:

$$y = f_k * x + \eta \quad (2.1)$$

siendo  $y$  la imagen observada,  $x$  la imagen nítida y  $f_k$  el filtro de *blur*, *PSF* o *kernel*<sup>1</sup>, que es una versión escalada de la forma de la apertura convolucionada con el patrón de difracción. El término  $\eta$  hace referencia al ruido introducido en el proceso de captura de la imagen, que se modelará como ruido gaussiano de media cero y varianza  $\sigma^2$ ,  $\eta \sim N(0, \sigma^2)$ .

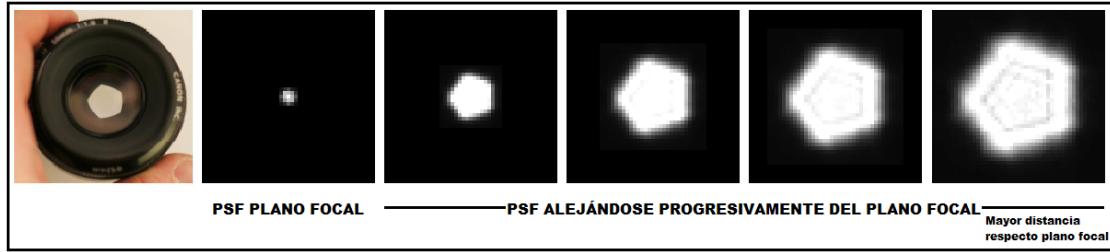


Figura 2.1: PSF a distintas profundidades para una lente Canon 50mm f1.8 II con apertura convencional parcialmente cerrada. (Adaptada de Levin *et al.*[10])

En la Figura 2.1 se muestra el objetivo Canon 50mm f1.8 II empleado a lo largo de todo el proyecto, con la apertura convencional parcialmente cerrada. Las imágenes a su derecha muestran la respuesta de esta apertura a una fuente de luz puntual a distintas profundidades. Esto es lo que se conoce como PSF (*Point Spread Function*) o filtro de *blur*  $f_k$  donde  $k$  representa la variación de la PSF según la profundidad. Por lo tanto, se debe tener en cuenta que la PSF varía conforme nos vamos alejando del plano focal, es decir varía en función de la profundidad y más específicamente con el grado de desenfoque. Esto implica que la cantidad de blur introducido será mayor cuánto mayor sea la distancia respecto al plano focal. Esta es la razón por la cual la captura de una imagen que contiene información a distintas profundidades no puede expresarse como la convolución de la imagen nítida con un único filtro ( $f_k$  en la Ecuación 2.1, variará en función de la profundidad  $k$  a la que se encuentre cada zona de la escena). De este modo la recuperación de la imagen nítida en el proceso de deconvolución, como veremos más adelante, deberá tener en cuenta que para cada zona de la imagen existe una única escala correcta de la PSF según su profundidad. Esta variación, además, es clave para la recuperación de profundidad por DFD, como se verá más adelante.

A continuación se introduce el concepto de desenfoque para comprender en primer lugar por qué sucede y en segundo lugar cómo puede ser controlado y usado para la estimación de profundidad y para el problema de *deblurring*. Para ello se representa el proceso de captura de imágenes en una cámara fotográfica considerando el modelo óptico simplificado de lente convergente delgada, ver Figura 2.2, donde se muestra cómo los rayos reflejados por el objeto alcanzan el sensor de la cámara.

En el proceso de captura de imágenes, mostrado en la Figura 2.2 cuando un objeto está situado sobre el plano focal de la lente todos los rayos de un punto de la escena

<sup>1</sup>Denominaremos de aquí en adelante a  $f_k$  como filtro, kernel o PSF (*Point Spread Function*) a distintas profundidades  $k$ , ya que indica la respuesta de la apertura a un objeto o fuente de luz puntual situado a distancia  $k$  respecto del plano focal.

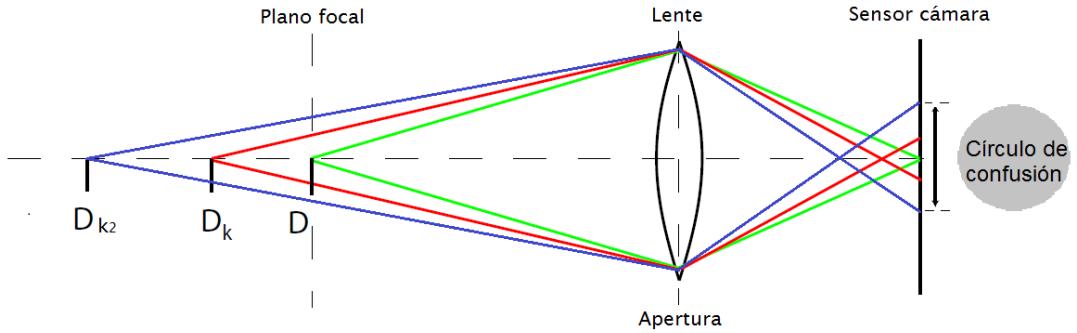


Figura 2.2: Modelo 2D de captura de imagen de una cámara con lente convergente delgada

convergen en un único punto del sensor, dando lugar a una imagen nítida (rayo verde). En cambio si sepáramos el objeto del plano focal y lo situamos a una distancia  $D_k$ , los rayos no convergerán en un único punto del sensor, sino en múltiples puntos (rayo rojo) dando lugar a una imagen desenfocada. Como vemos en la figura la imagen de un punto es, en estos casos, un círculo (el denominado círculo de confusión) que determinará el grado de desenfoque o *blur* de la imagen capturada. Cuanto mayor sea la distancia del objeto respecto del plano focal, mayor será el radio del círculo de confusión, y por lo tanto mayor el grado de desenfoque (rayos azules).

De este modo se establece una relación directa entre grado de desenfoque y distancia, al estar objetos a mayores distancias del plano focal caracterizados por un mayor grado de *blur*. De aquí, como se ha comentado, surge la idea de estimar profundidad a partir del desenfoque de una imagen, DFD (del inglés *depth from defocus*). Es importante resaltar que el patrón de desenfoque en el círculo de confusión vendrá también determinado por la forma y diámetro de la apertura. Por último, se define la profundidad de campo como el rango de profundidades en que no somos capaces de distinguir la existencia de *blur*, es decir el rango de profundidades para el que vemos la escena enfocada. Esta profundidad de campo varía con el diámetro de la apertura, de modo que aperturas más cerradas nos proporcionan una profundidad de campo mayor que aperturas más abiertas.

### 2.3. Codificando el desenfoque: análisis en frecuencia

El principal problema que surge para realizar la estimación de profundidad y la recuperación de una imagen nítida a la vez es que la respuesta en frecuencia de la apertura óptima para cada uno de los dos objetivos del proyecto es diferente.

El proceso de captura de una imagen (ver Ecuación 2.1) se puede expresar en el dominio de la frecuencia como:

$$Y(v, w) = F(v, w)X(v, w) + \zeta \quad (2.2)$$

siendo  $Y(v, w)$ ,  $X(v, w)$ ,  $F(v, w)$  y  $\zeta$  las transformadas de Fourier respectivas de la imagen observada  $y$ , la imagen original nítida  $x$ , el filtro  $f$  y el ruido  $\eta$ . Por lo tanto el problema de *deblurring* o recuperación de la imagen enfocada consiste en estimar la imagen original  $x$  dada  $y$ , obteniendo en el dominio frecuencial:

$$X(v, w) = \frac{Y(v, w)}{F(v, w)} - \frac{\zeta}{F(v, w)}, \quad (2.3)$$

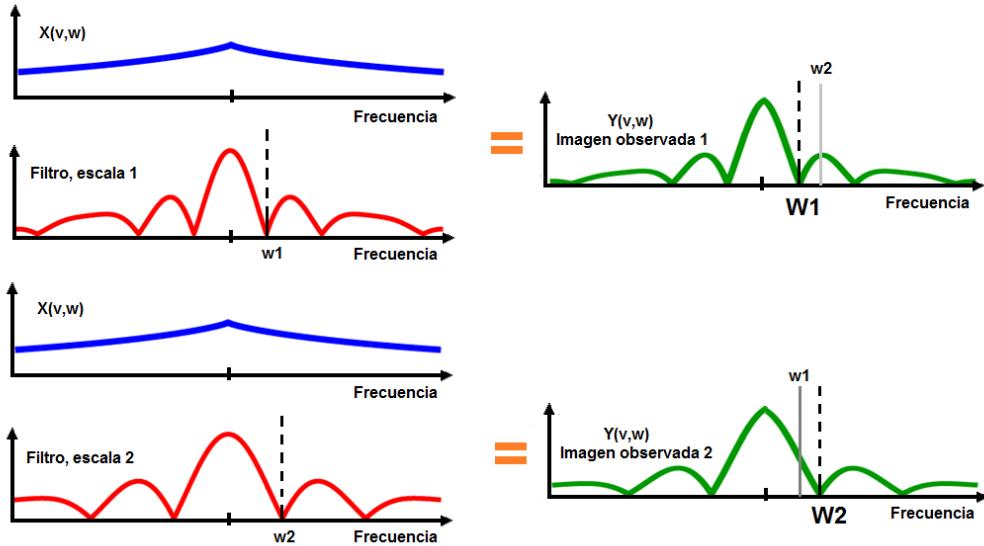


Figura 2.3: Corte 1D de las transformadas de Fourier de las imágenes capturas  $Y(v,w)$  para diferentes escalas del filtro. Las imágenes observadas mantienen cierta estructura de los filtros, preservando los pasos por cero del filtro correspondiente.

donde se puede ver que valores nulos en la respuesta del filtro  $F(v, w)$  hacen imposible recuperar la información a la vez que incrementan la contribución del ruido. El espectro frecuencial ideal para *deblurring* no posee pasos por cero, permitiendo que el filtro sea invertible, y su respuesta es de banda ancha y plana en frecuencia, evitando la pérdida de información en ciertas componentes frecuenciales.

En cambio cuando se precisa estimar profundidades, el espectro frecuencial ideal debe contener pasos por cero. Para comprender esto se ilustra en la Figura 2.3 la transformada de Fourier de un filtro en una dimensión para dos escalas distintas (i.e. dos profundidades), junto con la transformada de Fourier de la imagen original nítida  $X(v, w)$ . El proceso de captura de la imagen  $x$  modelado como una convolución, en el dominio frecuencial equivale a multiplicar ambas señales,  $X(v, w)$  y el filtro en la escala correspondiente. De este modo se obtienen las imágenes observadas  $Y(v, w)$  para cada una de las escalas. Las imágenes observadas  $Y(v, w)$  mantienen cierta estructura de los filtros y en particular se debe resaltar que muestran pasos por cero a las mismas frecuencias que su filtro correspondiente, la primera imagen observada mantiene el paso por cero en la componente frecuencial  $w_1$  mientras que la componente frecuencial nula para la segunda imagen observada es  $w_2$ .

Estimar la profundidad de una escena implica resolver el problema inverso. Se parte únicamente de una imagen capturada  $Y(v, w)$  y queremos saber qué escala del filtro se corresponde con esa imagen capturada, sabiendo que, si se conoce la escala del filtro, se conocerá la profundidad a la que se encuentra. Para ello se aplican dos filtros a escalas distintas sobre la imagen observada  $Y(v, w)$  para recuperar la imagen nítida  $X(v, w)$ , resultando la expresión de la Ecuación 2.3 pero sin considerar el ruido:

$$X(v, w) = Y(v, w)/F(v, w)$$

De este modo, si se trata del filtro correcto (ver Figura 2.4), sus pasos por cero ( $w_1$ ) coinciden con los pasos por cero de la imagen capturada. Resulta entonces el espectro de imagen  $X(v, w)$  mostrado a la izquierda de la figura, que se corresponde con un espectro de imágenes naturales (mayor contenido frecuencial en bajas frecuencias, disminuyendo

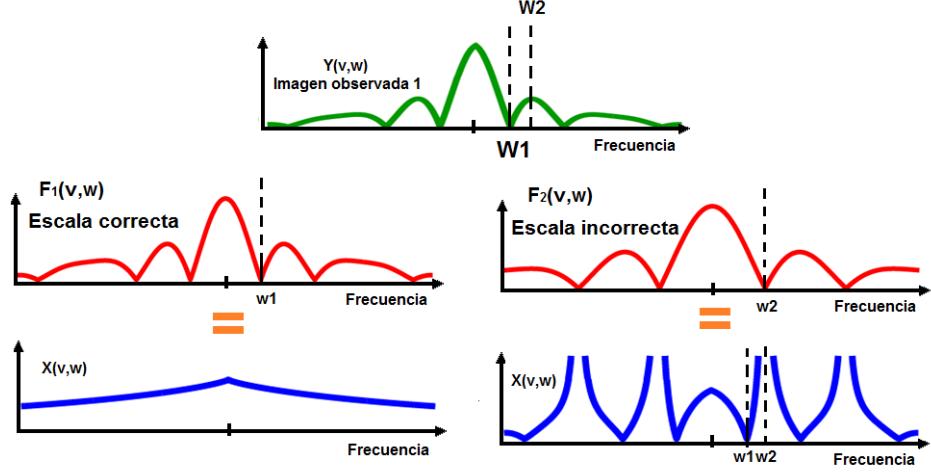


Figura 2.4: Identificación de la escala correcta a partir de la imagen observada. Espectro de la imagen en la izquierda obtenida con la escala correcta del filtro, espectro en la derecha obtenido con una escala incorrecta del filtro.

el contenido frecuencial para componentes de altas frecuencias). En cambio el filtro de escala incorrecta mostrado en la parte derecha de la figura, presenta pasos por cero en la componente frecuencial  $w_2$ , mientras que la imagen observada muestra valores no nulos para esa frecuencia. Esto supone que al realizar la división de  $Y(v, w)$  entre el filtro, el espectro resultante  $X(v, w)$  tiende a infinito para  $w_2$ , dando lugar a un espectro que no se corresponde con el esperado para una imagen natural, deduciéndose que la escala del filtro no es la correcta. Así, a partir de los pasos por cero de un filtro se puede identificar la escala correcta y con ello estimar la profundidad.

Existe pues, como hemos visto, un compromiso en la elección del filtro para conseguir tanto estimar profundidad como para recuperar imágenes enfocadas. En el primer caso se necesita que la respuesta del filtro presente pasos por cero, mientras que para *deblurring* se requiere que el filtro sea invertible (no presente valores nulos) a la vez que su respuesta sea de banda ancha y plana, preservando el contenido de altas frecuencias.

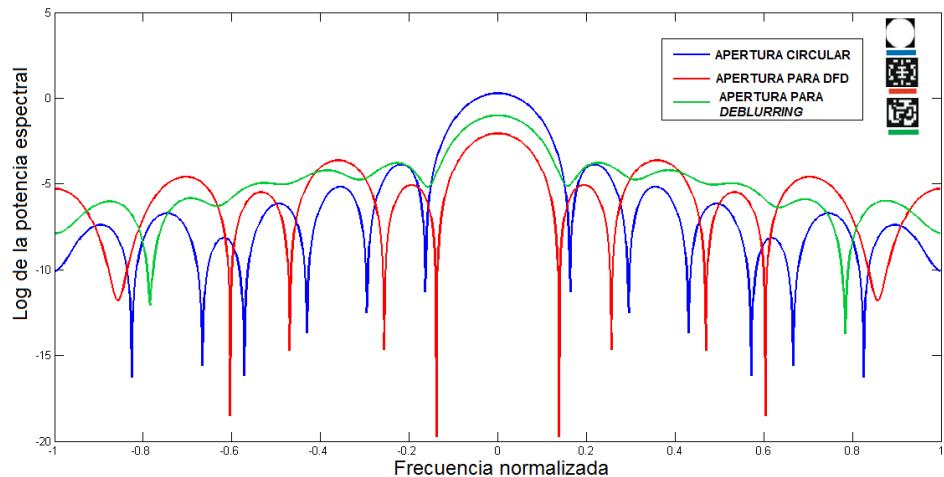


Figura 2.5: Logaritmo de la potencia espectral para la apertura circular, apertura optimizada para DFD [10] y apertura optimizada para *deblurring* [24].

Este compromiso en el diseño de las aperturas puede verse en la Figura 2.5. La potencia espectral de la apertura circular muestra varios paso por cero junto a una atenuación clara de las componentes frecuenciales altas, no resultando óptima para *deblurring*. La apertura diseñada para DFD presenta varios pasos por cero e intenta a su vez que su respuesta sea más plana, sin atenuar tanto las componentes de alta frecuencia. Por último la apertura diseñada para *deblurring* evita los pasos por cero y presenta una respuesta en frecuencia plana y de banda ancha.

## 2.4. Soluciones propuestas

Como se ha visto, el objetivo del presente trabajo es doble: recuperar a partir de fotografías tanto la escena enfocada como un mapa de profundidades. Para ello se decide en primer lugar plantear el problema de *depth from defocus* a partir de una única imagen capturada y utilizando una única apertura, partiendo del trabajo previo de Levin *et al.*[10]. La obtención de un mapa de profundidades a partir de una única imagen resulta un problema complicado ya que únicamente se dispone de la información de una imagen capturada. Se expondrán las limitaciones de este método, su fuerte dependencia con el sistema de captura y con la imagen capturada. Por ello se decide explorar una segunda opción que consiste en un método que emplea pares de aperturas codificadas [23]. En este segundo método se parte de dos imágenes capturadas con una pareja de aperturas codificadas y se usa la información del desenfoque relativo entre ambas capturas para recuperar la profundidad de la escena. En lo que a *deblurring* se refiere, la primera apertura está optimizada para profundidad, por lo que se emplea un método de deconvolución más favorable que contrarreste la respuesta poco apropiada de la apertura para ese objetivo. En el método de pares de aperturas codificadas su respuesta conjunta presenta un comportamiento mejor para *deblurring*, por ello se hace uso de una deconvolución de Wiener generalizada.

# Capítulo 3

## Apertura codificada óptima

### 3.1. Introducción

La primera de las opciones a explorar se corresponde con el diseño de una única apertura codificada para DFD, basado en el trabajo previo de Levin *et al.*[10]. La idea de este trabajo previo consiste en la inserción de una apertura codificada en la lente de la cámara, modificando el desenfoque producido por la lente. La codificación del desenfoque permitirá la extracción de información de profundidad y la recuperación de una imagen enfocada a partir de una única imagen capturada. Dado que dicha apertura estará optimizada para DFD se hará uso de un nuevo método de deconvolución para *deblurring*.

En primer lugar se realiza una búsqueda del patrón de apertura óptimo para DFD, para lo que se aplica un criterio de discriminación de profundidad (Sección 3.2). Una vez escogido el patrón de apertura óptimo se pasa a estudiar en primer lugar el problema de *deblurring* o deconvolución (Sección 3.3), y en segundo lugar la estimación de profundidad (Sección 3.4). Cabe destacar que para el proceso de deconvolución se hará uso de un modelo estadístico de imágenes y se introducirá el concepto de deconvolución sparse, realizando una comparativa entre los distintos métodos de deconvolución existentes. En concreto se analizarán las mejoras aportadas por el uso de un prior de derivadas sparse sobre el uso de prior de derivadas gaussianas.

El diagrama de flujo a seguir para la obtención del mapa de profundidades y la recuperación de la imagen enfocada se muestra en la Figura 3.1.

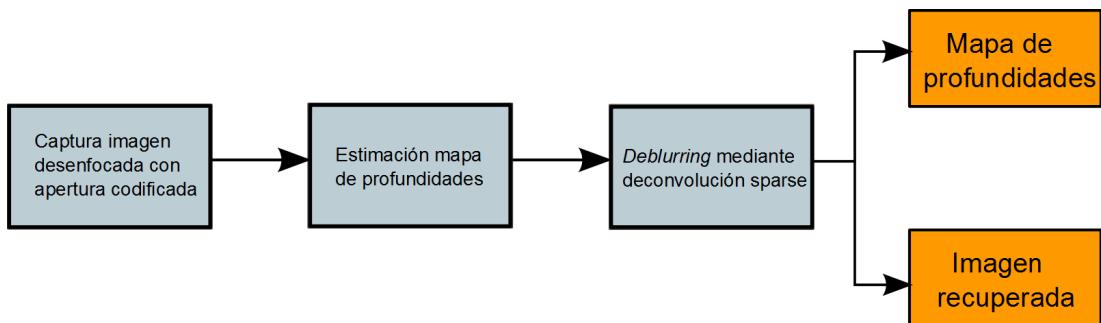


Figura 3.1: Diagrama de flujo a seguir para la obtención del mapa de profundidades y la imagen recuperada a partir de la imagen capturada con la apertura codificada.

### 3.2. Diseño y obtención de la apertura codificada

En el diseño de una apertura codificada existen dos principios básicos que el filtro debe cumplir:

1. El filtro debe ser fácilmente invertible de modo que la recuperación de la imagen nítida sea posible.
2. El filtro debe distinguir con claridad el desenfoque resultante de cada escala del filtro.

Teniendo en cuenta el amplio número de filtros posibles, la elección de la apertura óptima bajo estas dos condiciones no es un problema trivial. Nuestro objetivo de discriminar profundidades en una imagen hace que sea interesante resaltar el comportamiento de los filtros a distintas escalas. Como veremos ese será el punto de partida esencial.

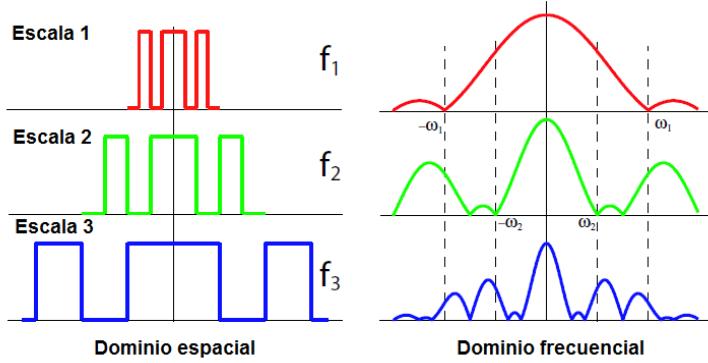


Figura 3.2: Ejemplo de filtro en 1D representando la relación del escalado en el dominio espacial del filtro con sus pasos por cero en frecuencia. (Adaptado de Levin *et al.*[10])

En la Figura 3.2 se muestra un filtro a tres escalas diferentes con sus respectivas transformadas de Fourier a la derecha. La idea principal es considerar la relación existente entre la escala del filtro en el dominio espacial y sus pasos por cero en el dominio frecuencial [16]. Por ejemplo, el filtro  $f_1$  muestra en su transformada de Fourier un paso por cero a  $w_1$ . Recordando las ecuaciones que rigen el procedimiento de captura en el dominio espacial:

$$y = f_k * x + \eta$$

y en el dominio frecuencial:

$$Y(v, w) = F(v, w)X(v, w) + \zeta$$

esto significa que si la imagen capturada  $y$  es resultado de la convolución con el filtro  $f_1$ , la transformada de Fourier de la imagen es nula para la frecuencia  $w_1$ , obteniendo  $Y(w_1) = 0$ . Por lo tanto, como se ha explicado en la Sección 2.3, a partir de las componentes frecuenciales nulas de la imagen observada podemos conocer la escala del filtro correspondiente y de ahí hacer una estimación de la profundidad.

Este argumento puede hacerse también en el dominio espacial. Si  $Y(w_1) = 0$  ciertas condiciones lineales deben satisfacerse lo que implica que la imagen  $y$  no puede ser un vector  $N$  dimensional arbitrario (siendo  $N$  el número de píxeles de la imagen). Tal y como

podemos observar en la Figura 3.2 la variación del escalado (frecuencia) en el dominio espacial se traduce en distintas componentes frecuenciales nulas en su transformada de Fourier. Cada escala del filtro define un subespacio lineal diferente de posibles imágenes desenfocadas. Por lo tanto, dada una imagen de entrada  $N$  dimensional, identificar la escala del filtro a la que fue blurreada (y a su vez la profundidad del objeto) se reduce en identificar el subespacio al que pertenece.

En teoría identificar las componentes frecuenciales nulas o equivalentemente encontrar el subespacio correcto es una tarea directa y simple. Sin embargo, en la práctica el problema se vuelve más complejo. En primer lugar el ruido presente en la toma de imágenes hace que ninguna componente frecuencial sea completamente nula (y por lo tanto que la imagen  $y$  no pertenezca a ningún subespacio en concreto). En segundo lugar, componentes frecuenciales nulas en la imagen observada  $y$  pueden deberse a que el contenido de la imagen original  $x$  sea nulo a esas frecuencias. Todo esto hace que el comportamiento de unos filtros de apertura sean más apropiados que otros para nuestro propósito.

Es importante recordar que siempre va a existir un compromiso en la elección de la respuesta en frecuencia entre el proceso de *deblurring* debido al desenfoque y el proceso de estimación de la profundidad. Para el problema de *deblurring* nos interesa que la respuesta del filtro de apertura presente el menor número posible de pasos por cero. Una respuesta con pocos pasos por cero permite que el filtro sea fácilmente invertible y facilita el proceso de deconvolución. En cambio si nuestro objetivo es la estimación de profundidad, tal y como hemos mostrado en la Figura 3.2, son los pasos por cero de la respuesta en frecuencia los que nos permiten discriminar profundidades. La presencia de éstos en la respuesta en frecuencia dificulta el proceso de invertir el filtro, ya que al perder información el proceso de deconvolución se vuelve muy sensible al ruido para esas frecuencias.

Como consecuencia de este compromiso se deberá sacrificar parte del contenido de la imagen para poder obtener información de profundidad de la escena. Afortunadamente, si sólo se sacrifica un pequeño número de frecuencias, el uso de priors de imágenes pueden reducir la sensibilidad al ruido, pudiendo hacer uso en la deconvolución de kernels de tamaño moderado ( $\leq 19$  píxeles).

Para el diseño del filtro de la apertura se hace uso de un modelo estadístico de imágenes reales, ver Apéndice B. Se estudian imágenes desenfocadas por un kernel específico, que conducen a la obtención de un criterio para evaluar la selectividad de cada filtro. Este criterio se aplica en una búsqueda aleatoria entre un amplio banco de filtros posibles para la elección de uno de ellos.

### 3.2.1. Criterio de selección del filtro

A partir del modelo estadístico de imágenes, que puede consultarse en el Apéndice B, se deriva el siguiente criterio de selección de filtro. Partiendo del proceso de captura de imagen modelado como una convolución:

$$y = f_k * x$$

El objetivo es conocer la verdadera escala  $k$  del filtro y con ello estimar la profundidad de esa parte de la escena y recuperar su enfoque. Resulta intuitivo que si las distribuciones

de la imagen desenfocada  $P_{k1}(y)$  y  $P_{k2}(y)$  a distintas profundidades  $k1, k2$  son similares será difícil distinguir entre ambas. Un método clásico de medida de distancias entre distribuciones es la divergencia de Kullback-Leibler (divergencia KL):

$$D_{KL}(P_{k1}(y), P_{k2}(y)) = \int_y P_{k1}(y)(\log P_{k1}(y) - \log P_{k2}(y)) dy \quad (3.1)$$

Un filtro de apertura que maximice esta distancia presentará una probabilidad  $P_{k1}(y)$  alta para una imagen desenfocada  $y$  de profundidad  $k1$ , mientras que la probabilidad  $P_{k2}(y)$  será baja (ya que la profundidad  $k2$  no se corresponde con la profundidad real).

Incluyendo la representación del modelo (Ecuaciones B.4, B.5) en el dominio de la frecuencia en la Ecuación 3.1, la divergencia KL puede expresarse como:

$$D_{KL}(P_{k1}, P_{k2}) = \sum_{v,w} \left( \frac{\sigma_{k1}(v, w)}{\sigma_{k2}(v, w)} - \log \left( \frac{\sigma_{k1}(v, w)}{\sigma_{k2}(v, w)} \right) \right) \quad (3.2)$$

donde  $\sigma_k(v, w)$  para cada profundidad  $k$  viene dado por:

$$\sigma_k(v, w) = |F_k(v, w)|^2 (\alpha|G_x(v, w)|^2 + \alpha|G_y(v, w)|^2)^{-1} + \eta^2 \quad (3.3)$$

donde  $G_x(v, w)$  y  $G_y(v, w)$  son las transformadas de Fourier de los filtros derivativos  $g_x = [1 - 1]$  y  $g_y = [1 - 1]^T$ , y  $F_k(v, w)$  la transformada de Fourier del filtro a la escala  $k$ . La Ecuación 3.2 muestra que la distancia entre distribuciones de dos profundidades distintas será mayor cuánto mayor sea el ratio de sus componentes frecuenciales. El ratio se maximiza para frecuencias  $v, w$  para las cuales  $F_{k2}(v, w)$  o  $F_{k1}(v, w)$  toma valores grandes. De modo que tal y como se ha nombrado anteriormente, son los pasos por cero del filtro los que permiten discriminar entre distintas escalas del mismo. Es importante que una escala posea pasos por cero a ciertas frecuencias  $v, w$ ; mientras que otra escala preserve contenido de la señal a esa misma frecuencia. Además, el hecho de que  $\sigma_k(v, w)$  pese el contenido frecuencial del filtro por el prior imagen, ver Ecuación B.5, indica que los pasos por cero son más discriminantes en los contenidos de baja frecuencia, donde el contenido de la imagen original es más significativo.

Para la elección de la apertura óptima para DFD se aplica el criterio de la Ecuación 3.2 a un amplio número de aperturas posibles [10], resultando elegida la apertura codificada mostrada en la parte derecha de la Figura 3.3 junto con un corte 1D de su transformada de Fourier. Para comprender la ventaja que nos aporta el uso de aperturas codificadas, se muestra también un corte 1D de la transformada de Fourier de la apertura circular convencional.

Mientras que las respuestas en frecuencia de la apertura codificada a distintas escalas presentan sus pasos por cero a diferentes frecuencias, en la apertura convencional los pasos por cero de las distintas escalas se confunden entre sí, teniendo lugar en las mismas frecuencias. Por lo tanto el uso de una apertura codificada nos va a permitir distinguir entre los nulos de distintas escalas, tal y como buscábamos, permitiendo conocer la escala correcta del filtro y por lo tanto estimar la profundidad de la escena. Además la información que se pierde para esa escala del filtro se conserva en el resto de escalas. Es importante resaltar que la respuesta en frecuencia de la apertura codificada tiene mayor ancho de banda presentando un comportamiento mejor a altas frecuencias que en el caso de la convencional.

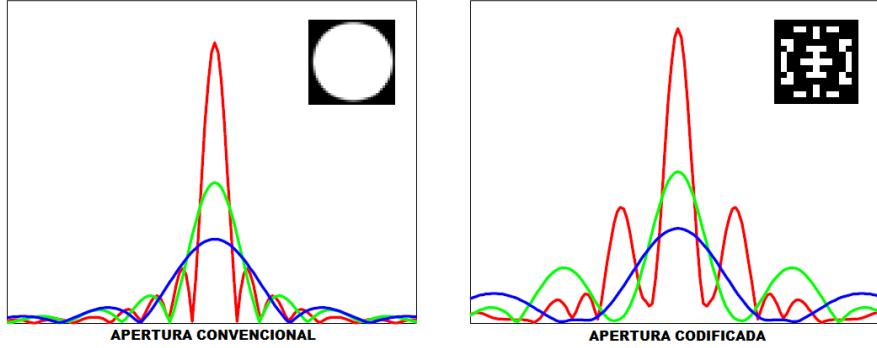


Figura 3.3: Corte transformadas de Fourier en 1D en tres escalas distintas de la apertura convencional (izquierda) y la apertura codificada escogida como óptima para DFD (derecha). La figura muestra cómo en el caso de la apertura convencional los pasos por cero a distintas escalas se confunden entre sí, mientras que para la apertura codificada tienen lugar a distintas frecuencias. (Adaptado de Levin[10] *et al.*)

### 3.3. Deconvolución sparse

Una vez identificada la escala correcta de blur en la imagen observada  $y$ <sup>1</sup>, se procede a eliminar el *blur* o desenfoque, reconstruyendo la imagen nítida de la escena a partir de la imagen desenfocada obtenida. Este proceso se denominará en adelante *deconvolución* o *deblurring*. En este apartado nos centraremos en explicar la deconvolución usando prior gaussiano y prior sparse. Una versión completa y detallada de los métodos de deconvolución que hacen uso de priors de imágenes naturales puede consultarse en el Apéndice A.

Basándose en el modelo probabilístico se obtiene:

$$P_k(x|y) \propto \exp\left(-\left(\frac{1}{\eta^2}|C_{f_k}x - y|^2 + \alpha|C_{g_x}x|^2 + \alpha|C_{g_y}x|^2\right)\right) \quad (3.4)$$

siendo  $C_{f_k}, C_{g_x}, C_{g_y}$  matrices de convolución  $N \times N$  del filtro de blur y de los filtros derivativos  $g_x = [1 - 1]$  y  $g_y = [1 - 1]^T$ , respectivamente. El problema de *deblurring* puede expresarse como, dada  $y$ , encontrar la máxima probabilidad de modo que  $x^* = \text{argmax} P_k(x|y)$ . Para poder resolver esto, es necesario introducir algún tipo de prior sobre la imagen  $x$ . En primer lugar se plantea la deconvolución usando un prior gaussiano. En este caso, el problema se reduce a uno de optimización de mínimos cuadrados

$$x^* = \arg \min \frac{1}{\eta^2}|C_{f_k}x - y|^2 + \alpha|C_{g_x}x|^2 + |C_{g_y}x|^2 \quad (3.5)$$

Minimizando la Ecuación 3.5 se busca por tanto la imagen  $x$  que minimice el error de reconstrucción  $|C_{f_k}x - y|^2$ . La solución óptima de la Ecuación 3.5 puede encontrarse resolviendo un conjunto de ecuaciones lineales sparse:  $Ax = b$  con

$$A = \frac{1}{\eta^2}C_{f_k}^T C_{f_k} + \alpha C_{g_x}^T C_{g_x} + \alpha C_{g_y}^T C_{g_y} \quad b = \frac{1}{\eta^2}C_{f_k}^T y \quad (3.6)$$

La Ecuación 3.6 se resuelve en el dominio frecuencial en unos pocos segundos para una imagen del orden de 4 megápixeles. Al considerar que la convolución es un operador cíclico aparecen artefactos en los bordes de la imagen, aunque no suele ser relevante en imágenes

---

<sup>1</sup>La identificación de la escala correcta del filtro se explica en la Sección 3.4

de gran tamaño. El problema de *deblurring* usando prior de derivadas gaussiano es simple y eficiente, pero su mayor defecto es que tiende a suavizar demasiado el resultado.

Para conseguir imágenes recuperadas con más contrastes se requiere el uso de un prior de imagen natural más fuerte, por ejemplo el uso de prior de derivadas sparse.

Para la obtención de la imagen  $x$  en el caso de uso de prior sparse se debe minimizar:

$$|C_{f_k}x - y| + \sum_{i,j} \rho(x(i,j) - x(i+1,j)) + \rho(x(i,j) - x(i,j+1)) \quad (3.7)$$

siendo  $\rho(z) = |z|^{0.8}$ . Mientras que el prior gaussiano distribuye las derivadas uniformemente sobre la imagen, un prior sparse opta por concentrar las derivadas en un pequeño número de píxeles, manteniendo la mayor parte de los píxeles constantes. El resultado son imágenes con bordes más nítidos, reduciendo el ruido y favoreciendo la eliminación de artefactos no deseados como el *ringing* (ruido alrededor de contornos o zonas marcadas que se asemeja a anillos). El problema del uso de prior de derivada sparse es que la optimización deja de ser un problema de mínimos cuadrados, ya que ya no se trata de un problema convexo. Para su optimización se usa un proceso iterativo de reponderación de mínimos cuadrados (*iterative reweighting least squares*), donde el peso de cada derivada se actualiza basándose en la solución de la iteración previa. En la Sección 3.5.3 se muestran resultados obtenidos al deconvolucionar imágenes capturadas junto un análisis y estudio de los resultados. Una descripción completa de los métodos de deconvolución basados en priors de imágenes puede consultarse en el Apéndice A.

### 3.4. Recuperación de profundidad

A la hora de estimar el mapa de profundidades se parte de una imagen desenfocada, capturada con la apertura codificada, y de las PSFs del sistema a distintas profundidades<sup>2</sup>. El primer paso es deconvolucionar, resultando  $k$  posibles imágenes recuperadas  $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k$ . Para cada profundidad  $k$ , se calcula el error de reconstrucción como:

$$e_k = y - f_k * \hat{x}_k, \quad (3.8)$$

siendo  $y$  la imagen observada y  $\hat{x}_k$  la imagen obtenida de la deconvolución de  $y$  con el filtro  $f_k$ . La imagen recuperada  $\hat{x}_k$  presentará una reconstrucción correcta de las zonas de la imagen donde  $k$  sea la escala correcta del filtro (es decir la PSF a esa profundidad). En cambio la reconstrucción en otras áreas de la imagen, cuya profundidad sea distinta, contendrán artefactos de *ringing* ya que no es la escala correcta de estas áreas. Esto hace que el error de reconstrucción en esas áreas sea elevado, pero más significativo en zonas con textura o bordes que en zonas sin textura. Para intentar resolver el problema de las texturas se calcula el error de reconstrucción en ventanas locales, es decir se calcula una aproximación de la energía  $\hat{E}_k(y(i))$  alrededor del píxel  $i$ , calculando el error de reconstrucción en una ventana local como:

$$\hat{E}_k(y(i)) \approx \sum_{j \in W_i} e_k(j)^2 \quad (3.9)$$

donde  $j \in W_i$  son los píxeles de esa ventana.

De este modo se podría estimar la profundidad para cada píxel  $d(i)$  como:

$$d(i) = \arg \min_k \lambda_k \hat{E}_k(y(i)) \quad (3.10)$$

---

<sup>2</sup>Las PSFs a distintas profundidades se obtienen mediante el método de calibración explicado en la Sección 3.5.1.

Como se puede ver en la Ecuación 3.10 en la práctica el error de reconstrucción en cada ventana local no es mínimo para la escala correcta, de modo que es necesario pesar la energía de cada ventana con un vector de pesos (un valor  $\lambda$  para cada profundidad). En la Sección 3.5.4 se explica cómo se plantea la obtención de estos valores y sus limitaciones.

Finalmente se reconstruye una imagen nítida escogiendo independientemente cada píxel de las  $x_k$  imágenes obtenidas de la deconvolución, de modo que la imagen recuperada puede expresarse como:

$$\hat{x}_{final}(i) = \hat{x}_{d(i)}(i), \quad d(i) = 1, 2, \dots k \quad (3.11)$$

De este modo, al reconstruir la imagen final nítida, cada píxel de la misma se toma de una de las  $\hat{x}_k$  imágenes obtenidas de la deconvolución, en concreto, de aquella que está a la profundidad adecuada dada por  $d(i)$ .

### 3.4.1. Obtención del vector de pesos

Tal y como hemos visto en la Ecuación 3.10, para que el error de reconstrucción sea mínimo en la profundidad correcta es necesario multiplicarlo por un vector de pesos  $\lambda = \{\lambda_i\} i = 1 \dots k$ . La obtención de este vector de pesos  $\lambda_i$  se debe realizar mediante un algoritmo de entrenamiento y aprendizaje partiendo de ventanas locales de profundidad conocida. Para disponer de un alto número (decenas de miles) de ventanas de profundidad conocida se capturaron imágenes de escenas con distintas texturas, fijando el enfoque en la distancia 0m. Las escenas capturadas tienen una única profundidad que es conocida. Y se captura una imagen para cada profundidad del rango determinado, es decir para cada profundidad  $k$ . Cada escena se descompone en  $N$  ventanas de tamaño  $n \times n$  píxeles. Y de este modo para cada ventana y para cada profundidad  $k$ , se calculan  $k$  errores de reconstrucción (vector  $E_k$ ).

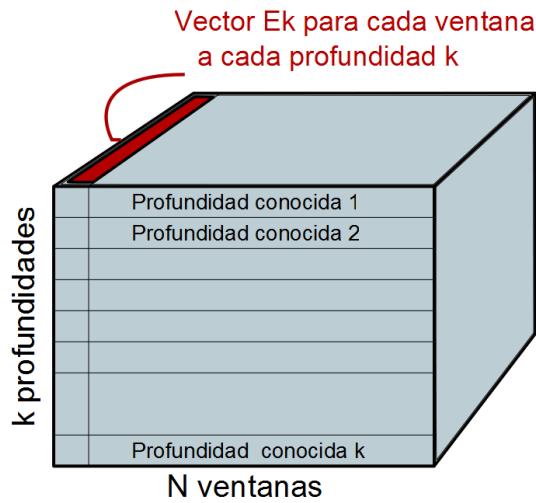


Figura 3.4: Representación gráfica del almacenamiento de los errores de reconstrucción en ventanas de profundidad conocida.

Para la obtención del vector de pesos partiendo de la matriz de errores de reconstrucción para 10.000 ventanas cuya profundidad es conocida se emplean dos métodos distintos.

El primero consistió en optimizar la siguiente función, que es convexa en el parámetro a optimizar (el vector  $\lambda_i$ ), mediante gradiente descendente (empleando la función fminunc

de Matlab):

$$\sum_{N \text{ ventanas}} (\log e^{\lambda_k E_{k_0}} - \log(\sum_k \lambda_k E_k)) \quad (3.12)$$

En segundo lugar se implementó un algoritmo genético <sup>3</sup> que parte de una población inicial S aleatoria de vectores de pesos de valores entre 0 y 1. De este modo, como se conoce la profundidad real de cada ventana, el criterio de evaluación del algoritmo genético es el número de aciertos. Los M vectores de pesos que resultan en un número mayor de aciertos será la población seleccionada y a partir de la cual se incrementará otra vez la población hasta S individuos mediante mutación y recombinación.

Los resultados de ambas propuestas como algoritmo de entrenamiento demuestran que el vector de pesos es muy dependiente tanto del sistema (parámetros de influencia en la captura) como de la imagen, siendo necesario calcular un vector de pesos para cada sistema y cada imagen capturada. Además se comprueba que las zonas sin textura de las imágenes no contienen información suficiente para que el error de reconstrucción varíe para distintas escalas del filtro (teóricamente error mayor para escalas incorrectas error menor para la escala correcta). Por todo esto se considera que el procesado de decenas de miles de ventanas de profundidad conocida resulta en un problema de complejidad elevada para que finalmente sea aplicable sólo para un sistema e imagen en concreto, y por ello se plantea explorar, en el Capítulo 4, un método que derive en mejores resultados y cuyo coste sea menor.

## 3.5. Resultados

Esta sección detalla en primer lugar el proceso de calibración, montaje del sistema y captura de imágenes llevado a cabo para la validación sobre soporte físico de la apertura. Después se muestra la recuperación de objetos nítidos en una escena de varias profundidades. Además se realiza una comparación entre los distintos métodos de deconvolución explicados en el Apéndice A y otros algoritmos tradicionales de deconvolución. Por último se muestran resultados y valoraciones de la obtención del mapa de profundidades.

### 3.5.1. Calibración

La calibración consiste principalmente en la obtención de las PSFs del sistema óptico para la apertura codificada, puesto que, como se ha explicado, esta PSF varía con la profundidad, el grado de desenfoque, y la forma y tamaño de la apertura. Para ello se imprime en material fotolitográfico y en alta resolución el patrón de la apertura óptima mostrada en la Figura 3.5. El patrón debe ser circular y de tamaño adecuado de modo que cubra completamente la lente, en este caso una Canon 50 mm f1.8 II. La inserción de la apertura codificada sobre la lente convencional del objetivo de la cámara junto con los pasos a seguir para desmontar el objetivo pueden consultarse en el Apéndice D. A lo largo de todo el proyecto se usa una cámara reflex Canon EOS 500D junto con el objetivo Canon 50 mm f1.8 II. La Figura 3.5 muestra dicha cámara y la apertura codificada insertada en el objetivo.

La parte más importante de la calibración del sistema es la obtención de la PSF del sistema óptico completo (cámara y apertura codificada) a distintas profundidades. Como se ha nombrado anteriormente, la PSF de un sistema óptico es su respuesta a un objeto o luz puntual. Por lo tanto para obtener las distintas PSF se hace uso de un array de diodos LED fabricado sobre un cajón negro donde se realizan pequeños agujeros del mismo tamaño que permitan obtener fuentes de luz puntuales. El array de LEDs (ver

---

<sup>3</sup>La estructura del algoritmo genético es similar a la utilizada para la obtención de pares de aperturas en la Sección 4.3, por ello no se considera necesario ni relevante explicarlo con detalle en esta sección.



Cuerpo Canon EOS 500D



Apertura codificada

Figura 3.5: Cuerpo de cámara reflex empleada y apertura codificada colocada sobre la lente del objetivo.

Figura 3.7) consta de diez puntos distribuidos espacialmente sobre el plano imagen. En realidad, y aunque no se ha comentado hasta ahora, la PSF varía también a lo largo del plano imagen debido a factores como la distorsión radial de la lente, de modo que las PSFs de cada extremo para una misma profundidad difieren ligeramente. En este proyecto no se tendrá en cuenta esta variación y se escogerá la PSF central superior como kernel de deconvolución para todo el plano imagen.



PSFs en la distancia en foco (puntos)



PSFs a distancia de desenfoque de 105 cm

Figura 3.6: Array de diodos LED empleado para la calibración de las PSFs

De este modo, sin modificar los parámetros establecidos y manteniendo la posición y el enfoque de la cámara fijo, se capturan distintas imágenes del array de LEDs a las profundidades de interés. Posteriormente, sobre estas imágenes capturadas se recortan las PSFs para cada profundidad, un ejemplo puede verse en la Figura 3.7. Estas PSFs deben ser procesadas antes de ser usadas en la deconvolución. En primer lugar se realiza una correcta umbralización de sus valores que corrija la difracción y el *glow* o destello del propio LED. Posteriormente se normaliza su energía a uno, esto es muy importante para no modificar la energía de la imagen capturada ya que valores de energía distintos a la unidad oscurecerían o aclararían la imagen. Como se verá en el apartado siguiente en ocasiones es necesario realizar un escalado de las PSF, este paso se realizará entre la umbralización y la normalización a energía unidad.

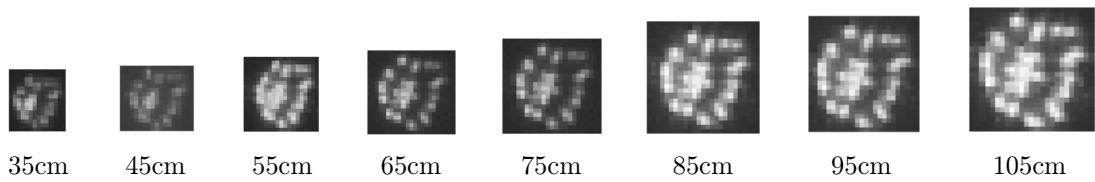


Figura 3.7: PSFs obtenidas sin umbralizar para distintas profundidades (apertura f 2.2, d=2m).

### 3.5.2. Captura de imágenes

Una vez calibrado el sistema se procede a la captura de imágenes con la apertura codificada. Para ello se crea un escenario con objetos a distintas distancias respecto del plano focal. Se escoge colocar objetos a nueve profundidades distintas, comenzando sobre el plano focal (distancia cero, objetos enfocados) y considerando una profundidad máxima de 105 cm respecto del plano focal. La separación entre las nueve profundidades es de 10 cm, excepto entre la distancia focal y la primera profundidad que es de 35cm. Esto se debe a que desde el plano focal hasta los 35 cm se considera que el grado de blur es muy pequeño (<4 píxeles) por lo que no se puede discriminar esas profundidades debido a la falta de estructura en el blur.

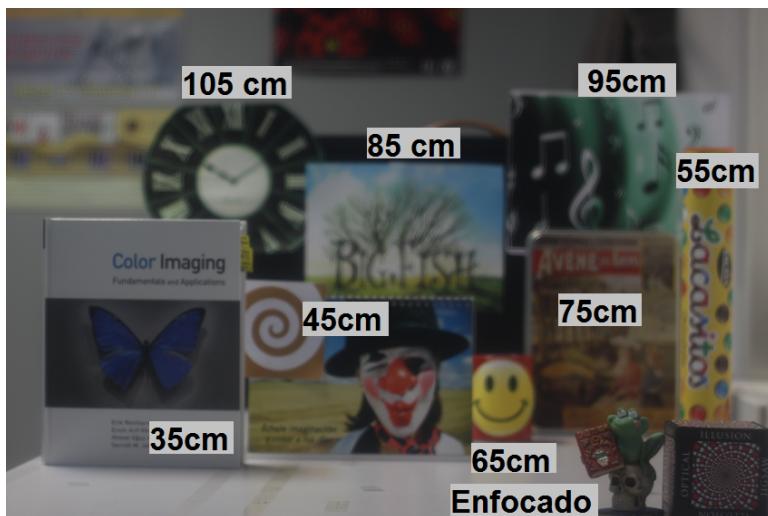


Figura 3.8: Ejemplo escena capturada con objetos distribuidos a lo largo de nueve profundidades.

En la captura de las imágenes es necesario calibrar y fijar los parámetros deseados al inicio y no cambiarlos a lo largo de todo el proceso de captura. Son varios los parámetros que influyen en el resultado de la imagen capturada. Principalmente seleccionaremos los siguientes parámetros:

- Tamaño de imagen: se establece en el menor que nos permite el modelo de la cámara, 2352x1568 píxeles.
- Diámetro de apertura ( $D$ )<sup>4</sup>, realizaremos capturas con distintos valores.
- Tiempo de exposición: variará en función del valor de apertura.

---

<sup>4</sup>Expresado en milímetros se denomina D, sino se expresa como número F.

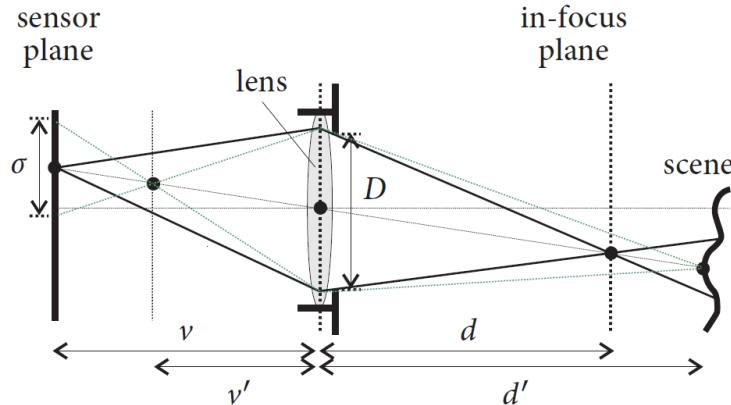


Figura 3.9: Geometría del diámetro de blur para lente delgada. (Adaptado de Hasinoff [7])

- Valor de sensibilidad ISO, se escoge un valor bajo, ISO 100, que permite obtener fotografías de buena calidad evitando la aparición de mucho ruido, ya que a mayor sensibilidad ISO mayor ruido presente en la imagen capturada.
- Distancia de la lente hasta el plano focal ( $d$ ), se realizan capturas a dos distancias de enfoque distintas observando su influencia.

En especial el diámetro de apertura (cómo de abierta o cerrada está) y la distancia de la lente hasta el plano focal influyen en el diámetro del *blur* o grado de desenfoque  $\sigma$ , que para una distancia de desenfoque  $d'$  viene dado por:

$$\sigma = D \frac{f|d' - d|}{(d - F)d'}, \quad (3.13)$$

donde  $f = 50mm$  es la distancia focal de nuestra lente,  $D$  es el diámetro de la apertura,  $d'$  es la distancia del objeto desenfocado a la lente y  $d$  la distancia de la lente al plano focal. Una representación gráfica de esto se muestra en la Figura 3.9.

Para el proceso de captura se enfoca a una cierta distancia  $d$ , que por tanto definirá el plano focal, de modo que los objetos sobre el plano focal aparecerán enfocados. La escena se capture mediante el uso de un disparador automático evitando cualquier tipo de movimiento de la cámara. De este modo se obtiene una imagen de la escena con grados de desenfoque distintos según la profundidad a la que se encuentre cada objeto (ver Figura 3.8).

Se capturó la escena para distintas combinaciones de parámetros de la cámara, que se corresponden, como se ha explicado, con distintos diámetros de *blur*. Las distintas capturas con sus respectivos parámetros se especifican en el Cuadro 3.1. Para cada captura se calcula el diámetro de blur teórico máximo (para la máxima profundidad) según la Ecuación 3.13. Se expresa el diámetro del blur en píxeles teniendo en cuenta que el tamaño del sensor de imagen de nuestra lente es de 22.3mm x 14.9mm y el tamaño de las imágenes capturadas se ha establecido en 2352x1568 píxeles. De este modo el tamaño de nuestro píxel en micras se pudo calcular como 22.3mm/2352 píxeles o 14.9mm/1568 píxeles, resultando un tamaño de 9.48 micras<sup>5</sup>. En el Cuadro 3.1 puede apreciarse que para la misma distancia  $d$  (distancia de la lente al plano focal) valores de apertura mayores (apertura más cerrada) resultan en una mayor profundidad de campo, y por lo tanto

<sup>5</sup>Este será el tamaño de los píxeles para la resolución escogida, el tamaño real de los píxeles de la cámara es distinto.

Cuadro 3.1: Parámetros para cada captura de imágenes

	<b>Apertura</b>	<b>Tiempo exposición</b>	<b>d</b>	<b>ISO</b>	<b>Máximo diámetro blur</b>
<b>Captura 1</b>	F 2.8	1/10	1.40m	ISO 100	29,88 píxeles
<b>Captura 2</b>	F 3.5	1/10	1.40m	ISO 100	23,9 píxeles
<b>Captura 3</b>	F 4.0	1/8	1.40m	ISO 100	20,92 píxeles
<b>Captura 4</b>	F 2.2	1/8	2m	ISO 100	21,16 píxeles
<b>Captura 5</b>	F 2.5	1/8	2m	ISO 100	18,6 píxeles

el diámetro de blur máximo es menor. Por otro lado es importante ver cómo la distancia  $d$  es inversamente proporcional al diámetro de blur (ver Ecuación 3.13). De este modo en las capturas 4 y 5 aumentando la distancia  $d$  se obtienen diámetros de blur máximos del orden de la Captura 3, a pesar de estar empleando valores de apertura menores (apertura más abierta, profundidad de campo menor). En la Sección 3.5.3 se mostrará con resultados reales la influencia de estos parámetros en el proceso de recuperación de imágenes nítidas.

### 3.5.3. Resultados del proceso de *deblurring*

A continuación se muestran los resultados obtenidos al deconvolucionar las imágenes capturadas con las PSFs obtenidas en la calibración para cada profundidad. El método de deconvolución utilizado es, en principio, la deconvolución sparse explicada en el Apéndice A, pero también se realizar un estudio comparando distintos algoritmos de deconvolución. Además, se analiza también la influencia de los distintos parámetros de captura en la calidad de la imagen enfocada resultante.

En primer lugar se debe notar que la deconvolución no resulta óptima para tamaños de PSF mayores de unos 19 píxeles. Esto se debe a que cuando la PSF es demasiado grande introduce artefactos como el *ringing* en la deconvolución.

La cámara reflex empleada no permite capturar tamaños de imágenes menores de 2352x1568 píxeles. Por lo tanto se opta por realizar un escalado de la imagen y las PSFs a tamaños menores, manteniendo siempre la relación entre las dimensiones de la imagen y la PSF. Esto se muestra en la Figura 3.10, que muestra, a la izquierda, la escena recuperada utilizando la imagen y la PSF sin reescalar, del tamaño original. A la derecha de la figura se muestra el resultado de escalar tanto la imagen como la PSF, obteniéndose una imagen de mayor calidad, sin *ringing*. Todas las imágenes que se muestran aquí han sido por tanto reescaladas a 1107x1694 píxeles, y las PSFs correspondientes se han escalado en la misma proporción.

Para las capturas 1, 2 y 3, se realiza una comparativa entre cuatro métodos distintos de deconvolución:

1. Deconvolución de Richardson-Lucy [17, 11]
2. Deconvolución de Wiener [4]
3. Deconvolución usando un prior de derivadas gaussianas (ver Apéndice A)
4. Deconvolución usando un prior de derivadas sparse (ver Apéndice A)

Se muestra la recuperación de un objeto situado a 45 cm para el caso de la Captura 1 (ver Figura 3.11), otro objeto a 85cm para la Captura 2 (ver Figura 3.12) y la recuperación de un objeto situado a 105 cm para la Captura 3 (ver Figura 3.13).



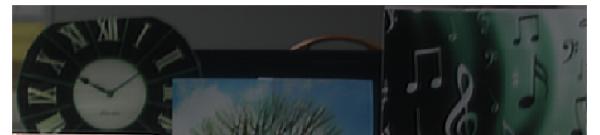
Imagen recuperada con PSF de tamaño 23x23



Imagen recuperada con PSF escalada a 13x13



Detalle del objeto recuperado (reloj, 105cm)



Detalle del objeto recuperado (reloj, 105cm)

Figura 3.10: *Ringing* debido al uso de kernel o PSF de tamaño elevado. Izquierda: Imagen recuperada con el kernel de tamaño original a 105cm, en el detalle del objeto se puede aprecia el *ringing* resultante. Derecha: Imagen recuperada al escalar el kernel a un tamaño menor, el objeto a 105cm aparece perfectamente recuperado sin presentar artefactos de *ringing*.

Se fijan los parámetros empleados en cada método, para la deconvolución de Richardson-Lucy se fija el número de iteraciones en diez, en el caso de la deconvolución de Wiener se establece el parámetro NSR (*Noise to signal power ratio* del ruido aditivo) en 0.005. En el caso de las deconvoluciones con uso de prior de derivadas se establece en ambos casos el término regularizador de *smoothing* en 0.004, y para la deconvolución sparse se establecen el número de iteraciones en 20. La deconvolución sparse lleva un tiempo de 4 minutos 46 segundos para una imagen de tamaño 1107x1694 píxeles en un procesador Inter Core i3 M370 @2.4 GHz. El tiempo de cálculo de los otros métodos no supera los 5 segundos. La deconvolución de Wiener, al realizar una estimación del ruido aditivo NSR, suele presentar imágenes recuperadas con mucho ruido. En cuanto a la deconvolución de Richardson-Lucy tiende a suavizar demasiado el resultado, y además suele presentar *ringing* en las imágenes recuperadas. Por último, la deconvolución que hace uso de prior gaussiano presenta también *ringing*, mientras que la deconvolución sparse al concentrar las derivadas en un número pequeño de píxeles da como resultado imágenes con bordes más marcados, reduce el ruido y ayuda a eliminar artefactos no deseados como el *ringing*. La recuperación de objetos para la Captura 1 (ver parámetros de captura en el Cuadro 3.1) no resulta óptima para distancias mayores de 65 cm ya que el diámetro de blur presente es elevado, y aunque recupera el enfoque comienzan a aparecer artefactos de *ringing*. En la Figura 3.11 se muestra la recuperación de un objeto situado a 45cm. En los ejemplos siguientes para la Captura 2 (ver Figura 3.12) y la Captura 3 (ver Figura 3.13), se muestran respectivamente objetos recuperados a 85cm y 105cm. En ambos casos se observa la mejor calidad del objeto recuperado mediante la deconvolución sparse, el objeto recuperado no presenta ni artefactos de *ringing* ni ruido como en los otros métodos, el resultado es una recuperación nítida de alta calidad. Finalmente las imágenes de la Captura 4 y Captura 5 presentan un grado de desenfoque demasiado elevado (presentan valores de apertura mayores, f2.2 y f2.5, resultando una profundidad de campo menor), por lo tanto los resultados obtenidos para la deconvolución sparse presenta *ringing* del mismo modo que sucedía a partir de cierta distancia de desenfoque para la Captura 1. En

la Figura 3.14 se muestra el resultado de la deconvolución sparse para ambas capturas. Además, en la Figura 3.15 se muestra cómo al deconvolucionar con una escalado de la PSF, la deconvolución introduce artefactos en aquellos objetos que fueron blurreados con una escala de la PSF mayor o menor.



Figura 3.11: Comparación de distintos algoritmos de deconvolución de un objeto situado a 45 cm del plano focal y capturado con la apertura codificada. (Captura 1, valor de apertura  $f = 2.8$ ,  $d = 1.40m$ ).



Figura 3.12: Comparación de distintos algoritmos de deconvolución para un objeto situado a 85 cm del plano focal y capturado con la apertura codificada. (Captura 2, valor de apertura  $f = 3.5$ ,  $d = 1.40m$ ).

Es decir, escalas incorrectas del filtro, introducen artefactos en el proceso de deconvolución, tal y como se expresaba en el criterio de selección del filtro.

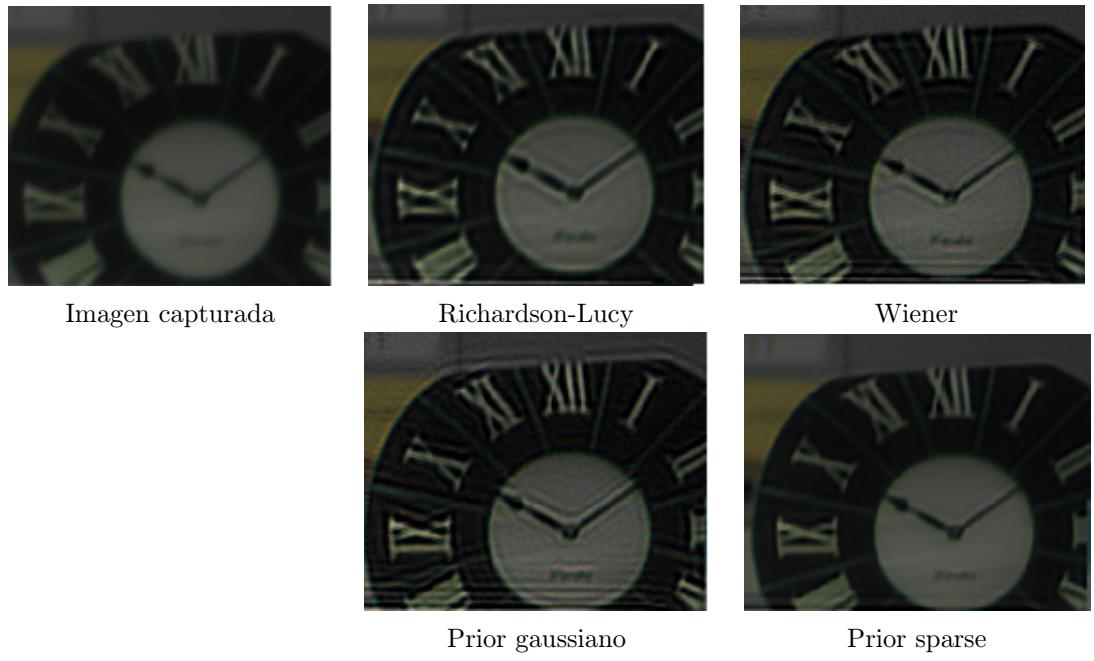


Figura 3.13: Comparación de distintos algoritmos de deconvolución para un objeto situado a 105 cm del plano focal y capturado con la apertura codificada. (Captura 3, valor de apertura f 4.0, d=1.40m).

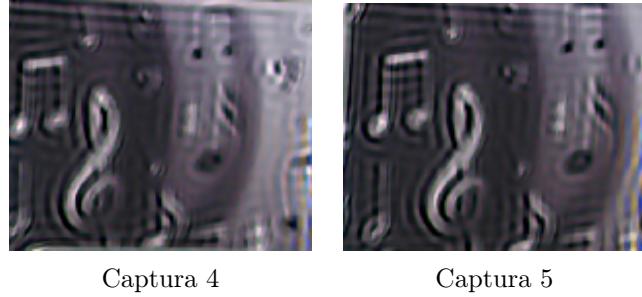


Figura 3.14: Detalles a 95 cm recuperados con *ringing* debido al elevado grado de desenfoque. Izquierda: detalle recuperado para la Captura 4. Derecha: detalle recuperado para la Captura 5.

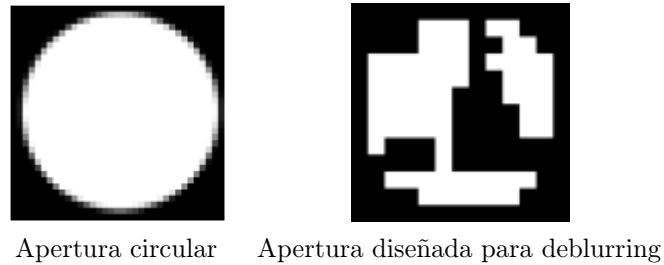


Figura 3.16: Aperturas empleadas para analizar la deconvolución sparse.

El procesado y análisis de las imágenes obtenidas en las cinco capturas muestra que a partir de cierto grado de desenfoque la deconvolución sparse deja de presentar mejoras en los resultados frente al resto de métodos de deconvolución. Por otro lado, se sabe que la apertura codificada empleada está diseñada para DFD, y por lo tanto su respuesta

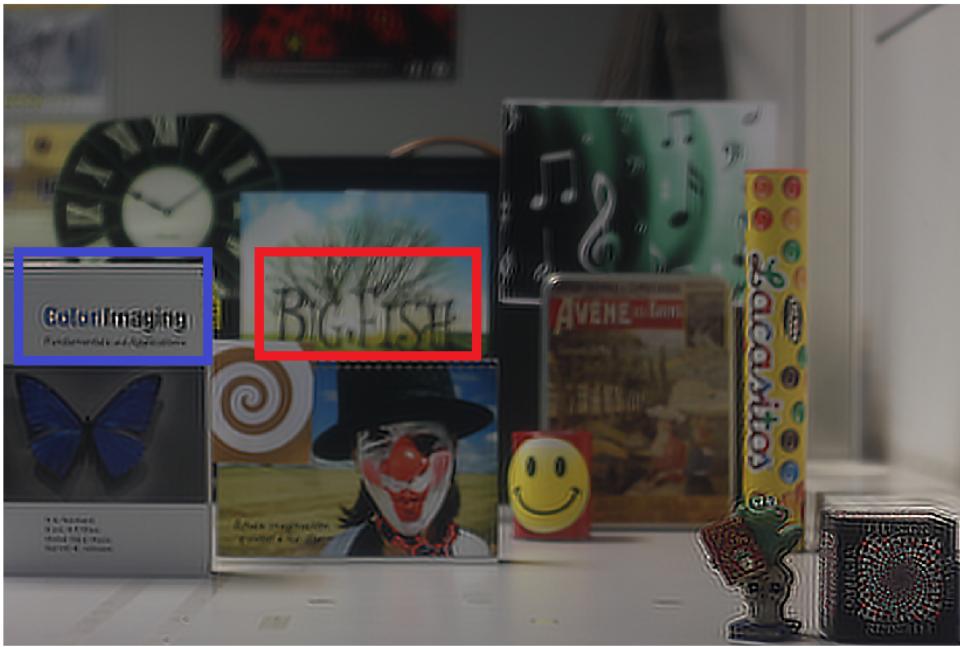


Figura 3.15: Aparición de artefactos en la deconvolución al usar un kernel de tamaño menor (caso de la zona señalada en rojo) o mayor (caso de la zona señalada en azul) del adecuado. La imagen está deconvolucionada con la PSF correspondiente a la profundidad del payaso (45cm), se puede apreciar el *ringing* introducido para el resto de objetos cuya profundidad es distinta y por lo tanto sus PSFs serán mayores o menores según corresponda.

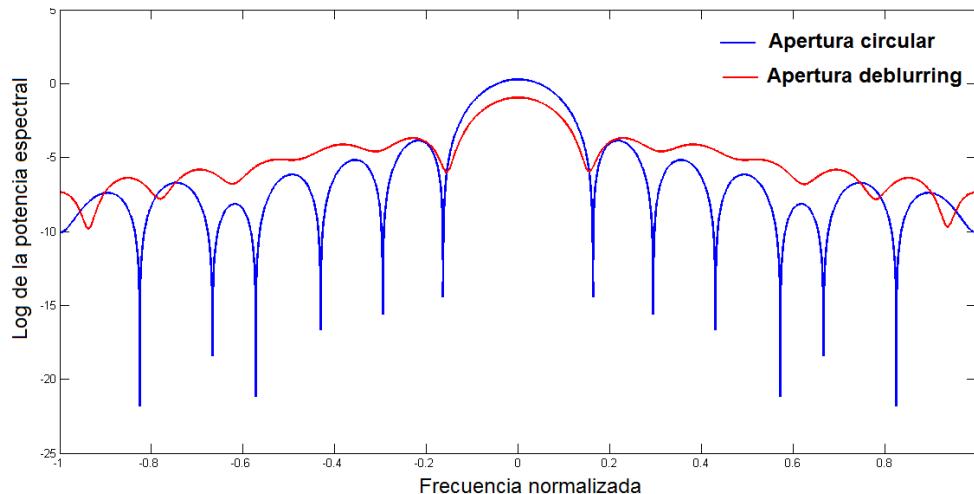


Figura 3.17: Izquierda: Logaritmo de la potencia espectral para la apertura circular y la apertura codificada diseñada para *deblurring* [24]. Derecha: Apertura circular y apertura diseñada para *deblurring*.

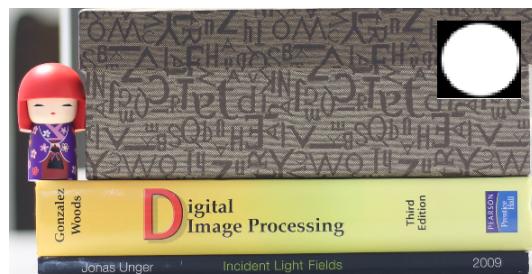
no resulta óptima para *deblurring*. Para comprender si la limitación viene dada por la deconvolución sparse o por la apertura codificada, se decide estudiar el comportamiento de la deconvolución sparse con una apertura diseñada para *deblurring* [24]. Además se compararán los resultados de la apertura diseñada para *deblurring* con el de la apertura circular convencional.

Para las dos aperturas propuestas en la Fig 3.16, la apertura circular convencional y una apertura diseñada para *deblurring*, se captura con cada apertura imágenes de objetos desenfocados a 40 cm y 70 cm de distancia respecto del plano focal (ver parámetros de captura en el Cuadro 3.2). En las siguientes figuras, Figura 3.18, Figura 3.19, Figura 3.20 y Figura 3.21, se muestran las imágenes enfocadas (*ground truth*), las imágenes desenfocadas capturadas y las imágenes recuperadas mediante deconvolución sparse y deconvolución de Wiener para cada apertura.

Se puede ver cómo la deconvolución sparse no resulta favorable cuando el grado de desenfoque es elevado, a pesar de que la apertura codificada sea óptima para *deblurring* (Figura 3.18, Figura 3.19). En casos de grado de desenfoque elevado la deconvolución de Wiener ofrece resultados mejores aunque en ningún caso la imagen resultante consigue la nítidez que se obtenía con la deconvolución sparse en imágenes con menor grado de desenfoque. Asimismo, la recuperación de los objetos en las imágenes capturadas con la apertura circular convencional muestran una calidad claramente inferior a los de la apertura diseñada para *deblurring*, demostrando nuevamente las aportaciones del uso de aperturas codificadas.

Cuadro 3.2: Parámetros para cada captura de imágenes

	<b>Apertura</b>	<b>Tiempo exposición</b>	<b>d</b>	<b>ISO</b>	<b>Diámetro teórico blur</b>
<b>d' = 1,4m</b>	F 2.0	1/10	1.0m	ISO 100	39,6 píxeles
<b>d' = 1,7m</b>	F 2.0	1/10	1.0m	ISO 100	69,39 píxeles



Objeto enfocado



Imagen capturada desenfocada



Recuperación deconvolución sparse



Recuperación deconvolución de Wiener

Figura 3.18: Apertura circular. Recuperación de objetos desenfocados situados a 40 cm del plano focal con deconvolución sparse y con deconvolución de Wiener.



Objeto enfocado



Imagen capturada desenfocada

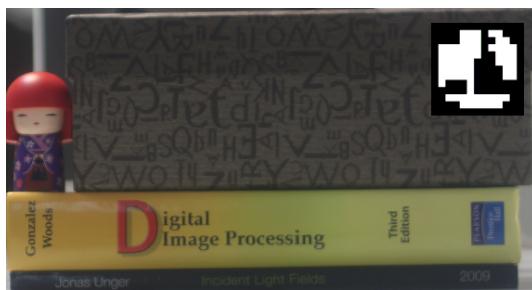


Recuperación deconvolución sparse



Recuperación deconvolución de Wiener

Figura 3.19: Apertura circular. Recuperación de objetos desenfocados situados a 70 cm del plano focal con deconvolución sparse y con deconvolución de Wiener.



Objeto enfocado

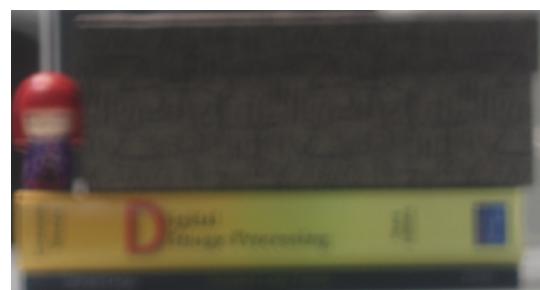
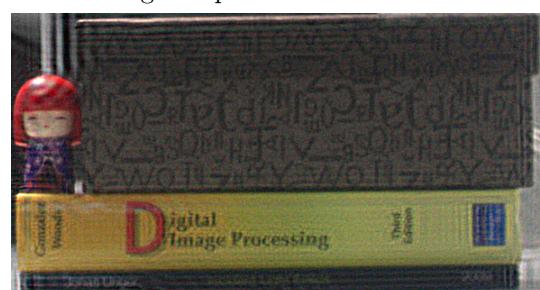


Imagen capturada desenfocada



Recuperación deconvolución sparse



Recuperación deconvolución de Wiener

Figura 3.20: Apertura codificada diseñada para deblurring. Recuperación de objetos desenfocados situados a 40 cm del plano focal con deconvolución sparse y con deconvolución de wiener.

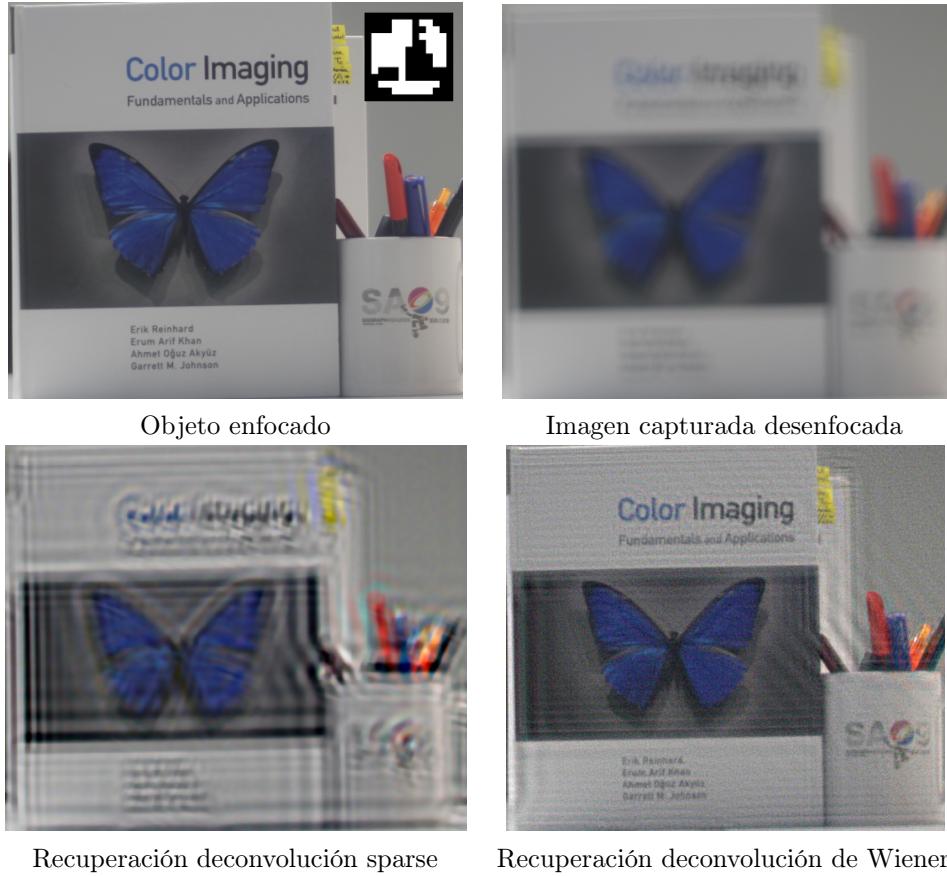


Figura 3.21: Apertura codificada diseñada para *deblurring*. Recuperación de objetos desenfocados situados a 70 cm del plano focal con deconvolución sparse y con deconvolución de Wiener.

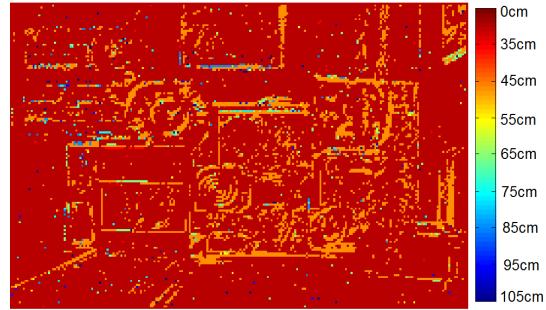
### 3.5.4. Resultados de la obtención del mapa de profundidades

Como hemos visto en la Sección 3.4, para estimar la profundidad de cada píxel hay que obtener el error de reconstrucción, en una ventana alrededor del píxel, para las distintas profundidades posibles y pesar dichos errores con una serie de pesos  $\lambda_i$ . La profundidad que dé error mínimo tras el promediado con los citados pesos será la correcta para cada píxel. A continuación se muestran los resultados obtenidos al realizar dicho cálculo sobre la escena para obtener su mapa de profundidades. La Figura 3.22 muestra dichos resultados. En primer lugar se muestra la escena capturada (parámetros de Captura 5) de la cual se quiere obtener el mapa de profundidades. A su derecha se muestra el mapa de profundidades obtenido promediando la energía del error de reconstrucción en ventanas locales de 10x10 píxeles pero sin aplicar el vector de pesos. El resultado es que el error de reconstrucción en la mayor parte de las ventanas locales de la imagen es mínimo para la profundidad menor (0cm, i.e. el plano focal, representado en color granate en la escala). Se ilustra así la necesidad de multiplicar el error en las ventanas locales por un valor de peso para cada profundidad.

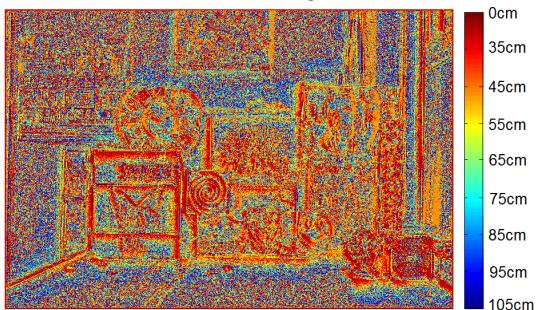
En la parte inferior de la Figura 3.22 puede verse a la izquierda el mapa de profundidades obtenido si se pesa el error a distancia 0 cm con un valor superior a la unidad mayor mientras que el resto de profundidades se pesan con la unidad. Es importante resaltar como las zonas con textura de los objetos y los bordes muestran un error mínimo para la misma profundidad, mientras que las zonas sin textura, como el fondo o la mesa, presentan valores de profundidad aleatorios. La imagen de la derecha muestra el mapa de profundidades obtenido con el vector de pesos completo.



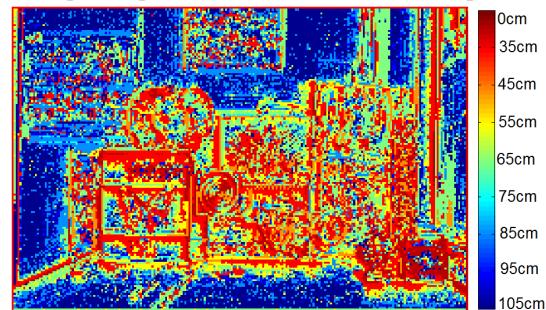
Escena original



Mapa de profundidades sin vector de pesos



Mapa de profundidades pesando distancia en foco



Mapa de profundidades con vector de pesos

Figura 3.22: Mapas de profundidades considerando y sin considerar vector de pesos.

Se deben tener en cuenta varios puntos de este método:

- El vector de pesos muestra una fuerte dependencia tanto con el sistema como con la imagen.
- Las zonas sin textura no contienen información que permita que el error de reconstrucción varíe para escalas de PSFs incorrectas.
- La cantidad de ventanas locales de profundidad conocida junto con el algoritmo de aprendizaje resulta en una complejidad que no se corresponde con la calidad de los resultados obtenidos.

Por estas razones se considera explorar un método más eficiente para la estimación de profundidades, que no presente una dependencia tan grande con el sistema y la imagen capturada. La opción a explorar escogida es el uso de pares de aperturas codificadas y se presenta en el Capítulo 4.

## Capítulo 4

# Pares de aperturas codificadas

### 4.1. Introducción

Tal y como hemos visto, el uso de aperturas circulares convencionales no resulta óptimo para la obtención de mapas de profundidades a partir de desenfoque. Además, la estimación del grado de desenfoque partiendo de una única imagen, como en el Capítulo 3, resulta un problema complicado. En general los métodos de estimación de mapa de profundidades y corrección de desenfoque basados en una única imagen necesitan hacer uso de priors de imágenes de gran peso en su deconvolución o incluir intervención del usuario. Para una estimación más robusta de la profundidad, existen varios métodos de DFD que capturan dos o más imágenes usando parámetros de la cámara con distinta información de desenfoque. Esto se puede realizar cambiando el enfoque, es decir, variando la distancia entre la lente y el sensor de la cámara, o modificando el patrón de la apertura. En este proyecto optaremos por la segunda opción.

Una técnica típica de DFD se basa en la captura de dos imágenes con un par de aperturas circulares. Una imagen se captura con la apertura circular abierta y otra con la apertura circular más cerrada. Una apertura menor nos proporcionará una mayor profundidad de campo, por lo que mayor rango de la escena aparecerá enfocado, mientras que el uso de una apertura mayor limitará la región enfocada de la escena. Por lo tanto, la imagen capturada con la apertura mayor mostrará mayor grado de desenfoque que la capturada con apertura menor, este desenfoque relativo será usado para la estimación de profundidad. Se ha de tener en cuenta que el desenfoque relativo estará muy influenciado por el patrón de la apertura, por lo que de nuevo hay métodos, como el que se usa en este proyecto, que optan por cambiar el patrón de la apertura circular por otro cuya respuesta en frecuencia sea más favorable. La mayoría de los patrones y/o diámetros de apertura usados años atrás se obtenían intuitivamente, para luego analizarlos en detalle en términos de sus funciones de transferencia óptica. En los últimos años el diseño de aperturas para corrección de desenfoque y DFD ha empezado a plantearse como un problema de optimización. En este proyecto se utiliza un algoritmo genético para la obtención del par de aperturas óptimo basándose en el trabajo de Zhou *et al.*[23], que presentan una métrica que indica cuándo un par de aperturas es óptimo para recuperación de profundidad.

### 4.2. Función objetivo

Como hemos dicho, el diseño del par de aperturas codificadas a utilizar se va a plantear como un problema de optimización. Esta sección explica la función objetivo a optimizar , que indicará, dado un par de aperturas, cómo de eficaz es a la hora de estimar profundidad.

Recordemos que para un objeto plano frontal, su imagen desenfocada  $f$  puede expresarse como:

$$f = f_0 * k^d + \eta \quad (4.1)$$

donde  $f_0$  es la imagen real enfocada,  $\eta$  es el ruido de la imagen asumido ruido blanco gaussiano de media cero y varianza  $\sigma^2$ ,  $N(0, \sigma^2)$ , y  $k^d$  es la PSF cuyo patrón está determinado por la apertura y cuyo tamaño  $d$  está relacionado con la profundidad. De este modo estimando el tamaño del *blur* de desenfoque podremos calcular la profundidad.

La Ecuación 4.1 puede expresarse en el dominio de la frecuencia como

$$F = F_0 \cdot K^d + \zeta, \quad (4.2)$$

donde  $F_0, K^d$ , y  $\zeta$  son las transformadas discretas de Fourier de  $f, f_0$ , y  $\eta$ , respectivamente. Se ha visto que una única imagen desenfocada es generalmente insuficiente para estimar la profundidad de la escena sin el uso de información adicional. Por ello este método procesa dos imágenes  $F_{i=1,2}$  de una misma escena capturadas con dos aperturas codificadas distintas, de modo que la PSF es diferente para cada imagen. Cada imagen tendrá distintas características de desenfoque y la combinación de esta información nos permitirá estimar profundidades. De este modo la captura de cada imagen  $i$  en el dominio frecuencial se puede expresar como

$$F_i = F_0 \cdot K_i^{d^*} + \zeta_i \quad (4.3)$$

donde  $K_i^{d^*}$  denota la transformada de Fourier de la  $i$ -ésima PSF con el tamaño de blur real  $d^*$  para esa profundidad. El objetivo es encontrar el tamaño de blur  $\hat{d}$  (relacionado con la profundidad) y la imagen enfocada  $\hat{F}_0$  resolviendo un problema de máximo a posteriori (MAP). Una descripción completa del proceso se puede consultar en el Apéndice C, donde se muestra que el problema deriva en la minimización de la función

$$E(\hat{d}|F_1, F_2, \sigma) = \min_{F_0} \sum_{i=1,2} \left\| \hat{F}_0 \cdot K_i^{\hat{d}} - F_i \right\|^2 + \|C \cdot \hat{F}_0\|^2, \quad (4.4)$$

donde  $C = \sigma \cdot \Psi$ , siendo  $\Psi$  una matriz de pesos. Para optimizar  $C$  se usará la ley  $1/f$  de imágenes naturales [19, 22]. Se define pues  $|C|^2 = \sigma^2/A$ , siendo  $\sigma^2$  la varianza del ruido y donde  $|A^2|$  se corresponde con la potencia espectral esperada en las imágenes naturales:  $|A(\xi)|^2 = \int_{F_0} |F_0(\xi)|^2 \mu(F_0)$ ; siendo  $\xi$  la frecuencia y  $\mu(F_0)$  la posibilidad de la medida de la muestra  $F_0$  en el espacio de imágenes. De este modo ahora se parte de dos imágenes capturadas  $F_1, F_2$ , y de la desviación estándar del ruido  $\sigma$ , es decir, en el problema de deconvolución se va a hacer uso de dos imágenes desenfocadas para la obtención de una única imagen enfocada  $\hat{F}_0$ . Se utiliza la deconvolución generalizada de Wiener, cuya expresión se puede obtener, para una  $\hat{d}$  dada, resolviendo  $\partial E/\partial \hat{F}_0 = 0$  para la Ecuación 4.4, obteniendo:

$$\hat{F}_0 = \frac{F_1 \cdot \bar{K}_1^{\hat{d}} + F_2 \cdot \bar{K}_2^{\hat{d}}}{|K_1^{\hat{d}}|^2 + |K_2^{\hat{d}}|^2 + |C|^2}. \quad (4.5)$$

$\bar{K}$  es el complejo conjugado de  $K$  y  $|X|^2 = X \cdot \bar{X}$ . Esta expresión se conoce como deconvolución generalizada de Wiener porque deriva de la deconvolución de Wiener, pero en este caso se parte de dos imágenes desenfocadas, cada una con una PSF distinta, y a partir de esta información se obtiene una única imagen enfocada.

Para que se obtenga un buen resultado en el proceso de deconvolución es muy importante que la suma de la potencia espectral del par de aperturas sea una respuesta frecuencial de banda ancha que cubra el espectro de frecuencias. Esto resulta más sencillo

de satisfacer usando más de un patrón de apertura, o lo que es lo mismo, disponiendo de múltiples imágenes desenfocadas con distintas PSF. Esto se ilustra en la Figura 4.2, que muestra la potencia espectral y la fase para una pareja de aperturas circulares (izquierda) y para el par de aperturas codificadas obtenido en el mencionado trabajo de Zhou *et al.*[23]. Este par de aperturas codificadas muestra una importante compensación tanto en potencia espectral como en fase. La compensación en fase conduce a una especie de efecto estereoscópico en la captura de imágenes, se obtiene un desenfoque relativo que permite favorecer la estimación de profundidades. Además de optimizar el par de aperturas codificadas para DFD, la potencia espectral conjunta (en rojo) presenta una respuesta en frecuencia de banda ancha, mejorando el proceso de *deblurring* y por tanto permitiendo recuperar una imagen enfocada de alta calidad.

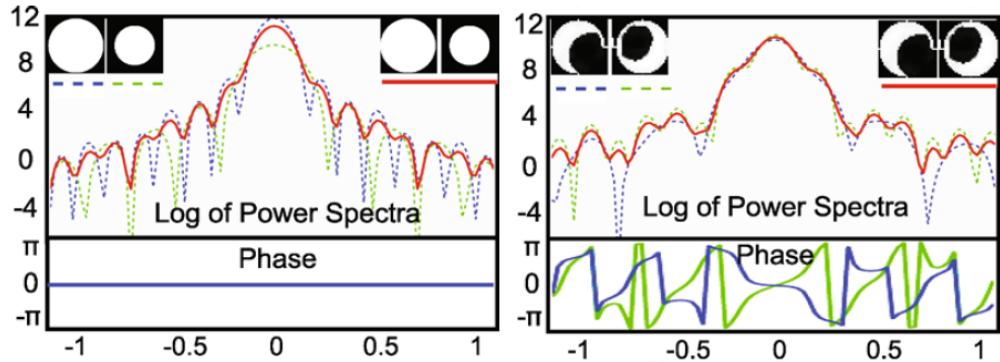


Figura 4.2: Potencia espectral y fase del par de aperturas convencionales (izquierda) y del par optimizado de aperturas codificadas (derecha).(Adaptado de Zhou *et al.*[23])

A partir de las ecuaciones 4.2, 4.4 y 4.5 se deriva el criterio de selección utilizado para la búsqueda del par de aperturas óptimas (para una descripción completa ver Apéndice C), resultando el siguiente criterio:

$$R = \min_{d \in D/d^*} M(K_1, K_2, d, d^*), \quad (4.6)$$

siendo  $M(K_1, K_2, d, d^*)$ :

$$M(K_1, K_2, d, d^*) = \left[ \frac{1}{n} \sum_{\xi} A \cdot \frac{|K_1^d K_2^{d^*} - K_2^d K_1^{d^*}|^2}{|K_1^d|^2 + |K_2^d|^2 + C^2} \right]^{1/2}, \quad (4.7)$$

con  $n$  el número de píxeles de la PSF.

En la Ecuación 4.6,  $D = \{c_1 d^*, c_2 d^*, \dots, c_l d^*\}$  es un conjunto de valores de distintos tamaños de blur (correspondientes por tanto con distintos valores de profundidad) para los cuales se evaluará la respuesta del par de secuencias. El tamaño de ese vector establece por tanto un compromiso entre el tiempo de cálculo del algoritmo de optimización y la precisión de la solución. En nuestra implementación, se toma un conjunto de valores suficientemente amplio y que a su vez su tiempo de cálculo sea aceptable,  $\{c_i\} = \{0, 1, 0, 2, \dots, 1, 5\}$  y  $d^* = 15$ .

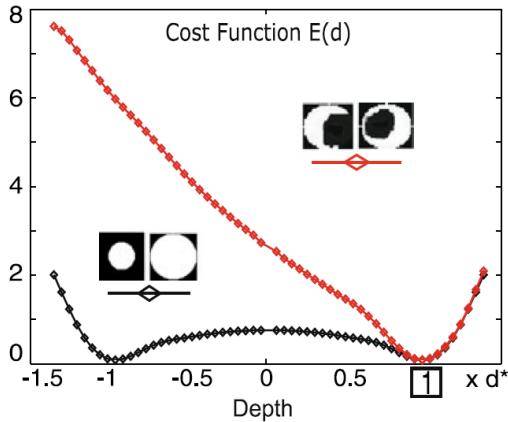


Figura 4.3: Función de coste  $E(d)$  para el par óptimo de Zhou (en rojo) y el par de aperturas circulares (en negro).  $E(d)$  presenta una mayor pendiente para el par óptimo, y por lo tanto la estimación del mínimo es más robusta frente al ruido. (Adaptado de Zhou *et al.*[23])

Un valor mayor de  $R$  en la Ecuación 4.6 indica que la función de energía (ver Figura 4.3) para DFD tiene una pendiente mayor , y por lo tanto su estimación será más robusta frente a ruido de la imagen y zonas sin textura. De acuerdo a la función de evaluación  $M$ , Ecuación 4.7, se evalúa cada secuencia para los distintos valores  $d \in D/d^*$ . Posteriormente se calcula el valor de  $R$  de esa secuencia que será el mínimo valor de  $M$ , que corresponde con el caso más desfavorable. Finalmente se seleccionarán aquellas secuencias cuyo valor obtenido de evaluar la función  $R$  sea mayor.

### 4.3. Algoritmo genético para obtención de pares de aperturas codificadas

Como ya hemos visto, la obtención de un patrón de apertura no es un problema sencillo. El criterio de evaluación de la apertura es habitualmente complicado y suele estar formulado en el dominio de Fourier. Además en el dominio espacial el patrón de la apertura debe cumplir ciertas imposiciones físicas:

- Los valores de transmitancia de la apertura deben tomar valores entre 0 y 1
- El patrón completo debe ser del mismo tamaño que la lente del objetivo (20mm en nuestro caso).
- Su resolución espacial debe ser lo más pequeña posible (i.e. se debe evitar que tenga píxeles de tamaño muy pequeño) para evitar efectos de difracción.

Por otro lado, la complejidad de la función de evaluación (ecuaciones 4.6 y 4.7) hace que no sea factible derivarla analíticamente. Además, el espacio de búsqueda de posibles aperturas es enorme, ya que para un patrón binario de resolución  $N \times N$  el número de posibles aperturas asciende a  $2^{N \times N}$ . Por estas razones, para resolver el problema de optimización, implementaremos un algoritmo genético que nos permita encontrar la pareja de aperturas óptimas de tamaño  $11 \times 11$ .

Los algoritmos genéticos se inspiran en la evolución biológica y su base genético-molecular. Estos algoritmos hacen evolucionar una población de individuos sometiéndola a acciones aleatorias semejantes a las que actúan en la evolución biológica (mutaciones y recombinaciones genéticas), así como también a una selección de acuerdo con algún

criterio, en función del cual se decide cuáles son los individuos más adaptados, que sobreviven, y cuáles los menos aptos, que son descartados.

En nuestro caso, los cromosomas, que serán vectores de 242 elementos, se van a corresponder cada uno con una pareja de aperturas (los 121 primeros elementos se corresponden con la apertura 1 del par (11x11) y los 121 elementos siguientes con la apertura2 (11x11)). De modo que los elementos del vector/cromosoma se corresponden con píxeles de la máscara/apertura, que serán binarios (transparentes u opacos).

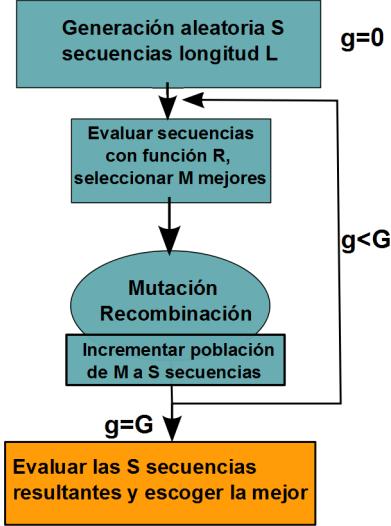


Figura 4.4: Esquema del algoritmo genético implementado

El proceso del algoritmo genético implementado se describe a continuación. Llamaremos  $g$  a la iteración (o generación) del algoritmo en que nos encontramos en cada caso.

1. Inicialización:  $g=0$ . Generación aleatoria de  $S$  secuencias binarias de longitud  $L$ .
2. Para  $g=1 : G$ 
  - a) *Selección*: Se evalúan las  $S$  secuencias con la función de evaluación (Ecuación 4.6). Se seleccionan las  $M$  secuencias mejores de las  $S$  secuencias disponibles.
  - b) Repetir recombinación y mutación hasta que la población (número de secuencias) crezca nuevamente de  $M$  a  $S$ .
    - *Recombinación*: Duplicar dos secuencias seleccionadas aleatoriamente de las  $M$  secuencias obtenidas en el punto 2.(a), alinearlas e intercambiar cada pareja de sus correspondientes pares de bits con una probabilidad  $c1$ , obteniendo dos nuevas secuencias.
    - *Mutación*: Para cada nueva secuencia generada, mutar parejas de sus bits con una probabilidad  $c2$ .
3. Evaluar las  $S$  secuencias finales con la Ecuación 4.6 y seleccionar la mejor (i.e. la de mayor valor de  $R$ ).

Nuestro objetivo es encontrar un par de aperturas de tamaño 11x11 óptimas. Por ello, nuestras secuencias del algoritmo genético serán de longitud  $L=242$  (dos aperturas

de longitud 121 píxeles cada una, es decir dos aperturas de 11x11). En el incremento de la población de M a S secuencias, establecemos la probabilidad de recombinación  $c1$  en 0.2, y la probabilidad de mutación  $c2$  en 0.05. En cuanto a la población inicial S (número de secuencias iniciales), población final M (número de secuencias finales) y número de generaciones G, se realizan varias ejecuciones variando estos parámetros, para determinar a partir de qué valores la convergencia del algoritmo es clara. En todas las pruebas la población final seleccionada es un 10 % de la población inicial.

Se seleccionan los pares de aperturas con mejor valor según el criterio de evaluación para cada configuración de parámetros, y se representan en la Figura 4.5, donde se puede observar que los pares de aperturas son complementarias, tal y como introducimos en el criterio de selección. En la Figura 4.6 se representan sus respectivas puntuaciones del criterio de evaluación (recordar que se buscan valores elevados de R(k)) para distinta población inicial S, población final M y número de generaciones G.

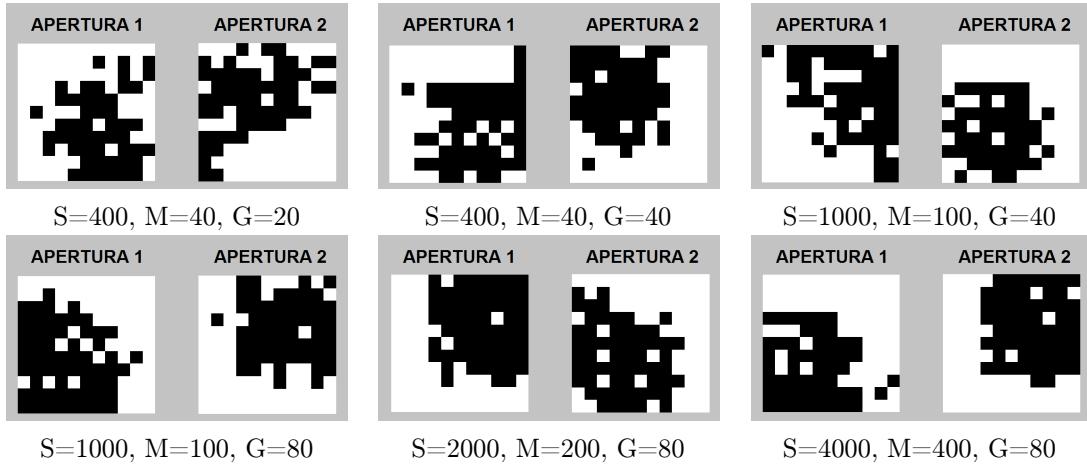


Figura 4.5: Pares de aperturas obtenidas para distintos parámetros iniciales del algoritmo genético: población inicial (S), población final (M) y número de generaciones (G).

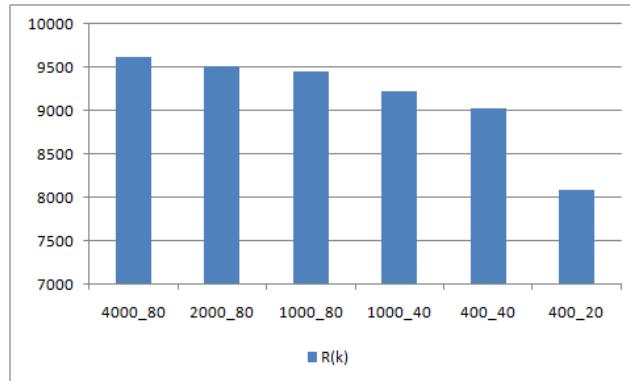


Figura 4.6: Convergencia del algoritmo genético para distintos parámetros iniciales. El eje X indica la combinación de población inicial (S) y número de generaciones (G) de cada ejecución, y el eje Y el R(k) correspondiente.

Debemos resaltar la carga computacional de este algoritmo. La función de evaluación requiere que para cada valor de blur  $d$  (15 valores) se calcule el escalado de los dos kernels, y sus respectivas transformadas de Fourier. Esto hay que hacerlo a su vez para cada secuencia (número de secuencias  $\leq 4000$ ) en cada generación; y repitiendo el proceso

para todo el número de generaciones (número de generaciones  $\leq 80$ ). La optimización para el caso más largo, S=4000 M=400 G=80, lleva un tiempo de cálculo elevado de 76,13 horas en un procesador Intel Core i7 950 @3.07 GHz, no resultando de interés para este proyecto la posible optimización del algoritmo genético,

#### 4.4. Recuperación de profundidad

Tras la obtención de pares de aperturas codificadas óptimas, se procede a estimar el mapa de profundidades  $U$  de la imagen. Para cada posible profundidad  $d \in D$ , se calcula la imagen recuperada  $\hat{F}_0^{(d)}$  según la Ecuación 4.5 obteniendo una imagen enfocada a partir de las dos imágenes desenfocadas. Para cada píxel, el residuo  $W^{(d)}$  entre las imágenes desenfocadas y la imagen recuperada  $\hat{F}_0^{(d)}$  indica una medida de cómo de cerca está  $d$  respecto de la profundidad real  $d^*$ :

$$W^{(d)} = \sum_{i=1,2} |IFT(\hat{F}_0^{(d)} * K_i^d - F_i)|, \quad (4.8)$$

donde IFT es la transformada inversa de Fourier en 2D. Para los pares de aperturas optimizadas, el valor de  $W^{(d)}(x, y)$  alcanza un mínimo para el píxel  $(x, y)$  si  $d$  coincide con la profundidad real  $d^*$ . Por lo tanto se puede obtener el mapa de profundidades  $U$  como:

$$U(x, y) = \arg \min_{d \in D} W^{(d)}(x, y), \quad (4.9)$$

y recuperar una imagen enfocada  $I$  como:

$$I(x, y) = \hat{F}_0^{(U(x, y))}(x, y). \quad (4.10)$$

#### 4.5. Resultados

A continuación se muestran los resultados obtenidos con pares de aperturas codificadas tanto para la estimación de profundidad como para *deblurring* o recuperación de imágenes nítidas.

Para ello se va a explicar en primer lugar la metodología usada para la validación y los distintos pares de aperturas empleados. Finalmente se muestran resultados de la estimación del mapa de profundidades y de la recuperación de imágenes enfocadas, realizando una comparativa entre ellos.

##### 4.5.1. Método de validación empleado

Para la validación física del par de aperturas se insertarían las máscaras de las aperturas codificadas en la lente del objetivo y se capturarían dos versiones de la escena con los mismos parámetros y misma posición de la cámara respecto de la escena. De este modo se obtienen dos imágenes  $F_1, F_2$  de la misma escena con distinto grado de desenfoque que posteriormente se procesarían para obtener el mapa de profundidades y la imagen enfocada, tal como se indica en la Figura 4.7. Por distintas cuestiones<sup>1</sup> se realiza una validación por simulación en vez de por soporte físico. Por lo tanto en vez de partir del par de imágenes capturadas con cada apertura, se parte de una única imagen de la escena que será desenfocada mediante simulación con cada una de las aperturas, obteniendo así  $F_1, F_2$ , tal y como se explica a continuación.

---

<sup>1</sup>Principalmente por la complejidad de obtener dos imágenes exactamente iguales con el par de aperturas. Entre una captura y otra sería necesario cambiar el objetivo de la cámara, e inevitablemente, con los medios que se disponen, la cámara sufriría un pequeño desplazamiento por lo que las imágenes capturadas no serán igual a nivel de píxel.

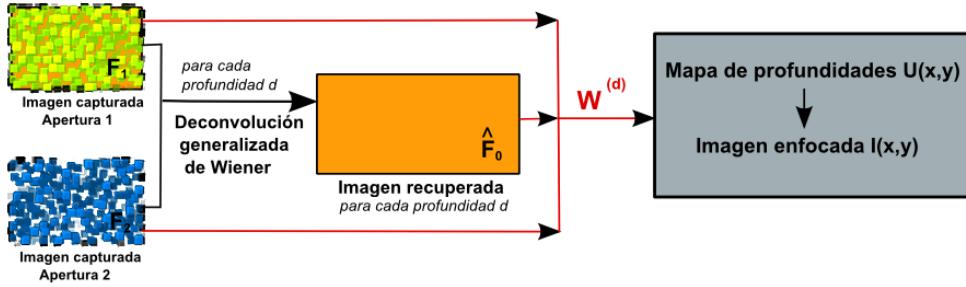


Figura 4.7: Diagrama de flujo del procedimiento de validación sobre soporte físico: obtención del mapa de profundidades  $U$  y la imagen enfocada  $I$  a partir de las dos imágenes capturadas con el par de aperturas.

En el proceso de captura de una imagen, la escena capturada se corresponde con la escena real convolucionada con la PSF a esa profundidad, tal y como indica la Ecuación 4.3. Por lo tanto cada zona de la imagen capturada tendrá un desenfoque proporcional a la distancia de esa zona al plano focal. De este modo las imágenes capturadas serán el resultado de la convolución de la escena original con la PSF a la profundidad real de cada zona de la escena. En la validación por simulación el desenfoque es introducido convolucionando distintas zonas de la imagen con distintas PSF del conjunto de profundidades  $d \in D$  consideradas. En una escena real el mapa de profundidades depende del tipo de la escena, sin seguir un patrón determinado, como se quiere mostrar en las imágenes capturadas de la Figura 4.7 con ventanas aleatorias a distintas profundidades. Por simplicidad, y sin pérdida de generalidad de los resultados, en nuestra simulación se opta por desenfocar la imagen de acuerdo a un mapa de profundidades horizontal. El procedimiento de validación por simulación puede verse en la Figura 4.8. Se parte de una imagen enfocada y se escoge un mapa de profundidades horizontal (podría ser otro cualquiera, pero desenfocar la imagen aleatoriamente es más costoso y no permite comprobar tan bien si la estimación del mapa de profundidades es correcta). De acuerdo al mapa de profundidades se escala adecuadamente para toda  $d \in D$  las distintas PSF de cada apertura y se obtienen las dos imágenes desenfocadas  $F_1$ , para la apertura 1, y  $F_2$  para la apertura 2. A partir de las dos imágenes desenfocadas, se procede a estimar el mapa de profundidades y la imagen enfocada tal y como se indica en la Sección 4.4. Debe tenerse en cuenta que en el proceso de validación por simulación se procede del mismo modo que en el de validación física, únicamente se tiene como información las dos imágenes desenfocadas. Lo único que cambia es la manera de obtener estas imágenes desenfocadas.

Habitualmente zonas de la imagen con poca textura no permiten estimar con precisión su profundidad. Esto sucede debido a que el residuo calculado en la Ecuación 4.8, en las zonas sin textura no se dispone de información suficiente, por lo tanto los residuos a distintas profundidades no varían. De este modo, resulta más complicado determinar la profundidad a la que se encuentran zonas sin textura o con textura suave.

Para la validación del par de aperturas se usa como escena una imagen que contiene una zona con una fuerte textura y otra con poca textura, ver Figura 4.5.1.

Se realiza un estudio sobre los resultados obtenidos por los seis pares de aperturas codificadas obtenidos con el algoritmo genético, ver Figura 4.5; y además se incluyen en el estudio el par de aperturas circulares, un par de aperturas obtenidas aleatoriamente y el par óptimo obtenido por Zhou *et al.* [23], ver Figura 4.10. Todas las aperturas son de tamaño 33x33, para poder realizar la comparación con el par óptimo de Zhou *et al.* Este último parte de dos aperturas 11x11 y realiza un escalado junto con optimización

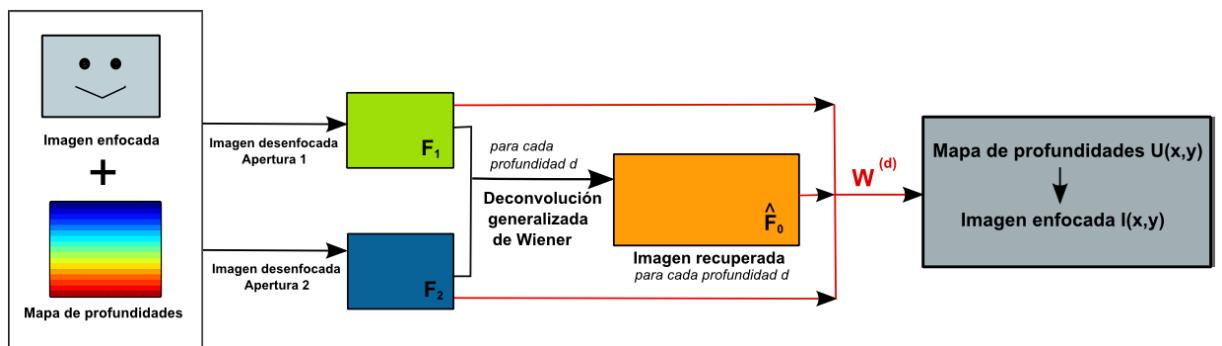


Figura 4.8: Diagrama de flujo del procedimiento de validación por simulación. Los detalles del proceso se explican en el texto.



Figura 4.9: Imagen usada para la validación del par de aperturas. Izquierda: textura fuerte. Derecha: textura suave.

de gradiente descendente que suavice la solución. Esto se realiza progresivamente hasta alcanzar un tamaño de apertura de 33x33. En nuestro caso se parte de las aperturas de 11x11 obtenidas con nuestro algoritmo genético y se realiza un escalado sencillo (*nearest neighbor*) hasta alcanzar el tamaño de 33x33.

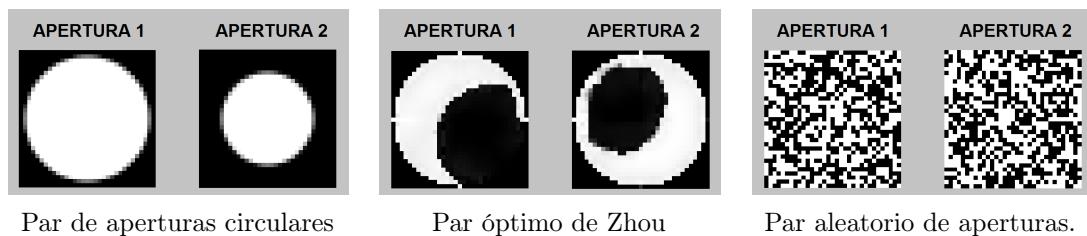


Figura 4.10: Pares de aperturas incluidas en la validación por simulación junto con las obtenidas en el algoritmo genético.

#### 4.5.2. Resultados de la obtención de mapas de profundidades

En la Figura 4.11 se muestra el porcentaje de error de cada par de aperturas al estimar el mapa de profundidades. El mapa de profundidades tiene trece profundidades posibles, por lo tanto el error máximo será que se estime como 1 la profundidad de un objeto situado a profundidad 13. En cambio un objeto situado a profundidad 6 su

error máximo de estimación será 7 (es decir, que se estime la profundidad como 13). De este modo el cálculo del porcentaje de error se realiza así, teniendo en cuenta el error máximo posible para cada banda de profundidad. El par de aperturas circulares junto al par obtenido de manera aleatoria muestran errores muy altos especialmente para las zonas de textura suave o poca textura. El par de aperturas circulares se comporta mejor para zonas con textura, pero aún así el porcentaje de error es mayor que para cualquier par de aperturas codificadas. Por lo tanto se muestra la mejora introducida al usar pares de aperturas codificadas frente a las aperturas convencionales. En el gráfico de la derecha se muestra la comparativa entre los pares de aperturas codificadas. En general se ve una tendencia a que el porcentaje de error disminuya, principalmente en las zonas con textura suave, conforme aumentamos la población inicial y número de generaciones. Nuestro par de aperturas con menor porcentaje de error se corresponde con el que obtenía mejor puntuación en el criterio de evaluación  $R(k)$  en la Figura 4.6, cuyos parámetros iniciales son  $S=4000$ ,  $M=400$ ,  $G=80$ . Las simulaciones muestran que este par de aperturas óptimo obtiene resultados mejores en la estimación del mapa de profundidades que el par óptimo de Zhou *et al.*

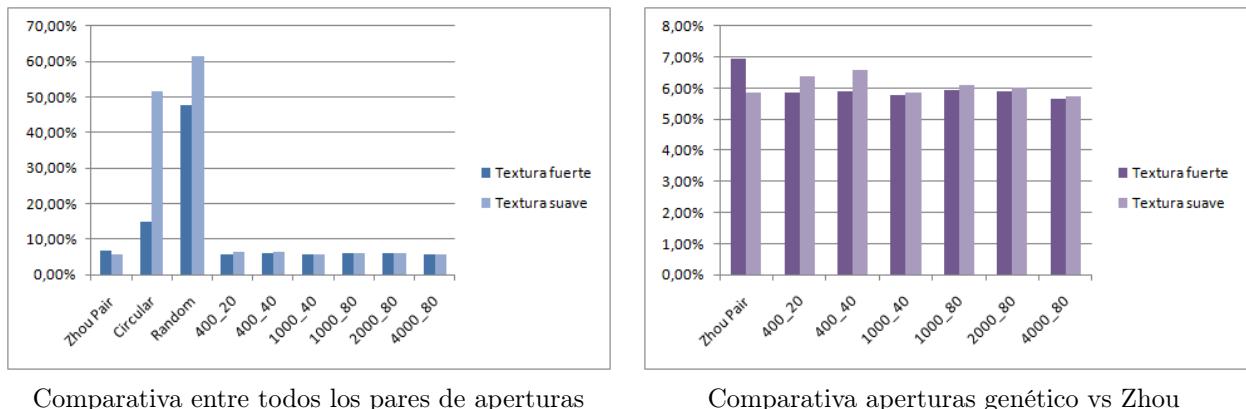


Figura 4.11: Porcentajes de error en el cálculo del mapa de profundidades para cada par de aperturas.

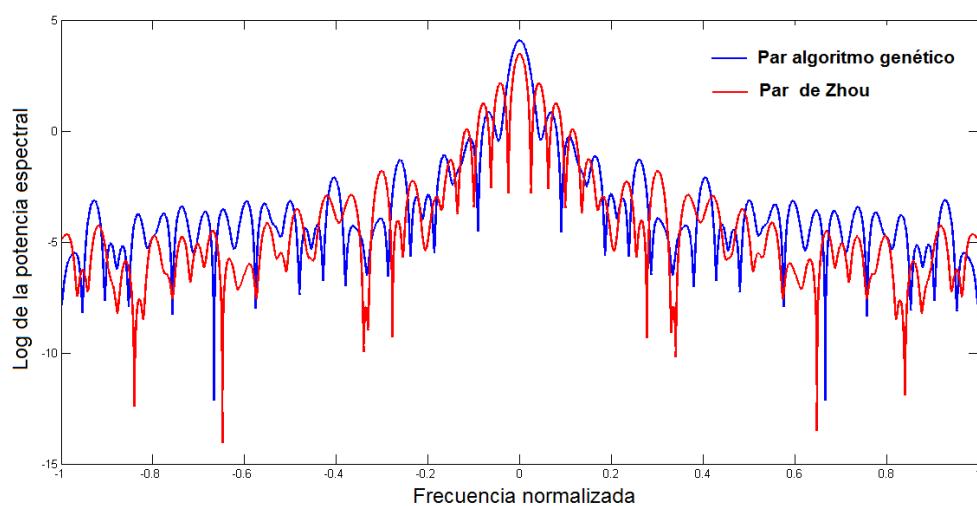


Figura 4.12: Logaritmo de la potencia espectral para el par óptimo de Zhou y nuestro par con mejor evaluación del genético ( $S=4000$ ,  $M=400$ ,  $G=80$ ).

A continuación se muestran los resultados obtenidos en la validación por simulación de estimación del mapa de profundidades para el par circular, el par óptimo de Zhou y nuestro par óptimo, ver Figura 4.13. El mapa de profundidades de la izquierda se corresponde con la parte de la imagen de textura fuerte mientras que el de la derecha es la parte de poca textura. Se ve cómo el uso de aperturas codificadas introduce una mejora muy notable en la estimación del mapa de profundidades respecto de la estimación del par de aperturas circulares. Asimismo vemos que el par óptimo obtenido con nuestro algoritmo genético realiza una mejor estimación de profundidad que el par óptimo de Zhou et al., especialmente en las zonas más alejadas (color granate).

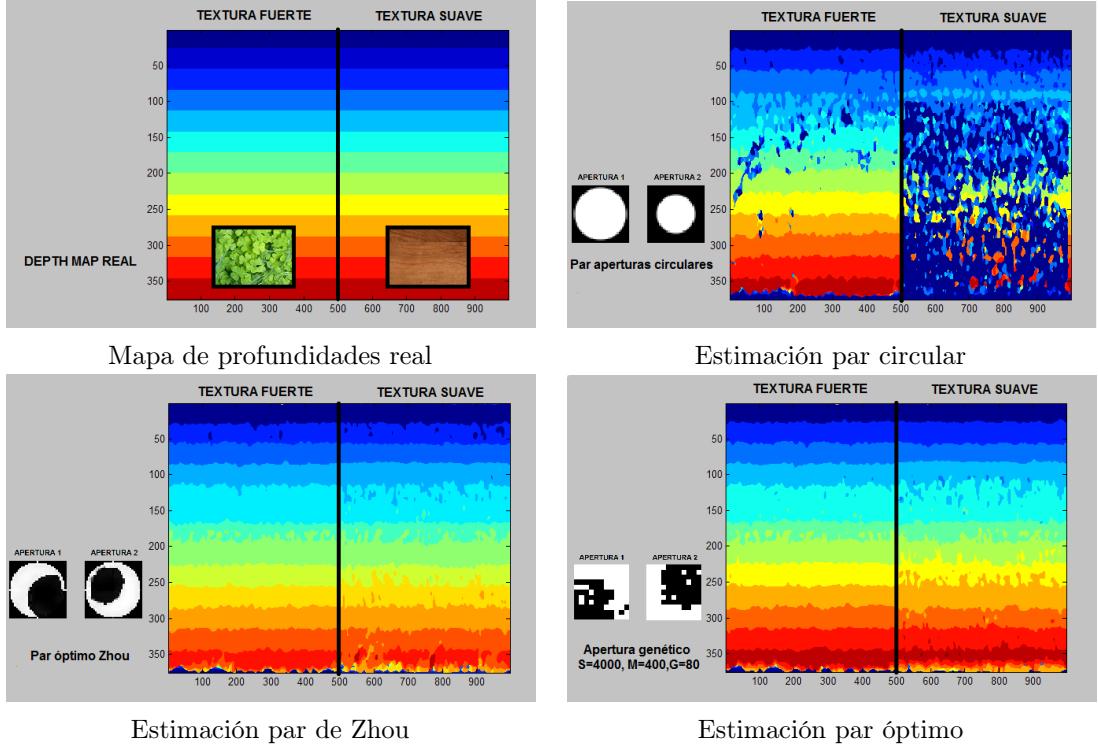


Figura 4.13: Estimación del mapa de profundidades para distintos pares de aperturas.

#### 4.5.3. Resultados del proceso de *deblurring*

La Ecuación 4.5 muestra que la recuperación de una imagen enfocada de alta calidad requiere:

- (a) una estimación precisa de la profundidad  $d$
- (b) que la potencia espectral del par de aperturas presente una respuesta en frecuencia de banda ancha.

Por simplicidad y por mantener separadas las distintas posibles fuentes de error se asume que la profundidad  $d$  es conocida y se evalúa por simulación la recuperación de imágenes enfocadas de las aperturas. Las aperturas utilizadas son: apertura circular, apertura de Levin *et al.*[10], par de aperturas circulares, par óptimo de Zhou *et al.*[23] y el par óptimo obtenido en este trabajo mediante el genético.

Para cada opción de aperturas codificadas se simulan las imágenes desenfocadas de la carta de resolución ISO 12233 mostrada en la Figura 4.14, mediante la convolución de la imagen original enfocada  $f_0$  con la PSF a la distancia  $d$  escogida. Partiendo de la imagen desenfocada (dos imágenes desenfocadas en el caso de pares de aperturas) y usando la deconvolución de Wiener, o la deconvolución generalizada de Wiener en el caso de pares

de aperturas, se recupera la imagen enfocada  $\hat{f}_0$ . En la Figura 4.14 se muestra para cada caso las imágenes desenfocadas y las imágenes recuperadas junto con un recorte a nivel de detalle que permita apreciar mejor los resultados de la deconvolución.

Sobre estos resultados se realizan dos métricas que evalúan la calidad de la imagen recuperada. En primer lugar se aplica la norma L2 (o distancia euclídea) que calcula la distancia euclídea entre píxeles correspondientes de la imagen original  $f_0$  y la imagen recuperada  $\hat{f}_0$ . El valor de píxel de las imágenes se encuentra entre cero (negro) y la unidad (blanco) siendo la distancia máxima entre píxeles igual a la unidad. Por lo tanto la distancia máxima entre dos imágenes es igual al número de píxeles de la imagen  $m \times n$ . Así, la norma L2 puede expresarse como:

$$d(\hat{f}_0, f_0) = \sum_{i=1, j=1}^{n, m} \|\hat{f}_0(i, j) - f_0(i, j)\|^2 \quad (4.11)$$

En el Cuadro 4.1 se pueden ver los resultados obtenidos al aplicar la norma L2 sobre las imágenes recuperadas. Cada imagen es de tamaño 512x512 píxeles, por lo tanto el error máximo según la norma L2 sería de 262.144. Para una visualización más clara se representa en la Figura 4.15 el porcentaje de error según la norma L2 para cada imagen obtenida con los distintos tipos de aperturas.

Cuadro 4.1: Resultados norma L2 para cada tipo de apertura.

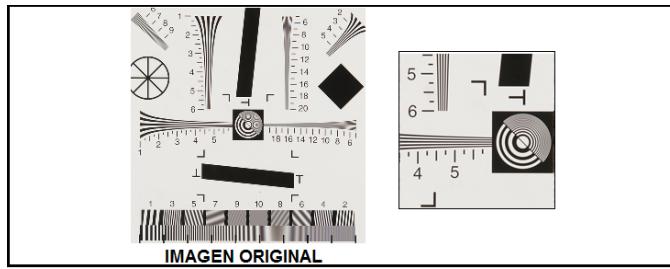
Apertura	Índice L2	Porcentaje de error
Par óptimo Zhou	529,55	0,20 %
Par óptimo genético	814,04	0,31 %
Par de aperturas circulares	1238,50	0,47 %
Apertura circular	2140,70	0,82 %
Apertura de Levin	1867,50	0,71 %

Los resultados de la normal L2 muestran que el par óptimo de Zhou junto con el par óptimo obtenido en el genético recuperan la imagen con menor error. Por otro lado el uso de la apertura convencional o la apertura de Levin no resultan óptimas para el problema de *deblurring*. Esto ya se ha comentado anteriormente, las respuestas en frecuencia de ambas no son óptimas para *deblurring*; además se debe tener en cuenta que la apertura de Levin se diseña bajo un criterio de discriminación de profundidad.

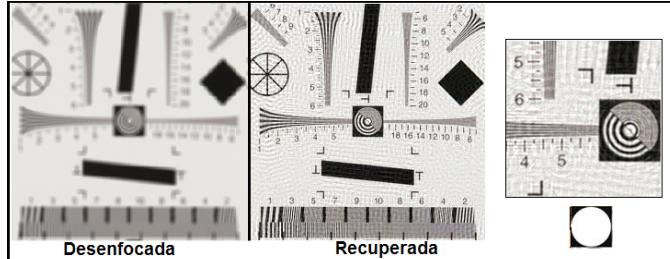
El segundo método empleado para la medida de similitud entre dos imágenes es la métrica SSIM (*Structural similarity index*). Esta métrica [21] se diseñó para mejorar los métodos de medida de calidad en imágenes tradicionales como PSNR (*Peak signal to noise ratio*) o MSE (*Mean squared error*) ya que se ha demostrado que son inconsistentes con la percepción del ojo humano. La métrica SSIM se calcula en varias ventanas de la imagen. La medida de dos ventanas  $x$  e  $y$  del mismo tamaño  $N \times N$  viene dada por:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4.12)$$

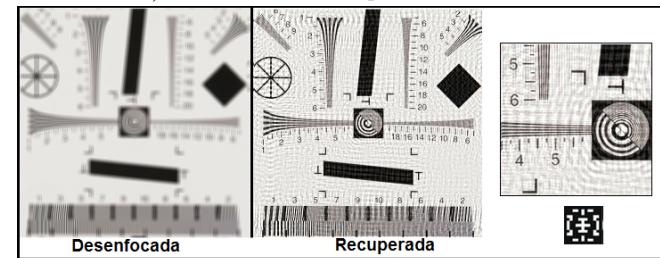
siendo  $\mu_x, \mu_y$  las respectivas medias de las ventanas  $x, y$ ;  $\sigma_x, \sigma_y$  las varianzas y  $\sigma_{xy}$  su covarianza. Los valores  $c_1, c_2$  son dos variables para estabilizar la división, donde  $c_1 = (k_1 L)^2$ ,  $c_2 = (k_2 L)^2$  siendo  $L$  el rango dinámico de los valores de píxel ( $L = 2^{\#bits per pixel} - 1$ ). Además se debe tener en cuenta que,  $-1 \leq SSIM \leq 1$ , donde es igual a la unidad para dos ventanas idénticas e igual a -1 para dos ventanas completamente opuestas. Los resultados de esta métrica mantienen la coherencia con la norma L2, siendo



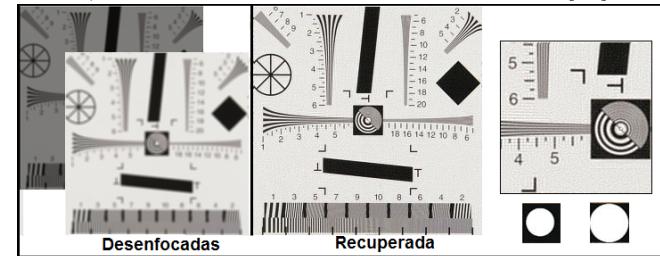
a) Imagen original



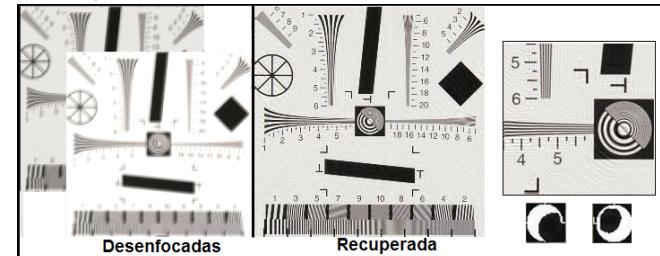
b) Resultados con apertura circular



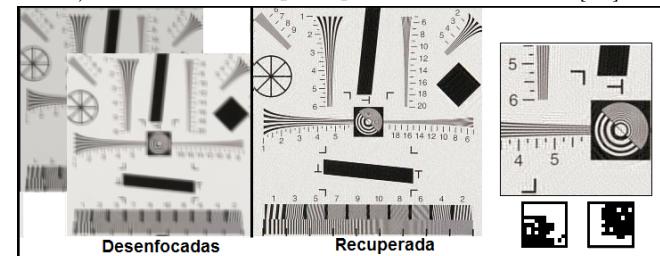
c) Resultados con apertura de Levin et al.[10]



d) Resultados con par de aperturas circulares



e) Resultados con par óptimo de Zhou et al.[23]



f) Resultados con par óptimo obtenido en el genético

Figura 4.14: Simulación de desenfoque y *deblurring* para distintas aperturas y pares de aperturas.

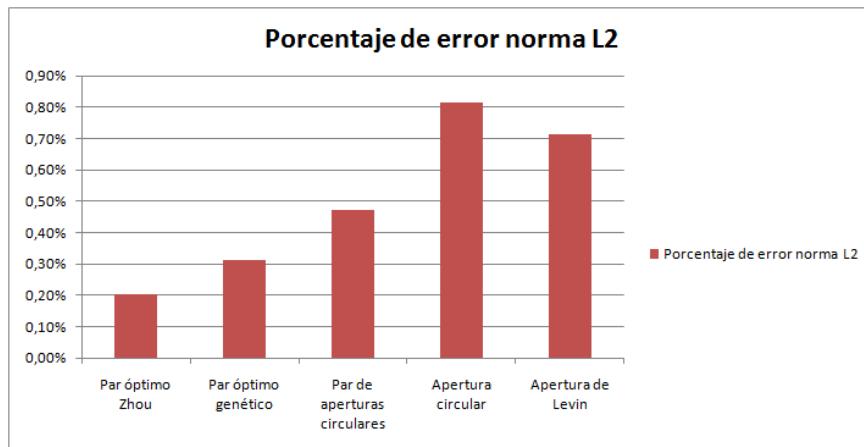


Figura 4.15: Porcentaje de error según la métrica L2 en la recuperación de imagen enfocada para cada tipo de apertura.

nuevamente los dos pares de aperturas codificadas (Zhou et al.[23] y óptima del genético) las dos con mejor valoración.

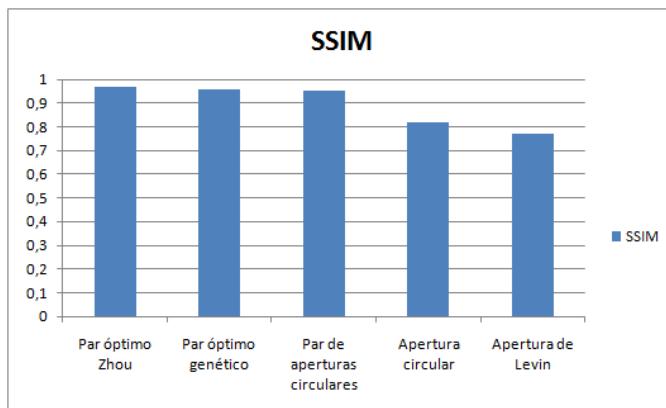


Figura 4.16: Resultados métrica SSIM para cada tipo de aperturas

# Capítulo 5

## Aperturas no binarias para *deblurring*

### 5.1. Introducción

Las aperturas codificadas empleadas en fotografía computacional de los trabajos de investigación más relevantes son binarias, es decir, únicamente consideran valores lógicos de 0 y 1 para la búsqueda del patrón óptimo de la apertura codificada. En la mayoría de los estudios, se indica que se limita el espacio de búsqueda a valores binarios principalmente por el coste computacional. Tal y como hemos visto, el número de posibles soluciones para encontrar una apertura binaria óptima de  $N \times N$  es de  $2^{N \times N}$ , por lo que ampliar la búsqueda incluyendo patrones no binarios implica aumentar el número de posibles soluciones, suponiendo un mayor coste computacional. Sin embargo, a parte de por simplicidad y menor coste computacional, no hay razones para limitar los valores de píxel de la apertura a blancos o negros. Una referencia de estudio en este aspecto es el trabajo de Veeraraghavan *et al.*[20] donde el autor cita ventajas de aperturas con valores continuos sobre aperturas de valores binarios (en su caso para la aplicación concreta de captura de *light fields*). Así pues, la última parte de este proyecto versa sobre la obtención y evaluación de aperturas con valores de píxel no binarios para la aplicación concreta de *deblurring*.

### 5.2. Obtención de aperturas no binarias optimizadas para *deblurring*

El estudio de aperturas no binarias se realiza en el contexto del trabajo de Zhou *et al.*[24] donde se realiza una búsqueda de aperturas binarias optimizadas para *deblurring*. Se parte en este trabajo de la estructura del algoritmo genético empleado para la búsqueda de aperturas binarias y se emplea la misma función de evaluación que da como resultado aperturas óptimas para *deblurring*. La descripción completa del criterio de evaluación empleado y del método de optimización se deja fuera de esta memoria por motivos de extensión y relevancia, pero puede consultarse en el Apéndice E. La función de evaluación de una apertura va a denominarse  $R(k)$ , pero esto no debe inducir a confusión con la función de evaluación descrita en el Capítulo 4. Son métricas diferentes, puesto que el objetivo y las características del problema también son diferentes (una sola apertura optimizada para *deblurring* en este caso, frente a una pareja de aperturas optimizadas para recuperación de profundidad y *deblurring* en el caso del Capítulo 4). Cabe destacar además que en este caso, a diferencia de en el capítulo anterior, se busca minimizar el valor de  $R(k)$ : valores menores de  $R(k)$  implican mejores aperturas. Para

limitar el espacio de búsqueda del algoritmo genético se limitan los posibles valores de los píxeles de la apertura a  $\{0, 0.5, 1\}$ , es decir, píxeles opacos, semitransparentes y transparentes respectivamente.

### 5.3. Análisis de aperturas no binarias

Se ha realizado un estudio sobre la calidad de las aperturas no binarias resultantes y el tiempo computacional para distintas ejecuciones del algoritmo genético, comparando estos resultados con los de aperturas binarias. Para ello se realizan distintas simulaciones variando los parámetros de población inicial  $N$  y del número de generaciones  $G$ , obteniendo siete combinaciones diferentes de estos parámetros. Para cada combinación de  $N$  y  $G$  se realizan tres ejecuciones del algoritmo, mostrando como resultado las medias de las tres ejecuciones. En todas las figuras que se muestran a continuación, el eje X muestra la población inicial  $N$  y el número de generaciones  $G$  de cada set de ejecuciones. El número de aperturas seleccionadas  $M$  es siempre del 10 % de la población inicial, la probabilidad de recombinación  $c1$  es 0.2 y la probabilidad de mutación  $c2$  se establece en 0.05. Todas las aperturas codificadas calculadas para esta comparativa, binarias y no binarias, tienen una resolución de  $7 \times 7$  para limitar el coste computacional del algoritmo.

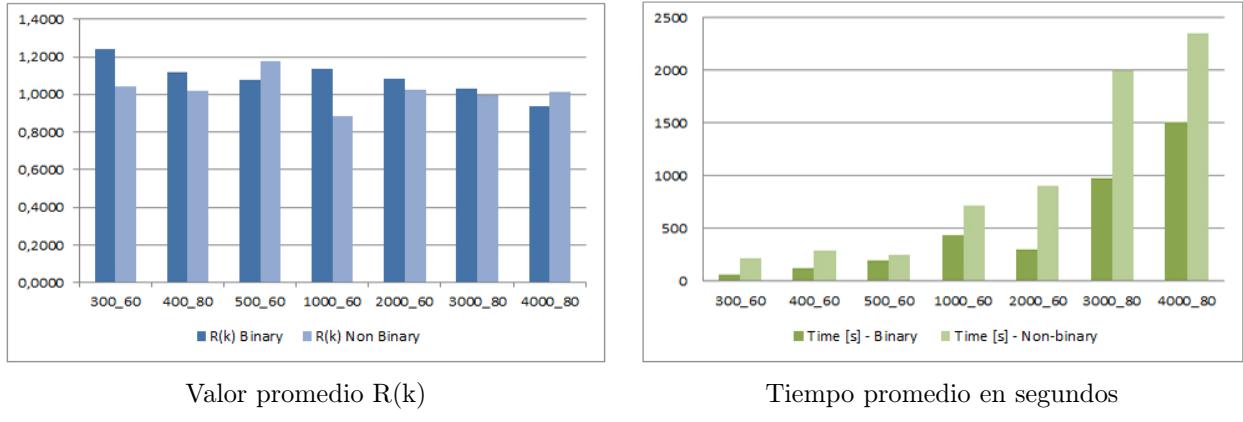


Figura 5.1: Valor promedio de la métrica  $R(k)$  (izquierda) y tiempo promedio en segundos hasta la convergencia del algoritmo (derecha); para aperturas binarias y no binarias y para distintos parámetros iniciales del algoritmo genético.

En la Figura 5.1 podemos ver cómo las aperturas binarias tienden a converger en general a valores ligeramente menores de  $R(k)$ , indicando una mejor respuesta y comportamiento (recordemos que en este caso se seleccionan las aperturas cuyo resultado del criterio de evaluación  $R(k)$  sea menor). Sin embargo, como era de esperar, las aperturas no binarias muestran tiempos de ejecución mayores. El cálculo de los tiempos mostrados se ha realizado ejecutando el algoritmo genético en un Intel Core i7 930 @ 2.80GHz.

Otro análisis realizado en la comparativa de aperturas no binarias vs. no binarias es una validación sintética. Para ello se simula el proceso de captura convolucionando una imagen nítida  $f_0$  con la apertura y añadiendo ruido gaussiano. Posteriormente aplicamos deconvolución de Wiener sobre la imagen desenfocada para recuperar la imagen  $\hat{f}_0$ . Para estimar la calidad de la imagen recuperada se aplica la norma  $L_2$ . Se realiza esto para 10 imágenes y se computa el valor de  $L_2$  promedio. Podemos ver una comparativa de los resultados de error de las aperturas binarias frente a las no binarias en la Figura 5.3, en general las aperturas no binarias se comportan mejor resultando un error menor en la mayoría de los casos. Se debe resaltar que valores bajos de la métrica de calidad de

una apertura  $R(k)$  no conllevan necesariamente a un menor error en la norma  $L_2$ . La explicación de esto es que  $R(k)$  evalúa cómo de óptima es la respuesta de una apertura para el conjunto de imágenes naturales pero esto no significa que resulte óptima para una imagen concreta [23].

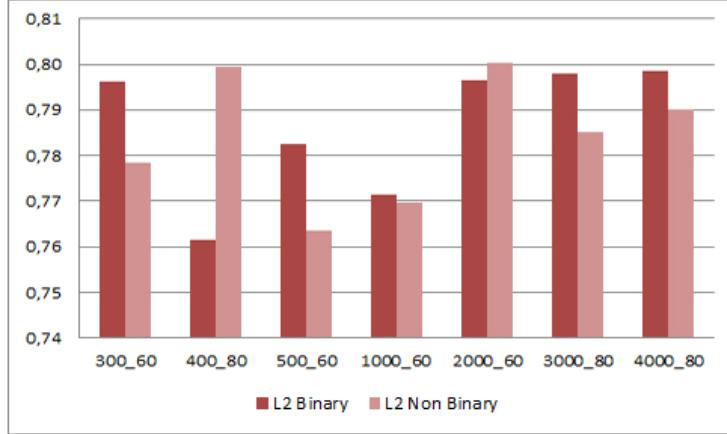


Figura 5.2: Valor promedio de la norma  $L_2$  para aperturas binarias y no binarias y para distintos parámetros del algoritmo genético. Los valores muestran el porcentaje respecto del error máximo.

## 5.4. Resultados y comparativa sobre imágenes reales

Por último se muestran pruebas físicas realizadas con la mejor apertura binaria y la mejor no binaria (ver Figura 5.3). Se capturan varias imágenes a distintas profundidades, se calibran sus PSF a esas profundidades con el array de LEDs siguiendo el procedimiento usado a lo largo del proyecto y explicado en la Sección 3.5.1. La Figura 5.3 muestra imágenes capturadas y recuperadas mediante deconvolución de Wiener para ambas aperturas codificadas. Se puede ver que a pesar de que la deconvolución no recupera completamente la imagen, la apertura binaria se comporta mejor en la recuperación.

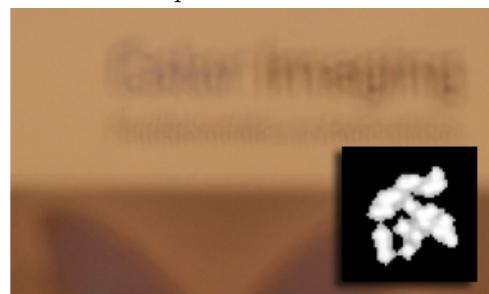
Los resultados obtenidos sobre el uso de aperturas no binarias no son en ningún caso concluyentes. Sería necesario realizar un estudio más completo, ampliando el número de aperturas analizadas, el estudio de convergencia, etc. Pero los primeros indicios muestran que no debe limitarse el espacio de búsqueda de las aperturas codificadas a espacios binarios, ya que la introducción de un tercer valor muestra aparentemente una mejora en los resultados. Por lo tanto queda como trabajo futuro la exploración más a fondo de aperturas codificadas no binarias, donde existan varios niveles de grises.



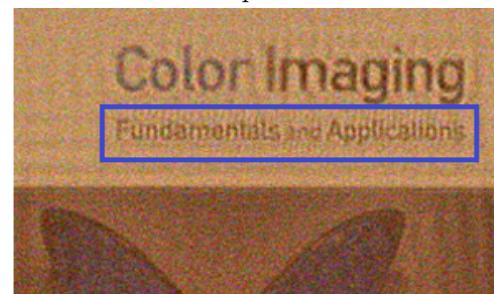
Apertura binaria



Recuperada



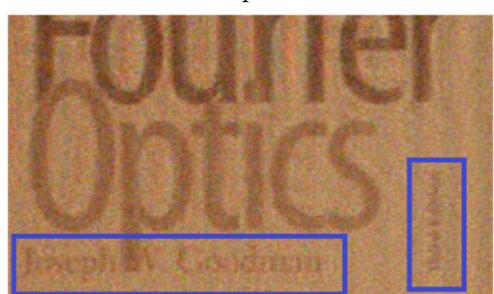
Apertura no binaria



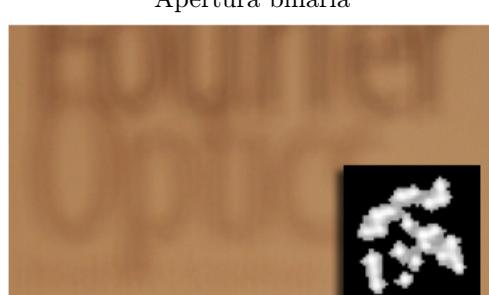
Recuperada



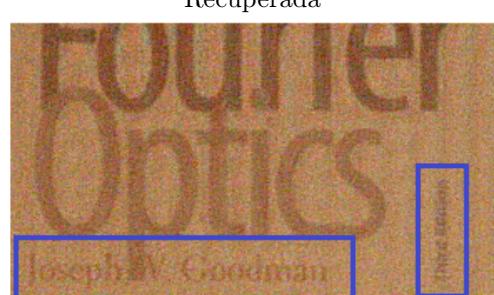
Apertura binaria



Recuperada



Apertura no binaria



Recuperada

Figura 5.3: Apertura binaria y no binaria con mejor resultado del criterio de evaluación del algoritmo genético. Se marcan las zonas recuperadas donde comprueba mejor la calidad superior del resultado de las aperturas no binarias.

# Capítulo 6

## Conclusiones

A continuación se muestran las principales conclusiones derivadas de los resultados de este trabajo. Además, siendo la fotografía computacional un campo de estudio reciente y de amplio contenido, se exponen varias opciones de trabajo futuro partiendo del trabajo realizado en este proyecto. Finalmente se realiza una valoración a nivel personal del proyecto.

### 6.1. Conclusiones del trabajo realizado

Los objetivos principales de este trabajo son tanto la obtención de información de profundidad de la escena capturada, como la recuperación de nítidez en zonas de la escena borrosas debido al desenfoque (*defocus deblurring*). Este doble objetivo se plantea unido, basándonos en la idea de obtener profundidad a partir del grado de desenfoque (DFD, *depth from defocus*). Para ello se ha propuesto como solución el uso de aperturas codificadas que nos permiten codificar ópticamente la luz antes de que ésta alcance el sensor de la cámara. De este modo, tras la posterior decodificación, se puede recuperar más información de la escena que con una apertura convencional.

La primera conclusión clara de este trabajo es la **mejora aportada por el uso de aperturas codificadas**.

- Hemos visto cómo la respuesta en frecuencia de la apertura convencional circular no resulta favorable mientras que el diseño de aperturas codificadas bajo un criterio de selección permite obtener respuestas en frecuencia adecuadas para nuestro objetivo.
- Además se ha visto que existe un compromiso claro entre una respuesta óptima para estimación de profundidad (presencia necesaria de pasos por cero en la respuesta en frecuencia) y una respuesta óptima para *deblurring* (respuesta en frecuencia de banda ancha y plana, sin pasos por cero de modo que el filtro sea invertible).

Por lo tanto si se desea obtener información de profundidad, se debe sacrificar en cierto grado el proceso de *deblurring*, ya que al tener que incluir pasos por cero en la respuesta en frecuencia, ésta dejará de ser óptima para *deblurring*. Aún así, hemos visto que con **el uso de priors de imágenes naturales** se pueden obtener resultados óptimos en la recuperación de enfoque de los objetos de la escena.

En cuanto al proceso de *deblurring*, en este trabajo se introduce el uso de la deconvolución sparse para la recuperación de objetos enfocados de una imagen capturada con una apertura codificada diseñada para DFD. A pesar de que la apertura codificada no es óptima para *deblurring* hemos visto cómo el uso de **la deconvolución sparse permite la recuperación de objetos nítidos** de la escena. Esto se debe a que la deconvolución

sparse concentra las derivadas en un número pequeño de píxeles resultando imágenes con bordes más marcados y reduciendo el ruido y otros artefactos como el *ringing*. Además se ha comprobado que la deconvolución sparse presenta mejores resultados que el resto de métodos de deconvolución, pero **sólo hasta cierto grado de desenfoque**. A partir de cierto diámetro de *blur* la deconvolución sparse no presenta mejoras e incluso se obtienen mejores resultados para el método de deconvolución de Wiener, tal y como se ha demostrado también con una apertura codificada diseñada para *deblurring*.

Del segundo objetivo del proyecto, la **obtención de información de profundidad**, podemos concluir que:

- Resulta en un método costoso para el caso de una única apertura. Se ha visto cómo el vector de pesos necesario para el error de reconstrucción muestra una fuerte dependencia tanto con el sistema como con la imagen.
- Se decide explorar el uso de pares de aperturas codificadas que permiten disponer de un método de estimación de profundidades más robusto, eficiente y aplicable a varios sistemas e imágenes.

El algoritmo genético escogido para la **obtención de pares de aperturas codificadas** ofrece buenos resultados, obteniendo una pareja de aperturas óptima para DFD y demostrando que sus resultados en simulación tanto para *deblurring* como para estimación de profundidad son de calidad comparable a trabajos previos (Zhou et al. [23])

Finalmente se ha explorado la introducción de valores no binarios (semitransparentes) en patrones de aperturas codificadas diseñadas para *deblurring*. Los resultados de este estudio, aunque en ningún caso sean concluyentes, muestran que **las aperturas codificadas no binarias presentan un mejor comportamiento que las binarias**. Por lo tanto se demuestra que no hay razones (excepto el coste computacional) para la limitación de los valores de la apertura a valores binarios (opacos y transparentes).

## 6.2. Trabajo futuro

El uso de aperturas codificadas en fotografía computacional permite un amplio abanico de estudio. Son varias las opciones de trabajo futuro que surgen de este proyecto. En primer lugar se plantea estudiar la estimación de profundidad mediante el uso de una única apertura con otros métodos que no presenten una dependencia tan fuerte con el sistema y la imagen como el estudiado en este trabajo.

Por otro lado, los resultados iniciales del estudio de aperturas no binarias muestran que se debería hacer un estudio más completo y concluyente sobre su mejor comportamiento. Para ello se debería realizar un estudio de convergencia más exhaustivo, y una comparación estructurada y amplia con las aperturas binarias.

Por último, se podría realizar la correspondiente validación por soporte físico del par de aperturas codificadas obtenidas. Para ello, como se ha comentado anteriormente, debería emplearse un sistema que nos permita obtener dos imágenes idénticas (sin que varíe la referencia de la cámara respecto de la escena al cambiar el objetivo) o en su lugar realizar un perfecto ajuste a nivel de píxel de las dos imágenes capturadas.

## 6.3. Conclusiones personales

La realización de este proyecto la valoro a nivel personal como una experiencia muy positiva. En primer lugar el estudio de aperturas codificadas para la fotografía computacional desde el primer momento me pareció un tema muy interesante. La realización de

este Proyecto Fin de Carrera me ha aportado tanto conocimientos teóricos como conocimientos prácticos. Pero principalmente valoro haber experimentado lo que es trabajar en una línea de investigación abierta, encontrarse con problemas reales cuya solución no siempre es sencilla ni directa, y que debe ser enfocada de una manera muy distinta a problemas clásicos más teóricos. Además este proyecto tiene la ventaja de incluir una parte de validación por soporte físico, donde se realizan pruebas reales, siendo siempre gratificante ver resultados aplicados a la realidad del trabajo realizado. Finalmente destacar que el estudio de las aperturas no binarias me ha permitido aportar una sección a un trabajo publicado en el Congreso Ibero-americano de Informática Gráfica (SIACG 2011), que resultó entre las tres mejores publicaciones del congreso.



## Parte II

# Bibliografía



# Bibliografía

- [1] Barrett, R., Berry, M., Chan, T.F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., and Der Vorst, H.V.: Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition. *SIAM, Philadelphia, PA.*(1994)
- [2] Caroli, E., Stephen, J., Cocco, G., Natalucci, L.,and Spizzichino, A.: Coded aperture imaging in X-and Gamma-ray astronomy. *Space Science Reviews*, 349-403. (1987)
- [3] Farid, H., and Simoncelli, E. P.: Range estimation by optical differentiation. *Journal of the Optical Society of America* 15, 1777-1786. (1998)
- [4] Gonzalez, R. C. and Woods, R. E.: Digital Image Processing. Addison-Wesley Publishing Company, Inc. (1992)
- [5] Gottesman, S., and Fenimore, E.: New family of binary arrays for coded aperture imaging. *Applied Optics*, 20, 4344-4352. (1989)
- [6] Grossmann, P.: Depth from focus. *Pattern Recognition Letters* 5, 1(jan.), 63-69. (1987)
- [7] Hasinoff, S. W.: Variable-aperture photography, Graduate Department of Computer Science University of Toronto, (2008)
- [8] Hasinoff, S. W., and Kutulakos, K. N.: Confocal stereo. In *European Conference on Computer Vision*, I: 620-634, (2006)
- [9] Hiura, S., and Matsuyama, T.: Depth measurement by the multi-focus camera. In *CVPR*, IEEE Computer Society, 953-961, (1998)
- [10] Levin, A., Fergus, R., Durand, F., Freeman, W.: Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics* 26, 3 (2007)
- [11] Lucy, L., B.:An iterative technique for the rectification of observed distributions. *Astronomical Journal*, Vol. 79, p. 745 (1974)
- [12] NayarS., Mitsunaga T.:High dynamic range imaging: spatially varying pixel exposures.In *Computer Vision and Pattern Recognition*, vol. 1,pp. 472479. (2000)
- [13] Olshausen, B. A., and Field, D. D.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381 (June),607-609 (1996)
- [14] Pentland, A. P.: A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.* 9, 4, 523-531, (1987)
- [15] Portilla, J., Strela, V., Wainwright, M., and Simoncelli, E. P.: Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing* (2003)

- [16] Premaratne, P.,and Ko, C.C.: *Zero sheet separation of blurred images with symmetrical point spread functions.**Signals, Systems, and Computers*, 1297-1299 (1999)
- [17] Richardson, W., H.: Bayesian-Based Iterative Method of Image Restoration, Vol. 62, Issue 1, pp. 55-59 (1972)
- [18] Roth, S., Black, M.J.: Fields of experts: A framework for learning image priors. In *CVPR* (2005)
- [19] Van der Schaaf, A., and Van Hateren, J.: Modelling the power spectra of natural images: statistics and information. *Vision Research*, 36(17), 2759-2770.(1996)
- [20] Veeraraghavan A., Raskar R., Agrawal A., Mohan A., Tumblin J.: Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*26 (2007)
- [21] Wang Z., Bovik A. C., Sheikh H. R., and Simoncelli E. P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, (2004)
- [22] Weiss, Y.,Freeman, W.T.: What makes a good model of natural images? In *CVPR* (2007)
- [23] Zhou, C., Lin, S., Nayar, S.: Coded aperture pairs for depth from defocus. In *ICCV* (2009)
- [24] Zhou, C., Nayar, S. K.: What are Good Apertures for Defocus Deblurring? In *IEEE International conference on Computational Photography* (2009)

# Parte III

## Apéndices



## Apéndice A

# Deconvolución basada en priors de imágenes naturales

### A.1. Introducción

Dada una imagen borrosa  $y$  y un filtro  $f$  el problema de re-enfoque o *deblurring* puede definirse como encontrar una imagen nítida  $x$  tal que:

$$y = f * x \quad (\text{A.1})$$

Existen varios métodos de uso común para resolver el problema de deconvolución, como el conocido algoritmo de Richardson-Lucy o la deconvolución de Wiener. Dadas las restricciones de estos métodos se propone emplear una técnica alternativa que hace uso de priors de imágenes naturales.

La convolución es un operador lineal y por tanto la Ecuación A.1 puede ser escrita como

$$y = C_f x \quad (\text{A.2})$$

donde  $C_f$  es la matriz de convolución  $N \times N$ , y las imágenes  $x$  e  $y$  están expresadas como vectores de  $N \times 1$  elementos. Expresando la convolución en el dominio frecuencial,  $F(v, w)$ , la transformada de Fourier de  $C_f$ , es una matriz diagonal:

$$Y(v, w) = F(v, w)X(v, w) \quad (\text{A.3})$$

siendo  $X$ ,  $Y$  las transformadas de Fourier respectivas de  $x$ ,  $y$ ; y con  $v$ ,  $w$  las componentes frecuenciales.

Si la matriz  $C_f$  es una matriz de rango completo y no hay ruido en el proceso de captura de la imagen, el problema de la deconvolución se reduciría a invertir  $C_f$  y definir  $x = C_f^{-1}y$ . En el dominio frecuencial  $X(v, w) = Y(v, w)/F(v, w)$ . Pero esto normalmente no es estable debido a las componentes frecuenciales nulas de  $F(v, w)$ . Incluso si  $|F(v, w)|$  no es exactamente cero pero su valor es pequeño, invertirlo lo hace muy sensible al ruido. Esto es, si el proceso de captura de la imagen observada  $Y$  incluye ruido,  $Y(v, w) = F(v, w)X(v, w) + \eta$ , y por lo tanto al querer obtener la imagen nítida  $x$  en el dominio frecuencial, se tiene:

$$\frac{Y(v, w)}{F(v, w)} = X(v, w) + \frac{\eta}{F(v, w)} \quad (\text{A.4})$$

si  $|F(v, w)|$  toma valores pequeños el proceso de deconvolución se ve más afectado por el ruido.

Para mejorar estas dificultades del problema resulta útil introducir un prior en el proceso de deconvolución, y buscar una imagen  $x$  que explique la imagen desenfocada  $y$

pero a la vez sea una buena imagen de acuerdo con el prior de imágenes naturales. Esto es, dada una imagen  $y$  se quiere encontrar el máximo a posteriori de  $x$ :

$$x = \arg \max_x P(x|y) \propto P(y|x)P(x) \quad (\text{A.5})$$

Si se asume que el ruido introducido por el proceso de captura se corresponde con un ruido gaussiano de varianza  $\eta^2$  se puede expresar la probabilidad como:

$$P(y|x) \propto e^{-\frac{1}{2\eta^2} \|x - C_f y\|^2} \quad (\text{A.6})$$

Para el prior se usa un conjunto de filtros  $g_k$  y preferentemente se busca que la respuesta de la imagen a este filtro sea pequeña:

$$P(x) = e^{-\alpha \sum_{i,k} \rho(g_{i,k} * x)} \quad (\text{A.7})$$

donde  $i$  recorre los píxeles de la imagen y  $g_{i,k}$  denota el filtro  $k$ -ésimo centrado en el píxel  $i$  para la función  $\rho$ . La elección más simple para los filtros son los filtros derivativos vertical y horizontal:  $g_x = [1 - 1]$  y  $g_y = [1 - 1]^T$ . Puede ser útil introducir derivadas de segundo orden, o filtros más sofisticados[18, 22]. La función  $\rho$  se selecciona para ser sparse, y en nuestra implementación se ha empleado  $\rho(z) = |z|^{0.8}$ . Sin embargo, por motivos computacionales, se comienza considerando priors Gaussianos, donde  $\rho(z) = |z|^2$ .

Tomando logaritmos de las Ecuaciones A.5, A.6, A.7, el problema de máximo a posteriori para  $x$  se reduce a minimizar:

$$\|y - C_f x\|^2 + w \sum_{i,k} \rho(g_{i,k} * x) \quad (\text{A.8})$$

donde  $w = \alpha\eta^2$ . Minimizando la Ecuación A.8 buscamos la imagen  $x$  que minimice el error de reconstrucción  $\|C_f x - y\|^2$ , donde se requiere que el prior sea lo más suave posible. En las siguientes secciones describimos varios métodos posibles de minimización de la Ecuación A.8 .

## A.2. Deconvolución basada en prior de derivadas gaussianas

Comenzaremos con el caso más sencillo, donde se considera que las derivadas siguen una distribución gaussina, es decir el prior de imagen es gaussiano,  $\rho(z) = |z|^2$ . Derivando la Ecuación A.8 respecto de  $x$  y estableciendo la derivada igual a cero, vemos que la solución óptima puede encontrarse resolviendo una serie de ecuaciones lineales sparse:  $Ax = b$  para

$$A = C_f^T C_f + w \sum_k C_{g_k}^T C_{g_k} \quad b = C_f^T y \quad (\text{A.9})$$

### A.2.1. Deconvolución dominio frecuencial

Resulta más sencillo resolver la Ecuación A.9 en el dominio frecuencial. Expresando la ecuación en el dominio frecuencial se obtiene

$$(|F(v, w)|^2 + w \sum_k |G_k(v, w)|^2) X(v, w) = F(v, w) * Y(v, w) \quad (\text{A.10})$$

y despejando la imagen de interés  $x$

$$X(v, w) = \frac{F(v, w) * Y(v, w)}{|F(v, w)|^2 + w \sum_k |G_k(v, w)|^2} \quad (\text{A.11})$$

Si no se usa ningún prior en el proceso de deconvolución ( $w = 0$ ) la ecuación se reduce a dividir  $Y(v, w)$  entre  $F(v, w)$ . La inclusión de priors hace que la solución  $X(v, w)$  tienda a cero en las componentes frecuenciales para las cuales  $|F(v, w)|$  es pequeña. Se debe resaltar también que la respuesta del filtro derivado es mayor para componentes de alta frecuencia y menor para componentes de baja frecuencia. Por lo tanto, el efecto del prior es más notorio en alta frecuencia, donde el contenido frecuencial de la imagen es menor y por lo tanto mayor la contribución del ruido. El problema de la deconvolución en el dominio de la frecuencia se resuelve eficientemente (tiempo de cómputo de segundos para una imagen de 2 Megapíxeles), el precio a pagar es que se asume que la convolución es cíclica por lo que los bordes de la imagen resultante serán erróneos. Además en la deconvolución en el dominio de la frecuencia aparecen ciertos artefactos en los bordes de los detalles de la imagen.

### A.2.2. Deconvolución dominio espacial

Para evitar asumir una convolución cíclica, se incluye en las matrices de convolución  $C_f, C_{g_k}$  únicamente píxeles válidos, es decir, se eliminan las filas correspondientes a los píxeles de los bordes de la imagen. Por tanto las matrices de convolución tendrán menos filas que columnas, dejando de ser diagonales en el dominio de la frecuencia. La Ecuación A.9 se resuelve numéricamente usando el algoritmo de Gradiente Conjugado [1]. El cuello de botella en cada iteración de este algoritmo es la multiplicación de cada vector residual por la matriz  $A$ . Afortunadamente la forma de  $A$  (Ecuación A.9) permite que se realice eficientemente como una concatenación de operaciones de convolución.

## A.3. Deconvolución basada en prior de derivadas sparse

La ventaja del uso de un prior de imagen Gaussiano es que el problema de optimización es convexo y la solución óptima puede ser derivada de manera sencilla como se mostraba en la Ecuación A.9. Sin embargo la realidad es que la distribución de los filtros derivativos en las imágenes naturales es sparse y no gaussiana. Las ventajas aportadas por el uso de priors sparse en aplicaciones de procesado de imágenes han sido demostradas en recientes estudios y publicaciones [15, 18, 22]. La diferencia principal es que mientras que un prior gaussiano distribuye las derivadas uniformemente sobre la imagen, el prior sparse concentra las derivadas en un número pequeño de píxeles, manteniendo la mayor parte de los píxeles constantes. El resultado son imágenes con bordes más marcados, reduciéndose además el ruido y permitiendo eliminar artefactos no deseados en la imagen como el *ringing*. La desventaja del uso del prior sparse, es que el problema de optimización deja de ser convexo. Por lo tanto para optimizarlo se usa un método iterativo reponderado de mínimos cuadrados (IRLS, Iterative re-weighted least squares).

En el método IRLS:

$$\sum_j \rho(A_{j \rightarrow} x - b_j) \quad (\text{A.12})$$

Expresando el problema de *deblurring* adecuadamente para la aplicación de IRLS, se establecen los vectores fila  $A_{j \rightarrow}$  como las filas de las matrices de convolución  $C_{g_k}, C_f$ . El método IRLS establece el problema como una secuencia de problemas de mínimos cuadrados., cada problema de mínimos cuadrados se responderá con la solución en el paso previo. La minimización de cada problema de mínimos cuadrados es equivalente a resolver un conjunto sparse de ecuaciones lineales.

Los pasos del algoritmo IRLS son:

- Inicialización: establecer  $\Psi_j^0 = 1$

- Repetir hasta que converja:

-Si  $\bar{A} = \sum_j A_{j\rightarrow}^T \Psi_j^{t-1} A_{j\rightarrow}$  y  $\bar{b} = \sum_j A_{j\rightarrow}^T \Psi_j^{t-1} b_j$ .  
 $x^t$  es la solución para  $\bar{A}x = \bar{b}$ .

-Establecer  $u_j = A_{j\rightarrow}x^t - b_j$  y  $\Psi_j^t(u_j) = \frac{1}{u_j} \frac{d\rho(u_j)}{du}$

El prior usado en la implementación es de la forma  $\rho(u_j) = |u_j|^{0,8}$ . Por lo tanto, el término de reponderación se reduce a:

$$\Psi(u_j) = \max(|u_j|, \xi)^{0,8-2} \quad (\text{A.13})$$

donde  $|u_j|$  se reemplaza con  $\max(|u_j|, \xi)$  evitándose así la división por cero.

Se debe resaltar que inicializar IRLS con  $\Psi_j^0 = 1$  es equivalente a inicializarlo con la solución del prior Gaussiano. Los pesos de reponderación  $\Psi(u_j)$  serán de gran peso en zonas de la imagen donde la derivada sea pequeña y de peso pequeño donde la derivada sea alta. Esto es, el efecto del proceso de reponderación es disminuir el efecto en zonas suaves de la imagen y aumentar la penalización de la derivada en los bordes de la imagen.

Como en este proceso la derivada en cada píxel se repondrá de distinta manera, esto implica que la Ecuación A.9 deja de ser una convolución y no puede ser resuelta en el dominio frecuencial, por lo tanto se está forzado a trabajar en el dominio espacial usando el algoritmo de Gradiente Conjugado. Se observa que la reponderación hace el sistema menos homogéneo, y ralentiza la convergencia de la solución numérica. Para el kernel de la apertura codificada, resolver el sistema de mínimos cuadrados reponderado requiere más Gradiente Conjugado que resolver el sistema no reponderado de la sección A.2.2. Experimentalmente, se ve que la calidad de los resultados obtenidos depende del numero de iteraciones del gradiente conjugado usadas.

## Apéndice B

# Modelo estadístico de imágenes

Las imágenes del mundo real tienen estadísticas muy distintas a matrices aleatorias de ruido blanco. Una propiedad estadística muy conocida de las imágenes es que poseen una distribución de derivadas sparse [13]. Se impondrá esta condición en la reconstrucción de la imagen. Sin embargo, en el diseño del filtro se asume por simplicidad una distribución Gaussiana. Esto es, el prior asume que las derivadas en la imagen nítida real  $x$  siguen una distribución Gaussiana de media cero.

$$P(x) \propto \prod_{i,j} e^{-\frac{1}{2}\alpha((x(i,j)-x(i+1,j))^2 + (x(i,j)-x(i,j+1))^2)} = N(0, \Psi) \quad (\text{B.1})$$

donde  $i, j$  indican los índices de los píxeles.  $\Psi^{-1} = \infty(C_{g_x}^T C_{g_x} + C_{g_y}^T C_{g_y})$  siendo  $C_{g_x}, C_{g_y}$  las matrices de convolución correspondientes a los filtros derivativos  $g_x = [1 - 1]$  y  $g_y = [1 - 1]^T$ . La constante  $\alpha$  se establece para que la varianza de la distribución coincida con la varianza de las derivadas en imágenes naturales ( $\alpha=250$ ). Expresando el prior en el dominio frecuencial se obtiene

$$P(X)\alpha e^{-\frac{1}{2}\alpha X^T \bar{\Psi}^{-1} X} \quad \text{donde } \bar{\Psi}^{-1} = \alpha \text{ diag}(|G_x(v, w)|^2 + |G_y(v, w)|^2) \quad (\text{B.2})$$

siendo  $v, w$  las variables del dominio frecuencial. Asumiendo una profundidad de escena constante  $k$ , la imagen observada desenfocada y con ruido puede modelarse como  $y = f_k * x + n$ . El ruido en los píxeles vecinos se asume independiente, siguiendo un modelo Gaussiano  $n \sim N(0, \eta^2 I)$  (se toma  $\eta = 0,005$ ).

Se define  $P_k(y)$  como la distribución de señales observadas bajo un grado de blur dado por  $f_k$ , es decir la distribución de imágenes que representan objetos a una profundidad  $D_k$ . El filtro de blur  $f_k$  transforma linealmente la distribución de la imagen nítida expresada en la Ecuación B.1, por lo tanto  $P_k(y)$  sigue siendo Gaussiana<sup>1</sup>  $P_k(y) \sim N(0, \sum_k)$ . Siendo la matriz de covarianza  $\sum_k$  una versión transformada del prior de la covarianza más ruido.

$$\sum_k = C_{f_k} \Psi C_{f_k}^T + \eta^2 I \xrightarrow[\text{de Fourier}]{\text{Transformada}} \bar{\sum}_k = C_{F_k} \bar{\Psi} C_{F_k}^T + \eta^2 I \quad (\text{B.3})$$

donde en el dominio frecuencial el prior es diagonal<sup>2</sup>.

De este modo la distribución de la imagen desenfocada en el dominio frecuencial resulta

$$P_k(Y) \propto \exp(-\frac{1}{2}E_k(Y)) = \exp(-\frac{1}{2}\sum_{v,w} |Y(v, w)|^2 / \sigma(v, w)) \quad (\text{B.4})$$

---

<sup>1</sup>Si  $X, Y$  son variables aleatorias y  $A$  una transformación lineal siendo  $X$  Gaussiano e  $Y = AX$ , entonces  $\text{Cov}(Y) = AC\text{ov}(X)A^T$ .

<sup>2</sup>Ya que (i) la transformada de Fourier de una matriz de convolución es una matriz diagonal y (ii) todas las matrices que forman  $\sum_k$  son o diagonales o matrices de convolución.

donde  $\sigma(v, w)$  son los valores de las diagonales de  $\bar{\Sigma}_k$ :

$$\sigma(v, w) = |F_k(v, w)|^2(\alpha|G_x(v, w)|^2 + \alpha|G_y(v, w)|^2)^{-1} + \eta^2 \quad (\text{B.5})$$

La Ecuacion B.5 representa la teoría expuesta anteriormente en la explicación de la Figura 3.2 sobre cómo los pasos por cero de la imagen (y por lo tanto del filtro correcto) permiten distinguir la escala correcta del filtro. Aquí vemos que si el filtro  $f_k$  tiene un nulo en la frecuencia  $(v, w)$  entonces  $\sigma(v, w) = \eta^2$ , que normalmente será un valor muy pequeño. Esto es, si el contenido frecuencial de la imagen observada  $Y(v, w)$  es significativamente mayor que cero, la probabilidad de que  $Y(v, w)$  se corresponda con la distribución  $P_k$  es muy pequeña. En otras palabras, si la imagen observada tiene contenido frecuencial a la misma frecuencia donde el filtro tiene un paso por cero, es muy improbable que sea la escala correcta  $k$  del filtro de blur.

## Apéndice C

# Criterio de evaluación par de aperturas

### C.1. Formulación de DFD (*Depth from defocus*)

Para un objeto plano frontal, su imagen desenfocada puede expresarse como

$$f = f_0 * k + \eta \quad (\text{C.1})$$

donde  $f_0$  es la imagen enfocada,  $\eta$  es el ruido de la imagen asumido ruido blanco gaussiano de media cero y varianza  $\sigma^2$ ,  $N(0, \sigma^2)$ , y  $k^d$  es la PSF (*point-spread-function*) cuyo patrón estará determinado por la apertura y cuyo tamaño  $d$  vendrá relacionado con la profundidad. De este modo estimando el tamaño del *blur* de desenfoque podremos calcular la profundidad.

La Ecuación C.1 puede expresarse en el dominio de la frecuencia como  $F = F_0 \cdot K^d + \zeta$ , donde  $F_0$ ,  $K^d$ ,  $\zeta$  son las transformadas discretas de Fourier de  $f$ ,  $f_0$ ,  $\eta$ , respectivamente. Una única imagen desenfocada es generalmente insuficiente para estimar la profundidad de la escena sin el uso de información adicional (basándonos en priors de imágenes o intervención del usuario [10]). Por ello haremos uso de dos imágenes  $F_i, i = 1, 2$  de una misma escena con distintas características de desenfoque, distintas PSFs para cada imagen.

$$F_i = F_0 \cdot K_i^{d^*} + \zeta_i \quad (\text{C.2})$$

donde  $K_i^{d^*}$  denota la transformada de Fourier de la  $i$ -ésima PSF con el tamaño de blur actual  $d^*$ . Nuestro objetivo es encontrar el tamaño  $\hat{d}$  y la imagen enfocada  $\hat{F}_0$  resolviendo un problema de máximo a posteriori (MAP)

$$\langle \hat{d}, \hat{F}_0 \rangle \propto \arg \max P(F_1, F_2 | \hat{d}, \hat{F}_0, \sigma) P(\hat{d}, \hat{F}_0) = \arg \max P(F_1, F_2 | \hat{d}, \hat{F}_0, \sigma) P(\hat{F}_0). \quad (\text{C.3})$$

Teniendo en cuenta la Ecuación C.2 se obtiene

$$P(F_1, F_2 | \hat{d}, \hat{F}_0, \sigma) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1,2} \| \hat{F}_0 \cdot K_i^{\hat{d}} - F_i \|^2 \right\}. \quad (\text{C.4})$$

El término del prior de imagen  $P(\hat{F}_0)$  puede ser expresado como  $P(\hat{F}_0) \propto \exp\{-\frac{1}{2}\|\phi(\hat{F}_0)\|^2\}$ , donde la función potencial  $\phi$  varía con la elección de los priors de imagen. Se escoge una función potencial lineal  $\phi(\hat{F}_0) = \psi \cdot \hat{F}_0$  que haga posible la discusión analítica posterior del problema, de este modo se tiene

$$P(\hat{F}_0) \propto \{ -\frac{1}{2} \|\psi \cdot \hat{F}_0\|^2 \}, \quad (\text{C.5})$$

donde  $\psi$  es la matriz de pesos. Distintas elecciones de  $\psi$  implican el uso de distintos priors de imagen, por ejemplo, si  $\psi$  es el filtro derivado en el dominio de Fourier, se está escogiendo el conocido prior de imágenes gaussiano.

La estimación del tamaño de *blur* se calcula mediante la  $\hat{d}$  que maximiza:

$$P(\hat{d}|F_1 F_2, \sigma) = \max_{\hat{F}_0} P(\hat{F}_0, \hat{d}|F_1, F_2, \sigma). \quad (\text{C.6})$$

Expresado como una función de energía logarítmica, el problema se convierte en la minimización de

$$E(\hat{d}|F_1, F_2, \sigma) = \min_{F_0} \sum_{i=1,2} \|\hat{F}_0 \cdot K_i^{\hat{d}} - F_i\|^2 + \|C \cdot \hat{F}_0\|^2, \quad (\text{C.7})$$

donde  $C = \sigma \cdot \psi$ . En vez de asignar una matriz de pesos específica  $C$ , se puede optimizar  $C$  haciendo uso de la ley 1/f de imágenes naturales [19, 22]; y definir  $|C|^2 = \sigma^2/A$ , donde  $|A^2|$  se corresponde con la potencia espectral esperada en las imágenes naturales:  $|A(\xi)|^2 = \int_{F_0} |F_0(\xi)|^2 \mu(F_0)$ ; siendo  $\xi$  la frecuencia y  $\mu(F_0)$  la posibilidad de la medida de la muestra  $F_0$  en el espacio de imágenes.

## C.2. Criterio de selección

Basándonos en la formulación del problema de *DFD*, *depth from defocus*, se busca un criterio para la selección de una pareja de aperturas cuya respuesta conjunta sea óptima para el objetivo de nuestro problema.

Para ello se deriva  $E(d|K_1^{d*}, K_2^{d*}, \sigma, F_0)$  sustituyendo C.2 y 4.5 en C.7. La distancia  $d$  estimada está relacionada con  $F_0$  y el nivel de ruido  $\sigma$ . Se puede integrar a lo largo de  $F_0$  haciendo uso de la ley 1/f de imágenes naturales:

$$E(d|K_1^{d*}, K_2^{d*}, \sigma) = \int_{F_0} E(d|K_1^{d*}, K_2^{d*}, \sigma, F_0) \mu(F_0), \quad (\text{C.8})$$

ésta puede ser reordenada y simplificada en

$$E(d|K_1^{d*}, K_2^{d*}, \sigma) = \sum_{\xi} \frac{A \cdot |K_1^d \cdot K_2^{d*} - K_2^d \cdot K_1^{d*}|^2}{\sum_i |K_i^d|^2 + C^2} + \sigma^2 \cdot \sum_{\xi} \left[ \frac{C^2}{\sum_i |K_i^d|^2 + C^2} + 1 \right], \quad (\text{C.9})$$

que expresa la energía correspondiente a una profundidad estimada  $d$ , dado el par de aperturas, el plano focal y el nivel de ruido. El primer término de la Ecuación C.9 mide la inconsistencia entre las dos imágenes desenfocadas cuando el tamaño de blur estimado  $d$  se aleja del tamaño de blur real  $d^*$ . El segundo término hace referencia al ruido de la imagen.

Por lo tanto la profundidad puede ser estimada con mayor precisión y certeza si  $E(d|K_1^{d*}, K_2^{d*}, \sigma)$  crece significativamente cuando el tamaño de blur estimado  $d$  se desvía del real  $d^*$ . Para asegurar esto, se evalúa el par de aperturas  $(K_1, K_2)$  en la distancia  $d^*$  y con un nivel de ruido  $\sigma$  usando

$$R(K_1, K_2|d^*, \sigma) = \min_{d \in D/d^*} E(d|K_1^{d*}, K_2^{d*}, \sigma) - E(d^*|K_1^{d*}, K_2^{d*}, \sigma)$$

$$= \min_{d \in D/d^*} \sum_{\xi} A \frac{|K_1^d K_2^{d^*} - K_2^d K_1^{d^*}|^2}{\sum_i |K_i^d|^2 + C^2} + \frac{\sigma^4}{A} \cdot \frac{\sum_i |K_i^{d^*}|^2 - \sum_i |K_i^d|^2}{(\sum_i |K_i^d|^2 + C^2) \cdot \sum_i |K_i^{d^*}|^2 + C^2} \quad (\text{C.10})$$

$$\approx \min_{d \in D/d^*} \sum_{\xi} A \cdot \frac{|K_1^d K_2^{d^*} - K_2^d K_1^{d^*}|^2}{|K_1^d|^2 + |K_2^d|^2 + C^2}, \quad (\text{C.11})$$

donde  $D = \{c_1 d^*, c_2 d^*, \dots, c_l d^*\}$  es un conjunto de valores de distintos tamaños de blur. En nuestra implementación, se toma un conjunto de valores suficientemente amplio y que a su vez su tiempo de cálculo sea aceptable, tomando  $\{c_i\} = \{0, 1, 0, 2, \dots, 1, 5\}$ .

De este modo el criterio de evaluación de aperturas es dependiente del tamaño de blur real  $d^*$  (distancia del objeto) y del nivel de ruido  $\sigma$ . Sin embargo como en la Ecuación C.10 el primer término domina sobre el segundo y  $C$  se puede despreciar respecto del resto de los factores; la Ecuación C.10 puede ser aproximada por C.11; por lo tanto la dependencia con  $\sigma$  desaparece en el criterio de evaluación (se considerará un valor de  $\sigma$  de 0.005).

Normalizando C.11 se obtiene

$$R(K_1, K_2 | d^*) \approx \min_{d \in D/d^*} \left[ \frac{1}{n} \sum_{\xi} A \cdot \frac{|K_1^d K_2^{d^*} - K_2^d K_1^{d^*}|^2}{|K_1^d|^2 + |K_2^d|^2 + C^2} \right]^{1/2}, \quad (\text{C.12})$$

siendo  $n$  el número de píxeles de la PSF. Tomando

$$M(K_1, K_2, d, d^*) = \left[ \frac{1}{n} \sum_{\xi} A \cdot \frac{|K_1^d K_2^{d^*} - K_2^d K_1^{d^*}|^2}{|K_1^d|^2 + |K_2^d|^2 + C^2} \right]^{1/2}, \quad (\text{C.13})$$

resulta

$$R = \min_{d \in D/d^*} M(K_1, K_2, d, d^*). \quad (\text{C.14})$$

El criterio de evaluación del par de aperturas usado en el algoritmo genético será el de la Ecuación C.14. Se tomará como  $R$  el mínimo valor de la función  $M$  evaluada para los distintos tamaños de blur  $d \in D/d^*$ . De entre los valores de  $R$  se tomará el mayor valor, ya que un valor mayor en  $R$  indica que la función de energía para DFD tiene mayor pendiente y por lo tanto la estimación será más robusta frente al ruido de la imagen y zonas sin textura.



## Apéndice D

# Cómo insertar una apertura codificada en una lente Canon EF 50mm f1.8 II

En este apéndice se muestra cómo se han introducido los patrones de aperturas codificadas en el objetivo de la cámara, pudiendo así modificar la apertura circular convencional.

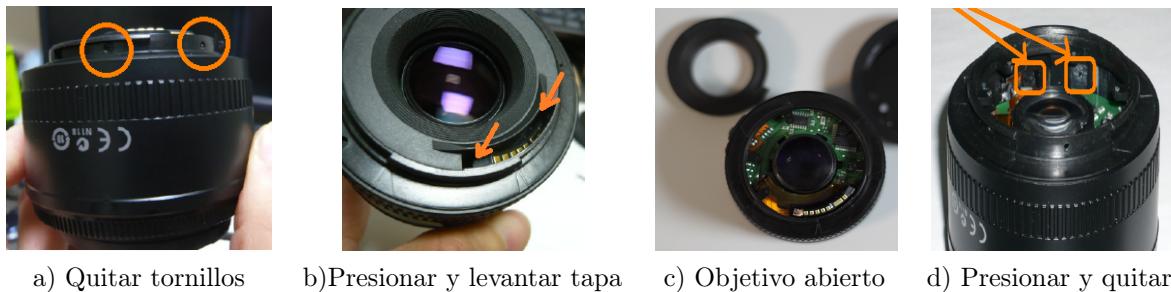
El objetivo empleado a lo largo de todo el proyecto es un Canon EF 50mm f1.8 II, usado sobre el cuerpo de una cámara reflex digital Canon EOS 500D, ver Figura D.1.

La elección de este objetivo se debe tanto a su precio asequible (se debe considerar el riesgo de estropear el objetivo al desmontarlo) como a su estructura simple formada por un pequeño número de partes y de material plástico, facilitando su apertura y montaje posterior.



Figura D.1: Canon EOS 500D con objetivo Canon 50mm f1.8 II.

A continuación se muestran los pasos a seguir para desmontar el objetivo e introducir la apertura codificada, ver Figura D.2. En primer lugar se quita la tapa trasera de protección del objetivo, quedando al descubierto los tornillos mostrados en D.2.a. Se quitan estos tornillos y se presiona en la zona señalada en la imagen D.2.b para después hacer saltar la tapa de plástico que cubre el objetivo. La imagen D.2.c muestra el objetivo abierto, ahora se deberá prestar especial atención a la pestaña que selecciona el modo manual o automático. Seleccionamos el modo manual, que nos permitirá girar el disco de enfoque, el objetivo descenderá y veremos los huecos señalados en la imagen D.2.d.



a) Quitar tornillos      b)Presionar y levantar tapa      c) Objetivo abierto      d) Presionar y quitar

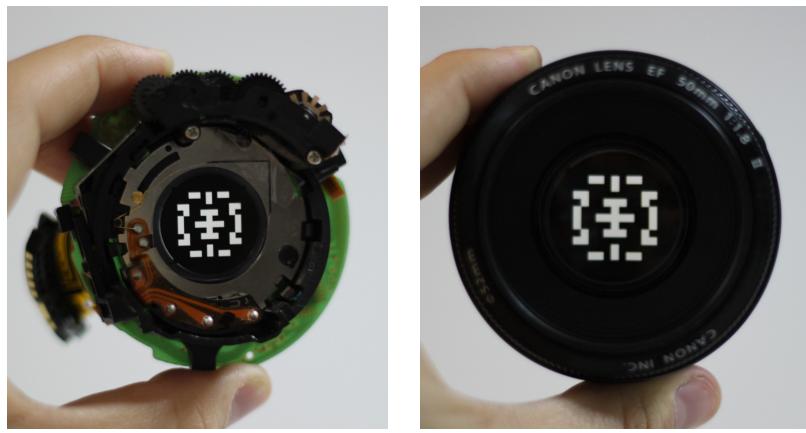
Figura D.2: Primeros pasos para desmontar el objetivo.

Se presiona en ambos lados, soltando así la parte de la carcasa que contiene la pestaña de modo manual y automático. De este modo se obtiene la situación mostrada en la Figura D.3.a. Por último quitamos el tornillo que nos permitirá abrir por completo el objetivo, separando esa pieza en tres. El conjunto de elementos que componen el objetivo se muestran en la Figura D.3.b.



a) Objetivo parcialmente desmontado      b)Objetivo completamente desmontado

Figura D.3: Componentes de la lente Canon EF 50mm f1.8



Colocada sobre la lente

Apertura codifica en objetivo

Figura D.4: Canon EF 50mm f1.8 II con apertura codificada.

En la FiguraD.4 se ve en la imagen de la izquierda la colocación del patrón de apertura sobre la lente del objetivo. El patrón de la apertura se imprime a alta resolución en

papel fotolitográfico, debiéndose corresponder el diámetro de la apertura impresa con el de la lente, en nuestro caso 20mm, para asegurarnos de que la apertura circular queda completamente modificada por el patrón escogido. De este modo se fija la apertura sobre la lente, fijando con cuidado una pestaña de la apertura en la zona que rodea la lente. Se vuelve a montar el objetivo. El resultado se ve en la imagen de la derecha, el objetivo completo con la apertura codificada insertada sobre la lente.



## Apéndice E

### Artículo: Coded Apertures for Defocus Deblurring

El estudio realizado sobre aperturas no binarias se incluyó en una publicación aceptada en el Congreso Ibero-americano de Informática Gráfica (SIACG 2011), resultando entre las tres mejores publicaciones del congreso. En esta publicación se realiza un estudio sobre aperturas codificadas óptimas para *deblurring*, además se incluye el citado trabajo sobre la introducción de valores no binarios en los patrones de aperturas codificadas. A continuación se adjunta el artículo citado.



# Coded Apertures for Defocus Deblurring

Belen Masia, Adrian Corrales, Lara Presa and Diego Gutierrez

Universidad de Zaragoza

---

## Abstract

The field of computational photography, and in particular the design and implementation of coded apertures, has yielded impressive results in the last years. Among their applications lies defocus deblurring, in which we focus in this paper. Following the approach of previous works, we obtain near-optimal coded apertures using a genetic algorithm and an existing quality metric. We perform both synthetic and real experiments, testing the performance of the apertures along the dimensions of depth, size and shape. We additionally explore non-binary apertures, usually overlooked in the literature, and perform a comparative analysis with their binary counterparts.

Categories and Subject Descriptors (according to ACM CCS): I.4.3 [Image Processing and Computer Vision]: Enhancement—Sharpening and deblurring

---

## 1. Introduction

In the past few years, the field of computational photography has yielded spectacular advances in the imaging process. The main idea is to code the light information in novel ways before it reaches the sensor, in order to decode it later and obtain an improved, enhanced or extended representation of the scene being captured. Several different strategies exist, from structured lighting, to new optical devices, to modulated apertures or shutters. In this work we focus on *coded apertures*. These are masks obtained by means of computational algorithms which, placed at the camera lens, encode the defocus blur in order to better preserve high frequencies in the original image. They can be seen as an array of multiple ideal pinhole apertures (with infinite depth and no chromatic aberration), whose location on the 2D mask is determined computationally. Decoding the overlap of all pinhole images yields the final image.

Some existing works interpret the resulting coded blur attempting to recover *depth from defocus*. Given the nature of the blur as explained by simple geometrical optics, this approach imposes a multi-layered representation of the scene being depicted. While there is plenty of interesting on-going research in that direction, in this paper we limit ourselves to the problem of *defocus deblurring*: we aim to obtain good coded apertures that allow us to recover a sharp image from its blurred original version. We follow standard approaches and pose the imaging process as a convolution between the original scene being captured and the blur kernel (plus a

noise function). In principle, this would lead to a blind deconvolution problem, given that the such blur kernel is usually not known. Assuming no motion blur nor camera shake, this kernel is reduced to the point spread function of the optical system. Traditional circular apertures, however, have a very poor response in the frequency domain: not only do they lose energy at high frequencies, but they exhibit multiple zero-crossings as well; it is thus impossible to recover information at such frequencies during deconvolution.

In this paper, we present several coded apertures with better frequency response, which allow us to recover information apparently lost to blur during the capture process. We follow the approach of previous works, and rely on the average power spectra of natural images to guide our optimization process, which is in turn performed by means of genetic algorithms. Once the coded apertures have been obtained, we show the feasibility of our results by printing them out on a photomask sheet and inserting them in an off-the-shelf camera. The captured blurred images are then deconvolved using Wiener deconvolution. We analyze the performance of our apertures as a function of shape, depth and size. We additionally modify our genetic algorithm to allow for non-binary masks, and perform a comparative analysis with their binary counterparts.

## 2. Previous Work

Coded apertures have been traditionally used in astronomy, coding the direction of incoming rays as an alternative to fo-

cusing imaging techniques [IIZ92]. Possibly the most popular patterns were the MURA patterns (Modified Uniformly Redundant Array) [GF89]. Veeraraghavan et al. [VRA<sup>\*</sup>07] showed how a 4D light field can be reconstructed from 2D sensor information by means of a coded mask. Placed at the lens, the authors achieve refocusing of images at full resolution, provided the scene being captured contains only Lambertian objects. Nayar and Mitsunaga [NM00], extended the dynamic range capabilities of the imaging system by placing a mask of spatially varying transmittance next to the sensor, and then mapping the captured information to high dynamic range.

Other works have proposed different coded apertures for defocus deblurring or depth approximation. To restore a blurred image, the apertures are designed to have a broadband frequency response, along with none (or distinguishable) zero-crossings in the Fourier domain. Hiura and Matsuyama [HM98] proposed a four-pinhole coded aperture to approximate the depth of the scene, along with a deblurred version of it, although their system required multiple images. Liang et al. [LLW<sup>\*</sup>08] use a similar approach, combining tens of images captured with Hadamard-based coded patterns. Levin et al. [LFDF07] attempted to achieve all-focus and depth recovery simultaneously, relying on image statistics to design an optimal aperture. Depth recovery is limited to a multi-layered representation of the scene. Last, the idea of spatial coding of the mask was transferred to the temporal domain by applying a coded exposure aimed at motion deblurring [RAT06].

In [ZLN09], the authors obtained paired apertures to recover both depth and focus from two images, using both genetic algorithms and gradient descent search. Last, a framework for evaluating coded apertures was recently presented [ZN09], based on the quality of the resulting deblurring and taking into account natural image statistics. Near-optimal apertures are obtained by means of a genetic algorithm. In this paper we follow the same approach, and analyze the obtained apertures along the size, depth and shape dimensions. Additionally, we extend our study by analyzing non-binary masks.

### 3. Optimal Aperture Design

Image blur due to defocus is caused by the loss of high frequency content when capturing the image. The capture process can be modeled as a convolution between the scene being captured and the point spread function (PSF) of the camera, which is defined as the response of the optical system of the camera to an impulse input in the spatial domain. Thus:

$$f = k_d * f_0 + \eta \quad (1)$$

where  $f_0$  is the real scene being photographed,  $f$  is the captured image,  $k_d$  is the PSF and  $\eta$  accounts for the noise introduced in the imaging process. Subscript  $d$  accounts for depth, since the PSF varies with depth or, more specifically,

with the degree of defocus (strictly speaking, it also varies spatially with the position within the image). We will assume that the noise follows a Gaussian distribution of zero mean and standard deviation denoted by  $\sigma$ ,  $N(0, \sigma^2)$ . By means of deconvolution, an approximation  $\hat{f}_0$  to the original sharp image can be obtained. Note that in the frequency domain the convolution becomes a multiplication, and Equation 1 can be written as:

$$F = K_d \cdot F_0 + \zeta \quad (2)$$

As Figure 1 shows, the PSF, and thus the response of the camera, is characterized by the pattern of the aperture. The response to a coded aperture can also be seen in Figure 2, which depicts the calibration array used in our physical experiments. Since, as mentioned, blur is caused by the loss of information at certain frequencies, the response of an aperture is better analyzed in the frequency domain. Figure 3 depicts a 1D slice of the power spectrum of different aperture patterns, computed by Fourier transforming the aperture (note that the y-axis is log-scale). This shows the magnitude of the response for different frequencies. Circular apertures exhibit zero crossings at several frequencies, and thus information at those frequencies is lost during the imaging process. Optimal apertures for deblurring therefore seek a smooth power spectrum, while keeping the transmitted energy as high as possible.



Figure 1: Left: *Images of the response to a point light of different apertures* (from top to bottom: focused aperture, defocused circular aperture -defocus depth = 90 cm- and one of our coded apertures -defocus depth = 90 cm-, shown in the right). A LED and black cardboard were used to create the point light. Right: *Canon EF 50mm f/1.8 lens with one of our coded apertures*.

#### 3.1. Aperture Quality Metric

Devising an aperture pattern whose frequency response is optimal can be done in different manners. In this paper we follow the approach of Zhou and Nayar [ZN09], which states the quality of an aperture pattern based on the quality of the deconvolution and on a prior model of natural images. In the following we briefly describe the metric and its foundation, and we refer the reader to the original paper for additional details.

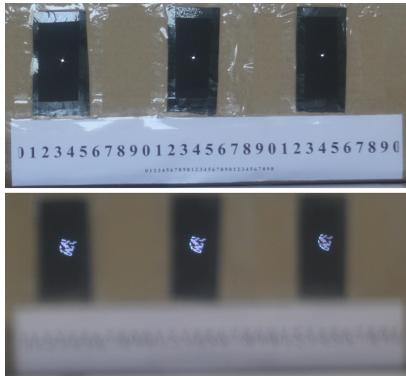


Figure 2: Our poor man’s LED array used to calibrate the PSFs of the apertures. Top: Focused image. Bottom: Image taken with one of our coded apertures at a defocus depth of 70 cm.

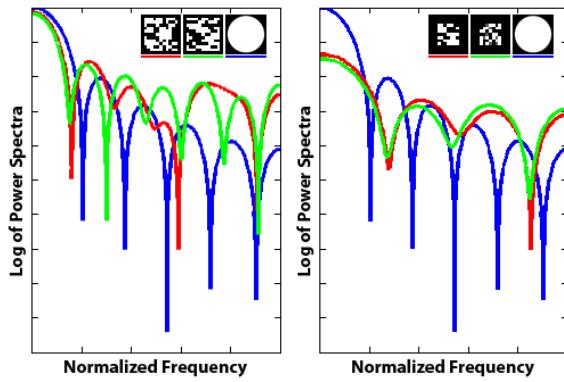


Figure 3: Power spectra comparison of different apertures with respect to a circular aperture (blue). Left: Our apertures for resolution  $11 \times 11$  and noise levels  $\sigma = 0.001$  (red) and  $\sigma = 0.005$  (green). Right: Our apertures for resolution  $7 \times 7$ , binary (red) and non-binary (green).

The quality metric chosen is the expectation of the  $L_2$  distance between the deconvolved image  $\hat{F}_0$  and the ground truth image  $F_0$  with respect to  $\zeta$ , which we want to be minimal (note that we have removed the subscript  $d$  for the sake of simplicity):

$$R(K, F_0, C) = \mathbb{E}_{\zeta} [\|\hat{F}_0 - F_0\|^2] \quad (3)$$

The recovered image  $\hat{F}_0$  can be obtained using Wiener deconvolution as follows:

$$\hat{F}_0 = \frac{F \cdot \bar{K}}{|K|^2 + |C|^2} \quad (4)$$

where  $\bar{K}$  is the complex conjugate of  $K$ , and  $|K|^2 = K \cdot \bar{K}$ .  $|C|^2 = C \cdot \bar{C}$  is the matrix of noise-to-signal power ratios (NSR) of the additive noise. Substituting this formulation in

Equation 3 we have:

$$R(K, F_0, C) = \mathbb{E}_{\zeta} \left[ \left\| \frac{\zeta \cdot \bar{K} - F_0 \cdot |C|^2}{|K|^2 + |C|^2} \right\|^2 \right] \quad (5)$$

and assuming that  $\zeta$  follows a Gaussian distribution with zero mean,  $\zeta \sim N(0, \sigma^2)$ :

$$R(K, F_0, C) = \left\| \frac{\sigma \cdot \bar{K}}{|K|^2 + |C|^2} \right\|^2 + \left\| \frac{F_0 \cdot |C|^2}{|K|^2 + |C|^2} \right\|^2 \quad (6)$$

Using a model of natural images as a prior, the expectation of  $|F_0|^2$  is

$$A(\xi) = \int_{F_0} |F_0(\xi)|^2 d\mu(F_0), \quad (7)$$

where  $\xi$  represents frequency and  $A$  can be approximated by averaging the power spectra of a number of natural images. This way the dependance on  $F_0$ , which is unknown, is circumvented, obtaining:

$$R(K, C) = \left\| \frac{\sigma \cdot \bar{K}}{|K|^2 + |C|^2} \right\|^2 + \left\| \frac{A^{1/2} \cdot |C|^2}{|K|^2 + |C|^2} \right\|^2 \quad (8)$$

The value of  $|C|^2$  which, for a given  $K$ , minimizes the value of  $R$  is  $|C|^2 = \sigma^2/A$ . Substituting this value in Equation 8 yields the sought quality metric, which depends only on the Fourier transform of the aperture pattern  $K$ , the estimated image noise  $\sigma$  and the average power spectra of natural images  $A$ :

$$R(K) = \frac{\sigma^2}{|K|^2 + \sigma^2/A} \quad (9)$$

### 3.2. Aperture Pattern Optimization

Once we have a way of evaluating a certain aperture with Equation 9, an optimization method can be used to obtain the minimum value of  $R(K)$  over the range of possible apertures. The space of possible apertures is infinite, since the aperture can be of different resolutions, and each pixel can in principle take infinite values. A priori the solution is limited only by physical restrictions, i.e. apertures with negative values are not realizable in practice and resolution is limited by the printing process. Resolution is additionally limited by diffraction effects, which appear as the size of the pixels in the aperture gets smaller, and hinder the performance of the aperture. Transmissivity is an additional issue to be taken into account when designing an aperture. Coded apertures typically have lower transmission rates than their circular counterparts, and the use of a longer exposure time to obtain an equivalent brightness to that of the circular aperture can cause other problems such as motion blur. This metric does not consider transmissivity when evaluating an aperture, but still it yields satisfactory results for the majority of cases.

#### 4. Experimental Setup and Results

In order to search for the best aperture pattern we have implemented a genetic algorithm which uses the quality metric described in Section 3 as evaluation function, resembling Zhou and Nayar’s work. The algorithm has the following scheme:

- *Initialization.* The initial population of  $N$  possible apertures is randomly generated. An aperture is defined by a vector of  $L$  elements, each element corresponding to a pixel.
- *Selection.* The quality metric of Equation 9 is used to evaluate the  $N$  possible apertures. They are then sorted according to this value and the best  $M$  apertures are selected.
- *Reproduction.* The selected  $M$  apertures, by means of crossover and mutation, populate the next generation. Crossover implies randomly selecting two apertures, duplicating them, and exchanging corresponding bits between them with probability  $c_1$ , obtaining two new apertures. Mutation ensures diversity by modifying each bit of the aperture with probability  $c_2$ .
- *Termination.* The two previous steps of reproduction and selection are repeated sequentially until the termination condition is met. We use a maximum number of generations  $G$  as stopping condition.

We have tested apertures of two different resolutions,  $11 \times 11$  and  $7 \times 7$  pixels (that is,  $L = 121$  and  $L = 49$ , respectively), while the rest of the parameters we used for the algorithm are  $N = 4000$ ,  $M = 400$ ,  $G = 80$ ,  $c_1 = 0.2$  and  $c_2 = 0.05$ . Since the optimal aperture depends on the noise of the image we have run the algorithm for different noise levels and tested the resulting apertures. Apertures designed for  $\sigma$  values of 0.001 and 0.005 proved to work best for a wide variety of images. Regarding the possible values the pixels in the aperture can take, we have experimented both with binary and non-binary apertures, but at this first stage we show results just for binary apertures. Results for non-binary apertures are discussed in Section 5.

From all the obtained apertures we have chosen three, and a conventional circular aperture, to perform our experiments. We chose the ones which we saw performed best over a wide variety of images. Two of them are  $11 \times 11$  apertures designed for noise levels of  $\sigma = 0.001$  and  $\sigma = 0.005$ , and the third one is a  $7 \times 7$  aperture designed for  $\sigma = 0.005$ . The three of them are depicted in Figure 4. For these apertures, we have performed both a synthetic validation and a validation with physical printed-out apertures.

The synthetic validation is done by simulating the capture process convolving a sharp image  $f_0$  with the aperture (plus noise) as in Equation 1, and subsequently using Wiener deconvolution to recover a deblurred image  $\hat{f}_0$ . The quality of the recovered image is measured using the  $L_2$  norm. We did this for 10 images and computed the average  $L_2$  value. Results are shown in Table 1 for the tested apertures. The minimum error is obtained by the  $7 \times 7$  aperture and the

two  $11 \times 11$  apertures perform very similarly, there is no significant difference, while, as expected, the circular aperture yields worse results. Another measure of the quality of

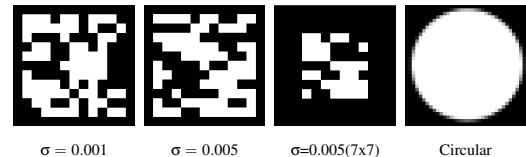


Figure 4: Apertures used in our experiments.

	<b>0.001</b>	<b>0.005</b>	<b>0.005 (7x7)</b>	<b>circular</b>
<b>L2 norm</b>	1.28	1.27	0.88	1.62

Table 1: Results of the  $L_2$  norm for different apertures. The table shows percentages with respect to the maximum error.

the apertures is given by their power spectrum, depicted in Figure 3. The  $11 \times 11$  apertures eliminate less frequencies than the circular aperture, and the  $7 \times 7$  aperture has an even smoother spectrum, which correlates with the  $L_2$  values previously obtained.

Experiments in real scenarios have been performed using a Canon EOS 550D with a Canon EF 50mm f/1.8 II lens shown (unmounted) in Figure 5. Our apertures were printed in a high resolution photomask sheet (see Figure 6 left) and inserted into the lens. The first step is the calibration of the



Figure 5: Camera and lens used in our experiments.

response of the camera (PSF) at different depths. We also calibrated the PSF for different image positions, since the response is spatially varying across the image plane. To do this we used an array of LEDs which we made as close as possible to point light sources with the aid of black cardboard. Figure 2 shows a close-up of the LED array. We locked the focus at 1 m and took an initial focused image, followed by images of the LEDs every 10 cm and until a distance of 2 m, thus having PSFs for defocus depths from 10 to 100 cm. For each position within the image and each depth, the actual cropped image of the LED served us as PSF, after appropriate thresholding of surrounding values which contain residual light. The resulting PSFs for three depths and the four tested apertures are shown in Figure 6 (right).

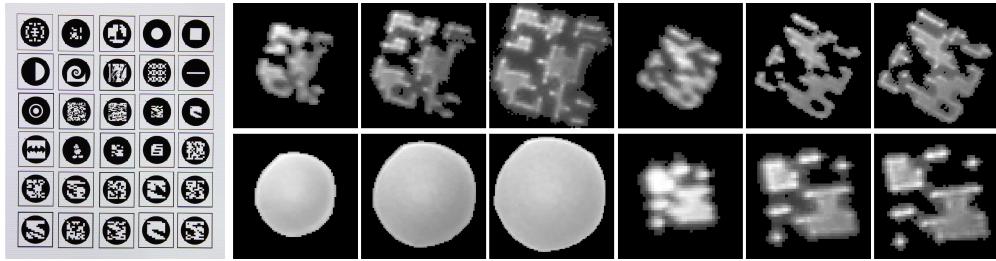


Figure 6: Left: Photomask sheet showing some of the apertures used. Right: PSFs at three different defocus depths (40, 70 and 90 cm) for the four apertures depicted in Figure 4.



Figure 7: Focused ground truth scenes.

Once calibration had been performed, images of three scenes at different depths were taken with each of the selected apertures. These images are then deblurred using the corresponding calibrated PSF by means of Wiener deconvolution. We used a NSR of 0.001 when deconvolving, since it gave the best results. The same exposure time and aperture was used for all the apertures, which results in some images being darker than others. Figure 7 shows the ground truth focused images of the three scenes, whereas Figure 8 depicts the defocused image captured with each aperture and the recovered image for the three different depths. Insets show the corresponding PSF. For all cases our apertures clearly outperform the circular one. The results of the other three apertures are fairly similar, with the  $7 \times 7$  aperture revealing more detail than the others in some regions. However, we believe this may be due to the fact that because of its smaller size, it offers a wider depth of field, thus causing less defocus blur for the same settings as the others. The ringing artifacts which can be observed are probably partially caused by inaccuracies of the calibrated PSFs. Additionally, and although very minor, some of the apertures exhibit slight diffraction effects which can also be the cause of artifacts due to misalignments of the color channels [VRA\*07].

## 5. Study of Non-Binary Apertures

Binary codes have the initial advantage of reducing the search space, and are usually preferred in the existing literature. However, there is no principled motivation to restrict the aperture pixel values to either black or white, other than apparent simplicity. A notable exception in this regard is the work by Veeraraghavan and colleagues [VRA\*07], where the authors report the advantages of continuous-valued aper-

tures, found by gradient descent optimization, over their binary counterparts. In this section we perform an analysis of non-binary apertures focused on our specific context and optimization method; in order to limit the search space of the genetic algorithm, we restrict the set of possible values to  $\{0, 0.5, 1\}$ .

We have studied the quality of the resulting aperture and the computation time for different executions of the genetic algorithm for the cases of binary and non-binary apertures. We have varied both the initial population  $N$  and the number of generations  $G$ , yielding seven different combinations of these two parameters. For each combination of parameters we have performed three executions of the algorithm, plotting the average values. For all the figures in this section, the x-axis shows the initial population  $N$  and the number of generations  $G$  of each set of executions. The number of selected apertures,  $M$ , is always a 10% of the initial population, the crossover probability  $c_1$  is set to 0.2 and the probability of mutation  $c_2$  is 0.05. All the calculated apertures have a resolution of  $7 \times 7$ . The reason of this is two-fold; first, computational cost of the algorithm is significantly reduced, and second, our previous experiments have shown that  $7 \times 7$  apertures yield results on par with (or better than)  $11 \times 11$  apertures. The value of  $\sigma$  (noise level) is set to 0.005 for all executions.

Figure 9 shows the average value of the quality metric to which the algorithm converged. Non-binary apertures tend to converge to slightly lower values of  $R(K)$ , potentially indicating a better performance. However, as expected, it also takes longer for non-binary apertures to converge to a stable value of  $R(K)$ . The execution times consumed until conver-

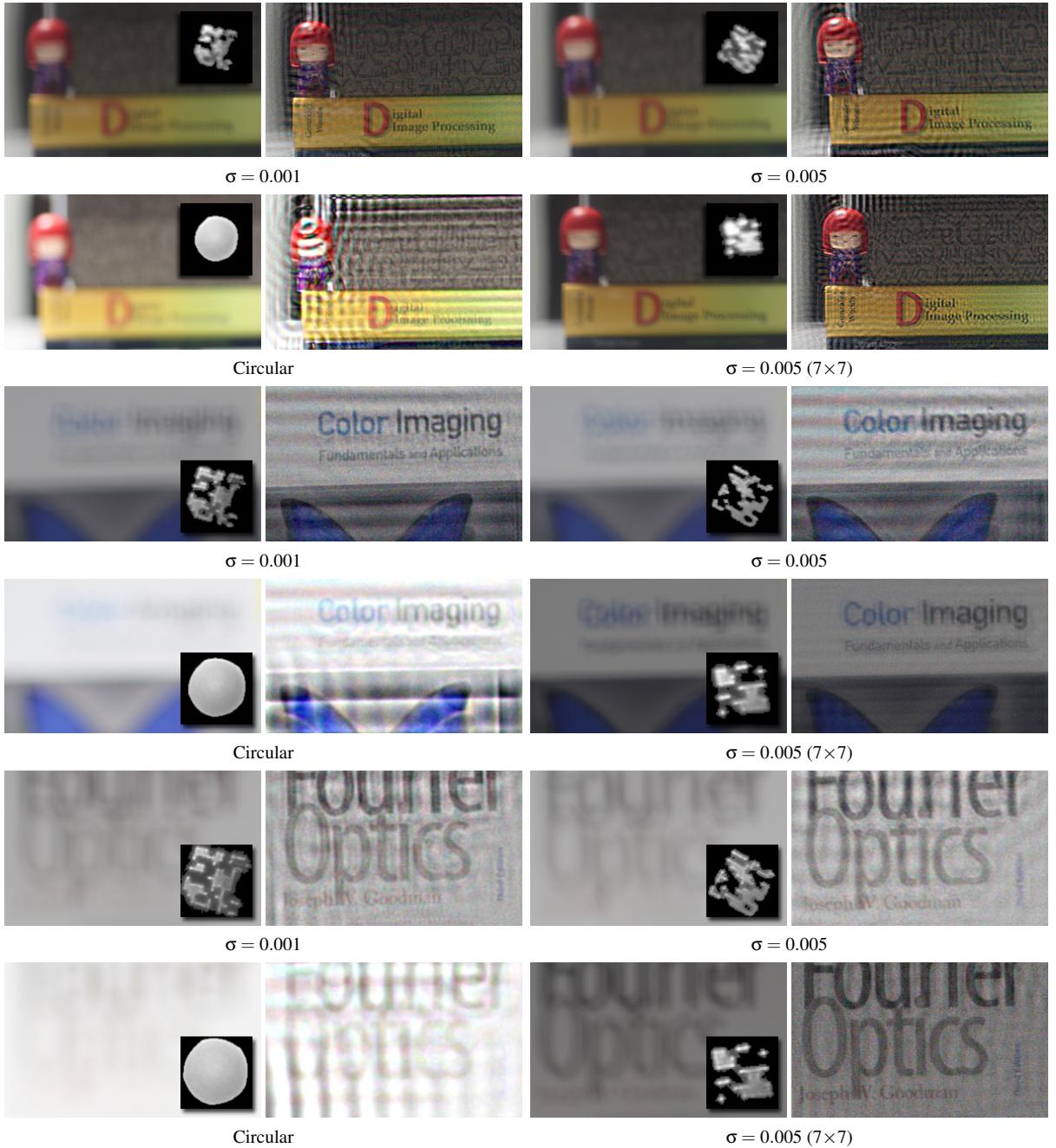


Figure 8: From top to bottom, each of the three scenes have been captured at a defocus depth of 40, 70 and 90 cm, respectively. For each pair of images, the left image shows the captured defocused image and the right image the recovered one. Insets depict the PSF of the aperture used in each case.

gence when running the algorithm on an Intel Core i7 930 @ 2.80GHz are shown in Figure 10.

For all the optimal apertures obtained in the different ex-

ecutions we have performed a synthetic evaluation similar to the one described in Section 4. We applied Equation 1 to an image  $f_0$  of the ISO 12233 resolution chart, to simulate

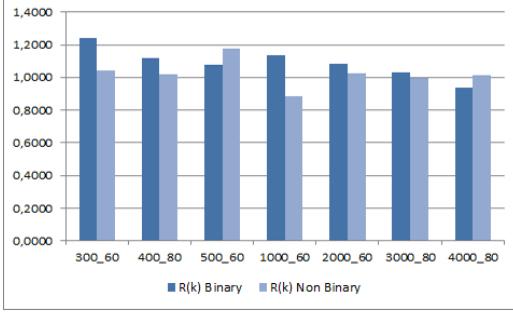


Figure 9: Average value of the quality metric  $R(K)$  for binary and non-binary apertures and for different initial parameters of the genetic algorithm.

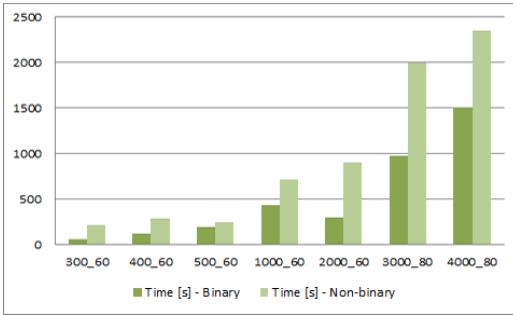


Figure 10: Average value of the time until convergence (in seconds) for binary and non-binary apertures and for different initial parameters of the genetic algorithm.

the capture process with the different apertures; we then performed Wiener deconvolution to recover the estimated sharp image  $\hat{f}_0$ . We have computed the  $L_2$  norm between  $\hat{f}_0$  and  $f_0$  and plotted the results in Figure 11. The non-binary apertures tend to behave better, the global tendency thus correlating with that of the quality metric  $R(K)$ . Nevertheless, this graph shows how lower values of  $R(K)$  not necessarily yield lower values of the  $L_2$  norm. This can be explained by the fact that  $R(K)$  is devised to give optimal performance over the entire space of natural images and thus may not be optimal for an image in particular [ZN09]. Figure 12 shows the image of the chart after convolution, and the recovered image for the best binary and non-binary apertures we obtained. For these two apertures we also plotted the power spectrum, shown in Figure 3 (right). Although both spectra are similar, overall the non-binary aperture has a more favorable response.

To further test the performance of binary vs. non-binary apertures, we printed out these two best apertures (shown in the insets of Figure 12) and captured real images with them. We have calibrated their PSF at different depths as explained in Section 4, and captured a set of images which we then have recovered using Wiener deconvolution. Figure

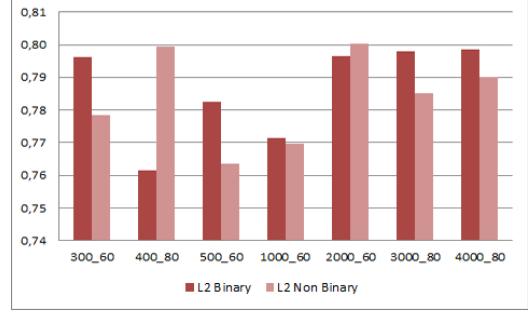


Figure 11: Average value of the  $L_2$  norm for binary and non-binary apertures and for different initial parameters of the genetic algorithm. Values show percentage with respect to the maximum error.

13 shows the results for a defocus depth of 70 cm and 90 cm. The corresponding ground truth focused scenes are shown in Figure 7. We can see how even though both recover detail to a great extent, the non-binary aperture performs better.

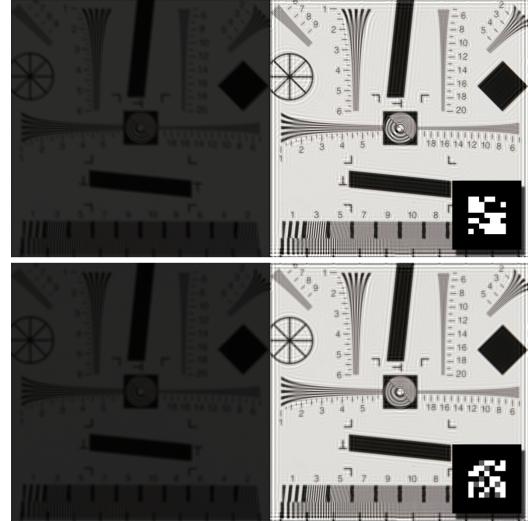


Figure 12: Top left: Defocused image of the ISO 12233 chart obtained using Equation 1 with the aperture shown in the inset in the right as convolution kernel. The aperture is the optimal binary aperture we obtained. Top right: Image recovered using Wiener deconvolution. Bottom row: Same for the optimal non-binary aperture, shown in the inset.

## 6. Conclusions and Future Work

In this paper we have introduced a comprehensive study of coded apertures for defocus deblurring, and implemented the full pipeline: from the genetic algorithms to obtain the codes, to their physical realization, and finally to the actual deblurring of out-of-focus images. We have analyzed the perfor-



Figure 13: For each pair of images, the left image shows the captured defocused image and the right image the deblurred image. Insets depict the PSF used in each case. Defocus depths are 70 cm (top scene) and 90 cm (bottom scene). The color of the images differs from those in Figure 8; please note that this is due to the fact that illumination conditions during the capture process were different, and not to the coded apertures themselves.

mance of the different patterns along several dimensions, namely shape, depth and size. For instance, we found that  $7 \times 7$  apertures are on par with, or outperform, higher resolution ones, which tend to be more computationally expensive to obtain. Additionally, we have extended previous works in the literature by lifting the binary restriction in our patterns, and allowing the genetic algorithms to add mid-gray to the binary (black or white) set of possible values. Although our results are not conclusive and more research needs to be carried out, initial findings suggest that there may be value in exploring continuous apertures, where several gray levels are allowed.

The inherent reduced light transmission when placing a modulating mask at the lens is also a factor that we would like to investigate further. By adding a term that maximizes transmission, we may come up with more efficient apertures. Similarly, finding coded apertures that optimize both defocus deblurring and depth is still an open problem where the community has barely scratched the surface. Last, we believe that the results shown in this paper show the viability and potential of this line of research, and we hope to raise awareness of this exciting field, fostering the creation of more research groups and potential collaborations.

## 7. Acknowledgements

We would like to thank the reviewers for their valuable comments. We also thank C. Zhou for his insights. This research has been partially funded by a Marie Curie grant from the 7th Framework Prog. (251415), the Spanish Ministry of Science and Technology (TIN2010-21543) and the Gobierno de Aragón (OTRI 2009/0411 and CTPP05/09). Belen Masia is

supported by a FPU grant from the Spanish Ministry of Education.

## References

- [GFS89] GOTTESMAN S., FENIMORE E.: New family of binary arrays new family of binary arrays for coded aperture imaging. *Applied Optics*, 20 (1989), 4344–4352. [2](#)
- [HM98] HIURA S., MATSUYAMA T.: Depth measurement by the multi-focus camera. In *IEEE Conference on Computer Vision and Pattern Recognition* (1998). [2](#)
- [ItZ92] IN 'T ZAND J.: *Coded Aperture Imaging in High-Energy Astronomy*. PhD thesis, University of Utrecht, 1992. [2](#)
- [LFDF07] LEVIN A., FERGUS R., DURAND F., FREEMAN W.: Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics* 26, 3 (2007). [2](#)
- [LLW\*08] LIANG C., LIN T., WONG B., LIU C., CHEN H.: Programmable aperture photography: multiplexed light field acquisition. *ACM Transactions on Graphics* 27, 3 (2008). [2](#)
- [NM00] NAYAR S., MITSUNAGA T.: High dynamic range high dynamic range imaging: spatially varying pixel exposures. In *Computer Vision and Pattern Recognition* (2000), vol. 1, pp. 472–479. [2](#)
- [RAT06] RASKAR R., AGRAWAL A., TUBMLIN J.: Coded exposure photography: Motion deblurring using shuttered shutter. *ACM Transactions on Graphics* 25, 3 (2006), 795–804. [2](#)
- [VRA\*07] VEERARAGHAVAN A., RASKAR R., AGRAWAL A., MOHAN A., TUMBLIN J.: Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.* 26 (July 2007). [2, 5](#)
- [ZLN09] ZHOU C., LIN S., NAYAR S.: Coded aperture pairs for depth from defocus. In *ICCV (oral)* (2009). [2](#)
- [ZN09] ZHOU C., NAYAR S. K.: What are Good Apertures for Defocus Deblurring? In *IEEE International Conference on Computational Photography* (2009). [2, 7](#)

