

Proyecto Fin de Carrera
Ingeniería en Informática
Curso 2010/2011 (Junio 2011)

FOTOGRAFÍA COMPUTACIONAL: DISEÑO DE APERTURAS CODIFICADAS PARA RECUPERACIÓN DE IMÁGENES DESENFOCADAS



Adrián Corrales Romeo
Directora: Belén Masiá Corcoy
Ponente: Dr. Diego Gutiérrez Pérez

Grupo de Informática Gráfica Avanzada
Departamento de Informática e Ingeniería de Sistemas
Escuela Superior de Ingeniería y Arquitectura
Universidad de Zaragoza



Universidad
Zaragoza



Departamento de
Informática e Ingeniería
de Sistemas
Universidad Zaragoza



A mis padres, JáCC y Glo, a mi segunda madre, Pilar, a mi hermano, Alberto, a una pequeña química que apareció sin avisar, a B, a D y a todos mis amigos.

Agradecimientos

Me gustaría dar las gracias a Belén Y Diego por ofrecerme la posibilidad de trabajar en un campo tan novedoso, siendo los primeros (a nivel nacional); pero además me gustaría agradecer su tiempo, ayuda y, muchas veces, paciencia a lo largo del desarrollo mi proyecto fin de carrera. También quiero agradecer a mis compañeros del grupo de trabajo (en el seno del GIGA) por sus aportaciones, y su buen humor, durante los días que hemos compartido lugar de trabajo. Finalmente, pero siempre importante, le doy las gracias a mi entorno por apoyarme en cada momento y, simplemente, por “estar ahí” (muchas veces desconociendo la relevancia de su existencia).

RESUMEN

La fotografía computacional es un campo emergente nacido de la unión de la óptica, el procesamiento de la señal, la visión por computador, la informática gráfica o incluso de la electrónica y el arte. En los últimos años este campo ha producido avances espectaculares en cuanto al procesamiento de imágenes se refiere.

La fotografía tradicional se basa en representar espacialmente, en una matriz de dos dimensiones, la escena real que se está observando en el momento de la captura. Uno de los problemas de este proceso es la aparición de zonas borrosas (blur) en las imágenes por falta de enfoque, movimiento (de la cámara o de la escena) u otros motivos. Esto es así puesto que se pierde la información necesaria para representar de manera correcta parte de la escena.

Las primeras aproximaciones para resolver este problema se basaban en el estudio de las estadísticas de las imágenes, con el objetivo de añadir información consiguiendo recomponer parte de esa información perdida. En los últimos años, de manera emergente, se han desarrollado nuevas técnicas que permiten codificar la información en el proceso de captura de imágenes.

Este trabajo, aprovechando el surgimiento de estas técnicas, presenta el uso de las *aperturas codificadas* como herramienta para codificar la información en el proceso de captura, consiguiendo eliminar blur y recuperar esas zonas borrosas. El trabajo se centra en la recuperación de blur por desenfoque (proceso comúnmente denominado *defocus deblurring*), por lo que se asumirá que las escenas capturadas no presentarán blur por movimiento; esto quiere decir que tanto la cámara como las escenas permanecerán inmóviles durante la captura de las mismas.

La primera parte de este trabajo consiste en obtener aperturas codificadas cuyo diseño sea óptimo (o casi óptimo) para el problema de *defocus deblurring*. Esto se realizará planteando el problema como uno de optimización, que se resolverá mediante un algoritmo genético. Se obtendrán, de esta manera, aperturas diseñadas para distintos niveles de ruido y con distinta resolución espacial, para su posterior análisis y validación. Estas aperturas se validarán inicialmente mediante simulaciones, simulando el proceso de captura para obtener imágenes con defocus blur, consiguiendo eliminar a posteriori la mayor parte del mismo. Después, una vez validado el proceso mediante simulación, se trasladará el problema a entornos reales con una cámara fotográfica y la impresión de las aperturas en material fotolitográfico. Finalmente se mostrará, en forma de resultados, el correcto funcionamiento del proceso y las limitaciones del mismo. Se ha realizado también, como parte de este proyecto, un estudio del estado del arte del campo de la fotografía computacional; estudio que se ha considerado de interés por tratarse de un campo de investigación de muy reciente aparición y por lo tanto con muy poca documentación existente.

Además, parte de la investigación realizada en este proyecto ha permitido la creación de un artículo aceptado en el Congreso Ibero-American de Informática Gráfica (SIACG 2011) que fue considerado como uno de los tres mejores artículos del congreso, siendo propuesto para su extensión y sumisión al Computer Graphics Forum, revista JCR.

Índice general

1. Introducción	1
2. Estado del arte	5
2.1. Definición de fotografía computacional	5
2.2. Clasificación en base al elemento modificado	5
2.3. Clasificación en base a la evolución temporal	6
2.4. Clasificación en base a la aplicación final	6
2.4.1. Captura de imágenes de alto rango dinámico (HDR)	7
2.4.2. Extensión o modificación de la profundidad de campo (DOF)	7
2.4.3. Re-enfoque (Deblurring)	8
2.4.4. Detección de profundidad	8
2.4.5. Detección de bordes	8
2.4.6. Flash/No-Flash	9
3. Defocus deblurring	11
3.1. Formulación matemática del problema	12
3.2. Trabajo previo	14
3.3. Solución adoptada	15
4. Obtención de aperturas codificadas óptimas	17
4.1. Elección del método de optimización	17
4.2. Implementación de un algoritmo genético	18
4.2.1. Representación	19
4.2.2. Operadores genéticos: cruce y mutación	20
4.2.3. Condición de parada	21
4.3. Convergencia del algoritmo	21
4.4. Aperturas obtenidas	22
5. Validación I: Simulaciones	25
5.1. Simulación del proceso de captura y recuperación de la información de la imagen	25
5.1.1. Proceso de captura o desenfoque de la imagen	25
5.1.2. Recuperación de la información de la imagen	26
5.2. Análisis del error obtenido: Norma L_2 e índice SSIM	28
5.2.1. Norma L_2	28
5.2.2. Índice SSIM	29
5.3. Espectro de potencia	30

6. Validación II: Experimentos reales	33
6.1. Configuración del sistema	33
6.2. Calibración del sistema	34
6.2.1. Metodología de obtención de la PSF del sistema	34
6.2.2. Construcción de un array de diodos LED	35
6.2.3. Captura de las imágenes de calibración	37
6.3. Captura y recuperación de las imágenes con blur	37
6.4. Resultados	38
7. Conclusiones y trabajo futuro	43
7.1. Conclusiones	43
7.2. Desarrollo del proyecto	44
7.3. Trabajo futuro	44
7.4. Conclusiones personales	44
A. Derivación matemática de la función de evaluación	51
B. Desmontaje del objetivo	53
C. Estudio de la convergencia del algoritmo genético	57
C.1. Supervivientes	57
C.2. Generaciones	58
C.3. Población	58
C.4. Cruzamiento	59
C.5. Mutación	59
D. Resultados de las pruebas reales	61
E. Artículo: Coded Apertures for Defocus Deblurring	69

Índice de figuras

1.1. Aparición de zonas borrosas en imágenes	2
1.2. Diagrama de bloques del trabajo realizado	3
3.1. Ejemplo de imágenes con <i>blur</i>	11
3.2. Concepto de profundidad de campo y círculos de confusión	12
3.3. Proceso, conceptual, de aparición de blur por desenfoque en imágenes	13
3.4. Apertura circular real de un objetivo <i>Canon</i>	13
3.5. Respuesta en el dominio de la frecuencia de una apertura circular	13
4.1. Esquema, conceptual, de un algoritmo genético	19
4.2. Ejemplo de apertura codificada binaria	20
4.3. Operador binario de cruce	20
4.4. Operador unario de mutación	21
4.7. Aperturas 11*11 obtenidas	23
4.8. Aperturas 7*7 y 20*20 obtenidas	23
5.1. Ejemplo, conceptual, del proceso simulado completo	25
5.2. Simulación de desenfoque de una imagen	26
5.3. Simulación de re-ensfoque de una imagen	27
5.4. Resultados, conceptuales, de la norma (o distancia) L_2	28
5.5. Resultados de la norma L_2 para diferentes aperturas	28
5.6. Resultados, conceptuales, de la medida <i>SSIM</i>	29
5.7. Resultados de la medida <i>SSIM</i> para diferentes aperturas	29
5.8. Comparación del espectro de potencia entre varias aperturas	31
6.1. Aperturas impresas en material fotolitográfico	33
6.2. Equipo utilizado para los experimentos reales	34
6.3. Obtención de PSFs para una apertura codificada	35
6.4. Tablero ideal para obtener PSFs de una apertura codificada a una distancia cualquiera	36
6.5. Array de diodos LED construido	36
6.6. Procedimiento utilizado para obtener las PSFs de una apertura	37
6.7. Primer ejemplo de los resultados obtenidos en los experimentos reales	39
6.8. Segundo ejemplo de los resultados obtenidos en los experimentos reales	39
6.9. Importancia de la profundidad de desenfoque en experimentos reales	40
6.10. Comparativa entre resultados obtenidos en los experimentos reales	41

1. Introducción

La fotografía computacional es un campo emergente nacido de la unión de la óptica, el procesamiento de la señal, la visión por computador, la informática gráfica o incluso de la electrónica y el arte. En los últimos años este campo ha producido avances espectaculares en cuanto al procesamiento de imágenes se refiere. La fotografía tradicional se basa en representar espacialmente, en una matriz de dos dimensiones, la escena real que se está observando en el momento de la captura. La fotografía computacional surge con el objetivo de obtener y explotar la información capturada, más allá que la simple representación en dos dimensiones de la escena, obteniendo una representación más completa del mundo real, traspasando las limitaciones de la fotografía tradicional.

Una vez realizado un estudio sobre el estado del arte de la fotografía computacional, que se muestra en la siguiente sección, se obtuvo una visión general acerca de nuevos algoritmos que explotan la información capturada con una cámara fotográfica. Con éstos se puede obtener información de una imagen (como la profundidad de los objetos de la escena), mejorar la calidad visual de una escena (creando una imagen en alto rango dinámico a partir de un conjunto de fotografías), cambiar la iluminación de una imagen o recuperar la información perdida en forma de zonas borrosas en una imagen.

Uno de los enfoques en fotografía computacional es codificar la información de la luz, antes de que incida en el sensor de la cámara fotográfica, de manera que se pueda decodificar dicha información a posteriori obteniendo una representación más completa de la escena capturada. Existen varios métodos de proceder dentro de este enfoque, pudiendo hacer una clasificación basada en la manera de codificar esa información: iluminación codificada y exposición codificada (en tiempo y en espacio).

En cuanto a lo que *iluminación codificada* se refiere, y como ejemplo del tipo de técnicas empleadas en fotografía computacional, una aplicación a un problema difícil de resolver en informática gráfica es la detección de los bordes de los objetos en una imagen. Al fotografiar un objeto utilizando el flash en la captura, se produce la proyección de dicho objeto en forma de sombra según la posición espacial del flash; en 2004 se presentó un método que, utilizando una cámara con varios flashes en distintas posiciones espaciales, conseguía detectar los bordes de los objetos de la escena fotografiada [26] (por composición de las capturas tomadas con los distintos flashes).

En cuanto a *exposición codificada*, una aplicación es recuperar la información perdida por la aparición de zonas borrosas en la imagen final provocada por movimiento en el momento de la captura (ya sea de la cámara o de un objeto de la escena) o por otros motivos (i.e. una mala

configuración de los parámetros de la cámara). Dicha codificación, como se ha dicho anteriormente, puede ser realizada en el tiempo o en el espacio.

El tiempo de exposición en una captura es el tiempo que permanece abierto el diafragma de la cámara durante el cual se graba la escena en el sensor de la misma. Cuando existe movimiento en una captura, los detalles de la escena (i.e. las letras de la matrícula de un vehículo) aparecen de manera borrosa en la imagen final y es imposible recuperar dicha información. En 2006, mediante aperturas y cierres del diafragma (con una secuencia obtenida computacionalmente) durante el tiempo de exposición en la captura, se consiguió recuperar la información de detalle perdida en una escena con objetos en movimiento [24]. Sin embargo, la aparición de zonas borrosas en una imagen puede ser provocada por otros motivos distintos al movimiento (i.e. desenfoque a la hora de hacer la captura). En 2007 se presentó un método [36] cuyo objetivo era recuperar la información perdida en una imagen, mediante la introducción de un patrón en el sistema de lentes de una cámara fotográfica, codificando la información en el espacio durante el tiempo de exposición de la captura.



Figura 1.1: *Aparición de zonas borrosas en imágenes*. Izquierda: *zonas borrosas por movimiento del objeto capturado*. Centro: *zonas borrosas por desenfoque, debido a una mala elección de los parámetros de la cámara*. Derecha: *zonas borrosas por exposición, debido a las limitaciones de campo de la cámara*.

Lo que se persigue en este trabajo es recuperar imágenes, como la imagen central de la Figura 1.1, con zonas borrosas por desenfoque, proceso usualmente denominado *defocus deblurring*. Aunque existen, como se ha visto, varios métodos para codificar la información de la luz, este trabajo se centra en la codificación de la exposición en el espacio. Este concepto se denomina *aperturas codificadas* y será la herramienta que permita, en este trabajo, recuperar la información perdida en las zonas borrosas de una imagen. El objetivo principal será obtener una apertura codificada, óptima, que permita resolver dicho problema.

Las aperturas codificadas son patrones, obtenidos por medio de algoritmos computacionales, cuyo objetivo es mejorar las características que poseen las aperturas (aproximadamente circulares¹) de las cámaras fotográficas. Estos patrones se colocan en el sistema de lentes (objetivo) de una cámara fotográfica para modificar la respuesta frecuencial de la apertura. Cuando el objetivo es *defocus deblurring*, mediante el uso de la apertura se pretende mantener la amplitud de las altas frecuencias de la imagen original. Algunos de los trabajos existentes, simultáneamente, intentan recuperar mapas de profundidad de una escena interpretando la codificación obtenida en la captura, aunque existe un compromiso a la hora de elegir el problema a resolver², ya que

¹La apertura viene definida por el tamaño del diafragma (en el momento de capturar una imagen) de una cámara fotográfica, que está compuesto por unas hojas solapadas y ajustables; esto hace que la forma de la apertura no sea circular sino poligonal como se puede ver en la Figura 3.4.

²Recuperar información perdida u obtener mapas de profundidad de una imagen.

las aperturas diseñadas para resolver un problema no son óptimas para resolver el otro. Existen algunos enfoques interesantes en esa dirección [11, 15, 16] pero, dada la naturaleza de las zonas borrosas en la imagen, esas aproximaciones imponen una representación multi-capa de la escena representada.

En la elaboración de este trabajo se ha seguido un enfoque que define el proceso de desenfoque como una *convolución* entre la escena original capturada y la respuesta de la apertura codificada utilizada, con la adición final de ruido. En principio, esta aproximación conduciría a un problema de deconvolución ciega, dado que la respuesta de la apertura codificada, por lo general, es desconocida. Se asume que la cámara queda estática en las capturas para que no se produzca desenfoque por movimiento, por lo que la respuesta efectiva de la apertura codificada queda definida por la función de dispersión del punto o PSF (*point spread function*) del sistema óptico. En este sentido, las aperturas circulares que poseen todas las cámaras fotográficas tienen una muy mala respuesta en el dominio de la frecuencia puesto que, aparte de diezmar la amplitud de las altas frecuencias, poseen numerosos pasos por cero para varias frecuencias. Esto impide recuperar la información (en dichas frecuencias) durante el proceso de recuperación de las zonas borrosas mediante deconvolución.

El trabajo realizado presenta varias aperturas codificadas que poseen mejor respuesta, en el dominio de la frecuencia, que una apertura circular; por ello, estas aperturas permiten recuperar la información perdida en el proceso de captura de la imagen. Siguiendo el enfoque de trabajos anteriores [36], este trabajo plantea el problema como un problema de optimización resuelto mediante el uso de un algoritmo genético que obtiene una apertura codificada. Una vez obtenida una serie de aperturas codificadas de distinta resolución y para diferentes niveles de ruido, se procede a su análisis y evaluación mediante simulaciones. Una vez validadas mediante simulación, se procede a imprimirlas en una lámina de material fotolitográfico. Finalmente, se insertan en el objetivo de una cámara real para concluir la validación con experimentos reales. El proceso entero se muestra en la Figura 1.2. Además, se analiza el rendimiento de las aperturas en función de su forma y tamaño, además de en función de la profundidad de la escena capturada.

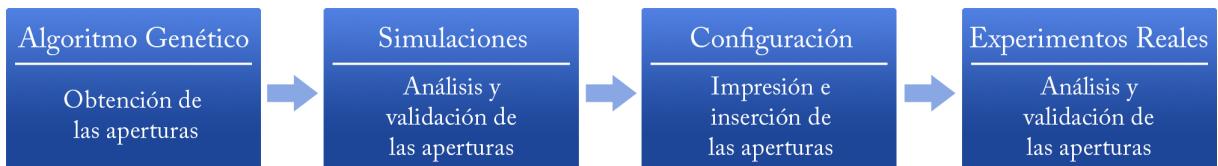


Figura 1.2: *Diagrama de bloques del trabajo realizado.*

2. Estado del arte

En este apartado se va a exponer el estado del arte de la fotografía computacional, como tarea de exploración y análisis previo al inicio del proyecto, mostrando el trabajo realizado por distintos investigadores en este campo. Más que un análisis detallado y exhaustivo, que escapa al alcance de esta memoria, este estado del arte pretende ofrecer un vistazo general del campo.

2.1. Definición de fotografía computacional

Como se ha visto en la introducción, la fotografía computacional es un campo de trabajo de reciente aparición que surge con el objetivo de obtener y explotar la información capturada de una escena, más allá que la simple representación en dos dimensiones de la misma. Haciendo uso de nuevas técnicas se pretende obtener una representación más completa del mundo real, traspasando las limitaciones de la fotografía tradicional.

Mediante la creación de nuevos algoritmos y métodos se pretende llevar más allá los límites de la fotografía tal y como la conocemos logrando, entre otras cosas, mejorar la calidad visual de una escena, cambiar la iluminación de una imagen o recuperar información perdida por desenfoque. Obteniendo, incluso, imágenes imposibles mediante el uso convencional de una cámara fotográfica como son las fotografías panorámicas o efectos de video como es el “tiempo bala” aparecido en distintas películas (*The Matrix*) y videoclips (*The Rolling Stones*).

2.2. Clasificación en base al elemento modificado

Los cuatro elementos, principales, que componen la fotografía tradicional son: *sistema de lentes* encargado de hacer llegar la información de una escena al sensor; *sensor fotosensible* que recibe información de la escena; *obturador* que permite controlar la luz que llega al sensor; fuente externa de luz o *flash* usado para iluminar una escena. Estos elementos han sido explorados para crear nuevas metodologías de captura de imágenes y en función de cual de estos elementos es modificado se puede esbozar una primera clasificación:

- Sistema de lentes: modificando la forma de la apertura en campos como la astronomía [34], añadiendo complejidad al sistema de lentes [21] o, incluso, prescindiendo del mismo [38] a la hora de capturar una imagen.

- Sensor: modificando la forma del sensor dando lugar a la *Plenoptic Camera* [9], que permite capturar la información de la luz de manera n-dimensional [28], o modificando el funcionamiento del sensor para aumentar el rango dinámico [13] o el rango espectral de la imagen capturada.
- Obturador: se puede controlar la luz que llega al sensor de tal manera que se puedan obtener imágenes nítidas de un objeto en movimiento controlando la apertura y cierre del obturador [24] u obtener una imagen en alto rango dinámico combinando varias fotografías tomadas de manera casi simultánea.
- Iluminación: mejorar y controlar la iluminación de la escena a capturar permite obtener representación más completa de la misma, como la forma tridimensional de un objeto mediante la utilización de varios flashes estratégicamente posicionados [26], o haciendo uso de la técnica llamada “flash/no-flash” [6].

2.3. Clasificación en base a la evolución temporal

Haciendo uso de la definición establecida por Ramesh Raskar [25] se puede hacer una clasificación de la fotografía computacional, en base a su evolución en el tiempo, apareciendo tres fases principales: fotografía *épsilon*, fotografía *codificada* y fotografía *en esencia*.

- Fotografía épsilon: construir una cámara mejorada en cuanto a los parámetros tradicionales se refiere (rango dinámico, campo de visión o profundidad de campo). Debido a la reducida capacidad de una cámara, la escena se re-construye mediante la utilización de varias fotografías (cada una capturada con épsilon variaciones de los parámetros de la cámara).
- Fotografía codificada: construir herramientas que traspasen las capacidades de esta cámara mejorada. El objetivo es codificar la información capturada de la escena (en una o muy pocas imágenes) de tal manera que su decodificación permita una descomposición versátil de la escena en base a sus parámetros físicos.
- Fotografía en esencia: el siguiente paso, en este momento sin explorar, sería olvidar que una cámara fotográfica debe imitar al ojo humano. Intentando recuperar mayor información sobre la escena (la esencia de la escena), a parte de los parámetros físicos de la misma, la idea es poder generar nuevas formas de expresión artística visual y comunicación.

2.4. Clasificación en base a la aplicación final

La aplicación final para la que se desarrollan métodos o técnicas de captura de la información de las escenas conforma otra clasificación posible. A continuación se exponen, en base a la aplicación final, las técnicas más destacadas.

2.4.1. Captura de imágenes de alto rango dinámico (HDR)

Diseño del sensor

Una primera aproximación es el uso de varios elementos sensores con diferentes niveles de sensibilidad, combinando las distintas medidas obtenidas por cada elemento sensor en el chip antes de dar lugar a la imagen HDR, dando lugar a varias patentes [10, 29, 32]. Uno de los problemas de este enfoque es la reducción de la frecuencia de muestreo espacial (pudiendo aparecer efectos de *aliasing*) y la pérdida de resolución en la imagen final.

Un enfoque diferente se propuso en [3], donde se mide el tiempo que le cuesta al sistema llegar a la saturación de cada zona de la imagen capturada. De esta manera, ese tiempo codifica la información de alto rango dinámico, al ser inversamente proporcional a la luminosidad de cada zona de la imagen.

Existen enfoques más novedosos y flexibles donde la exposiciones varían a través del espacio de la cámara [1, 22]. Usando patrones con diferentes niveles de sensibilidad la flexibilidad del sistema aumenta aunque, de nuevo, la resolución espacial disminuye y pueden aparecer efectos de *aliasing*. Las distintas mediciones capturadas se combinan creando una imagen HDR.

Exposición múltiple

Otra aproximación para obtener imágenes en HDR es capturar varias imágenes usando, para cada una, diferentes niveles de exposición. La idea principal es resaltar las zonas oscuras con niveles de exposición altos (saturando las zonas claras de la escena) y al contrario con niveles de exposición bajos. De esta manera, cuando una misma escena es capturada con distintos niveles de exposición, se pueden combinar las imágenes seleccionando cada píxel de aquella imagen en la que no aparece muy oscuro ni saturado. Esta técnica se llama sucesión de exposiciones y ha sido utilizada por numerosos autores [4, 17]. Estos autores asumen que el brillo medido por el sistema de captura (cámara) se relaciona linealmente con los valores de la escena real pero, casi siempre, existe una relación no lineal entre ambos.

Debido a este motivo algunos autores han propuesto métodos para estimar la función de respuesta radiométrica de una cámara, obteniendo mapas de radiancia (donde los valores de los píxeles en la imagen final son proporcionales a los valores de la escena real) utilizados, por ejemplo, en aplicaciones médicas [5, 18, 19].

2.4.2. Extensión o modificación de la profundidad de campo (DOF)

Los dispositivos de captura de imágenes tienen limitaciones relacionadas con la configuración física existente en el dispositivo en el momento de realizar una captura. Entre ellas se encuentra la profundidad de campo definida por el tamaño del diafragma utilizado en el momento de la captura. Aunque el concepto se explicará en la Sección 3, la profundidad de campo determina, a grandes rasgos, qué objetos de la escena aparecerán enfocados y cuáles desenfocados en la imagen final.

En el año 2006 se desarrolló una cámara que permite obtener imágenes enfocadas a distintas profundidades [28]. Para desarrollar el dispositivo, se insertó una matriz de “micro-lentes” entre

el sensor y el sistema de lentes de tal manera que cada micro-lente obtiene información de manera independiente a las demás, reordenando los rayos incidentes se consigue enfocar la imagen a distintas profundidades. Esta implementación permite enfocar digitalmente una imagen y variar la profundidad de campo, lo que ofrece posibilidades muy altas de personalización de las imágenes finales con el uso de una cámara que funciona exactamente igual que una cámara convencional (desde el punto de vista del usuario).

En el año 2008, se presentó un dispositivo que extiende la profundidad de campo a un nivel superior pudiendo, además de enfocar objetos a distintas profundidades en una escena, eliminar ópticamente un objeto a una profundidad (DOF discontinuo) o incluso variar la inclinación de la profundidad de campo (DOF inclinado) [20]. La idea principal para crear este dispositivo es dar libertad de movimiento al sensor, posibilitando variar la profundidad de campo sin cambiar el tamaño del diafragma del objetivo.

2.4.3. Re-enofoque (Deblurring)

La aparición de zonas desenfocadas o *blur* degrada significativamente la calidad de la imagen final. Por ello, en los últimos años, algunos investigadores han propuesto distintos métodos intentando disminuir el efecto ocurrido en este tipo de imágenes. Las primeras aproximaciones se basaban en estimar la función de dispersión del punto (PSF, explicada en la Subsección 3.1), del sistema de captura para usarla en el proceso de re-enofoque [2, 33].

Si bien es cierto que la mayoría de técnicas desarrolladas se han basado en paliar el desenfoque por presencia de movimiento, recientemente se ha presentado una solución para las imágenes con desenfoque por falta de enfoque [36]. Mediante el uso de una apertura codificada se consigue recuperar parte de la información que, a simple vista, se había perdido en el proceso de captura de la escena.

2.4.4. Detección de profundidad

Varios autores han realizado estudios acerca de la obtención de la profundidad de una escena fotografiada, mediante el uso de una apertura codificada óptima diseñada en base a las estadísticas de las imágenes [15], utilizando parejas de aperturas codificadas de alta resolución [35] o analizando la variación del desenfoque aparecido en los bordes de la imagen [37].

2.4.5. Detección de bordes

En el año 2003 se presentó un prototipo basado en una cámara con un flash múltiple con el objetivo de detectar los bordes de los objetos de una escena [26]. El método desarrollado obtiene cuatro fotografías de la misma escena iluminadas desde distintas posiciones (arriba, abajo, derecha e izquierda) con el fin de detectar las sombras proyectadas por los objetos de la escena. Además demostraron que esta técnica se podía utilizar para detectar bordes en vídeos mediante el uso de una secuencia muy rápida de los cuatro flashes.

2.4.6. Flash/No-Flash

En 2001 se exploró por primera vez la idea de capturar la misma escena con y sin el uso del flash de la cámara (técnica “flash/no-flash”) [6]. Con este método consiguen obtener una imagen de la escena como si sólo estuviese iluminada con el flash (sin luz ambiental), estiman las funciones de reflectancia de los objetos de la escena y estiman la distribución espectral de la iluminación ambiente. Bajo este mismo enfoque, otros autores [7, 23] han desarrollado métodos de combinación de pares de imágenes de escenas con poca iluminación ambiental, ya que las imágenes sin flash poseen altos niveles de ruido y las imágenes tomadas con flash alteran el color de los objetos de la escena. Recientemente se ha presentado un prototipo de cámara con un flash que utiliza luz infrarroja y ultravioleta fuera del rango visible, llamado *flash oscuro* [14].

3. Defocus deblurring

En general, cuando se utiliza una cámara fotográfica se pretende obtener una copia de la escena que se está visualizando, pero en múltiples ocasiones las imágenes capturadas poseen zonas borrosas. Estas zonas, generalmente indeseadas, se denominan *blur* y con *deblurring* se define el proceso de eliminación de la mayor parte (idealmente del total) de esas zonas borrosas. Éste es uno de los motivos por los que se comenzaron a estudiar distintas técnicas de re-ensfoque de imágenes o *deblurring techniques*. Aunque, en la actualidad, una gran cantidad de equipos fotográficos son muy sofisticados, el blur en las imágenes sigue siendo un problema de gran interés de estudio.

Los motivos de este problema son por un lado humanos, como el movimiento de la cámara o la ausencia de un enfoque correcto a la hora de hacer una captura. Por otro lado, existen motivos inherentes a la cámara como son la resolución del sensor o la presencia de filtros en el mismo. Por ello, en ciertas ocasiones es inevitable obtener imágenes como las mostradas en la Figura 3.1, que muestra dos ejemplos de imágenes con blur. Los problemas inherentes al equipo son más complicados de subsanar (siendo la mejora en la imagen casi inapreciable) por lo que, generalmente, la fotografía computacional trabaja en la eliminación del blur por movimiento y del blur por desenfoque.

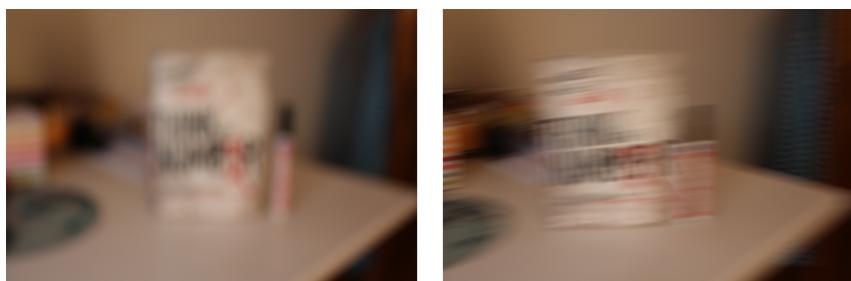


Figura 3.1: Izquierda: *imagen con blur por desenfoque*. Derecha: *imagen con blur por movimiento*.

El motivo de la existencia de blur por desenfoque (*defocus blur*) deriva del tamaño de la apertura utilizada para la captura. Su diámetro establece la *profundidad de campo* en el momento de la captura, que es el intervalo, en el eje Z (perpendicular al plano imagen y al plano del sensor), en el que los objetos de la escena aparecerán enfocados en la imagen capturada (apareciendo desenfocados los objetos que estén localizados fuera de dicho intervalo). Conforme el diámetro del diafragma disminuye en tamaño la profundidad de campo aumenta, siendo el caso extremo la *pinhole camera* en la que el diámetro de diafragma es un punto. En una *pinhole camera* la profundidad de campo es infinita.

El enfoque de las cámaras fotográficas establece, para una profundidad d (distancia, en el eje Z, desde el sensor a un punto de la escena), un plano P perpendicular al eje Z cuyos puntos están perfectamente enfocados; los sucesivos planos imagen (perpendiculares al eje Z) irán perdiendo enfoque a medida que vayan alejándose del plano P . Esto se debe a que todos los rayos provenientes de un punto del plano P inciden en un mismo punto del sensor de la cámara, no ocurriendo así con los puntos que se encuentran a otras profundidades. La aparición de blur en la imagen grabada se debe a que la información proveniente de esos puntos se distribuye por un área circular del sensor, como se puede ver en la Figura 3.2, creando los llamados *círculos de confusión*.

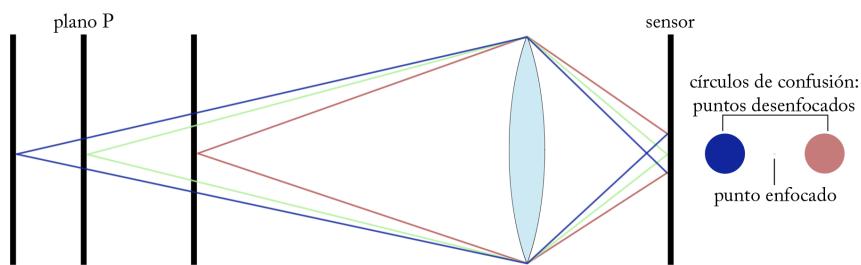


Figura 3.2: *Concepto de profundidad de campo y círculos de confusión. Los puntos del plano P están perfectamente enfocados. Los puntos de otros planos (fuera del intervalo definido por la profundidad de campo) están desenfocados, creando los llamados círculos de confusión.*

3.1. Formulación matemática del problema

El problema de aparición de blur en una imagen está estrechamente relacionado con la forma de la apertura utilizada para la obtención de la misma. La siguiente ecuación describe matemáticamente el proceso de captura de una imagen:

$$f = f_0 * k_d + \eta , \quad (3.1)$$

donde:

- f es la imagen desenfocada capturada,
- f_0 es dicha imagen perfectamente enfocada (i.e. la escena real),
- k_d es la respuesta de la apertura utilizada a una profundidad d ,
- η es el ruido de la imagen.

La respuesta k_d , de la apertura utilizada varía espacialmente en la escena según las coordenadas 2D de dicha escena y la profundidad a la que se encuentra dicha escena respecto a la cámara. Asimismo, no varía sólo con la profundidad absoluta sino también con la distancia entre el plano focal y la escena. Esta respuesta se denomina PSF (*point spread function/función de dispersión del punto*) ya que codifica la respuesta del sistema cuando la entrada es un impulso (i.e. un punto).

$$\begin{array}{c}
 \text{Image} = \text{Image} \cdot \text{Blur Kernel} + \eta \\
 \mathbf{f} = \mathbf{f}_0 \cdot \mathbf{k}_d + \eta
 \end{array}$$

Figura 3.3: Proceso, conceptual, de aparición de blur por desenfoque en imágenes.

Se asume que el ruido η sigue una distribución gaussiana de media 0 y una desviación estándar dada por σ , $N(0, \sigma^2)$. De manera conceptual, la ecuación 3.1 se puede expresar como muestra la Figura 3.3.

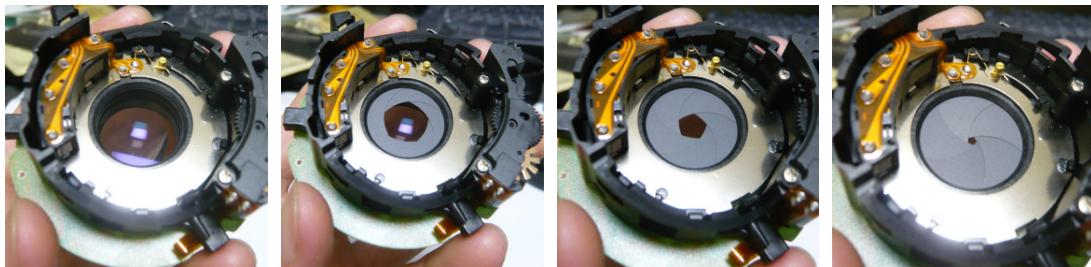


Figura 3.4: Apertura circular con distintos tamaños de diafragma.

En las cámaras fotográficas (ya sean analógicas o digitales) la apertura tiene forma aproximadamente circular, como se puede ver en la Figura 3.4. En el dominio de la frecuencia, la respuesta de una apertura circular disminuye la amplitud de las altas frecuencias en la imagen obtenida, lo que implica la pérdida de los detalles en la escena, causando o incrementando el blur en la imagen (como se verá a lo largo de este trabajo); por lo que el paso de la ecuación 3.1 al dominio de la frecuencia facilitará el análisis de dicha respuesta.

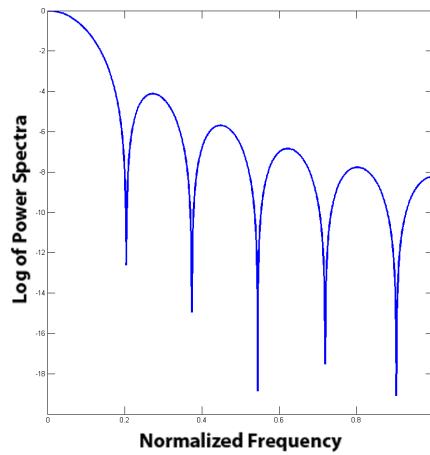


Figura 3.5: Respuesta en el dominio de la frecuencia de una apertura circular. El eje X muestra frecuencia normalizada, nótese que la escala del eje Y es logarítmica.

En la Figura 3.5 se puede ver la respuesta, en el dominio de la frecuencia de una apertura circular, se trata del espectro de potencia (*power spectrum*). Éste muestra el modo en que la

apertura va a afectar a la amplitud de la señal (la imagen capturada) en cada una de sus frecuencias. En el caso de una apertura circular, además de existir una disminución de la señal en las altas frecuencias, se observan numerosos pasos por cero para varias frecuencias (picos verticales). Esto impide, una vez que existe blur en la imagen capturada, recuperar dicha pérdida de información como se verá en la Sección 5.3. La disminución de amplitud empeora además el SNR de la señal, dificultando la recuperación de la información perdida.

3.2. Trabajo previo

En el comienzo de este proyecto se estudiaron las distintas causas del blur, intentando abordar el problema desde distintos enfoques, con el objetivo principal de recuperar la mayor parte de la información perdida en una imagen capturada. La mayoría de los trabajos se basan en el operador de *deconvolución* para recuperar la información perdida; habiendo cuatro metodologías principales no excluyentes entre sí:

- métodos de probabilidad Bayesiana,
- métodos iterativos basados en alcanzar una solución de máxima probabilidad (*maximum likelihood*),
- métodos de deconvolución a ciegas (*blind deconvolution*),
- aperturas codificadas.

Entre los métodos Bayesianos destaca el algoritmo Pixon utilizado, sobretodo, en aplicaciones astronómicas. Entre los métodos iterativos cabe destacar el algoritmo Richardson-Lucy, basado en alcanzar una solución de máxima probabilidad. Los métodos de deconvolución a ciegas parten del desconocimiento de la apertura utilizada, pretendiendo extraer una apertura válida a partir de la información existente en la imagen. Sin embargo, con el paso de los años aparecieron nuevos enfoques, basados también en el operador de deconvolución, como es la introducción de una apertura codificada en el sistema de lentes de una cámara.

Los métodos anteriores trabajan añadiendo información a las imágenes capturadas en las que parte de la información se ha perdido de manera irreversible. Mediante el estudio de las estadísticas de las imágenes o la obtención de una PSF consiguen recomponer parte de esa información perdida. Siguiendo una línea de acción distinta, se comenzó a investigar de qué manera se podía modificar la apertura de la cámara a la hora de capturar una imagen, entendiendo que es una de las principales causas de desenfoque. La idea era obtener nuevas aperturas que obtuviesen una respuesta en el dominio de la frecuencia favorable y, de este modo, poder recuperar la mayor parte posible de información perdida en la imagen. Este fue el nacimiento de lo que se conoce por *aperturas codificadas para defocus deblurring*.

Tradicional y principalmente, las aperturas codificadas se han utilizado en astronomía, codificando la dirección de los rayos incidentes, siendo posiblemente los patrones MURA (Modified Uniformly Redundant Array) [8], cuya respuesta en el dominio de la frecuencia tiene una forma casi plana, los más conocidos. En los últimos años se han empezado a utilizar, en el campo de la

fotografía computacional, las aperturas codificadas como herramienta para distintas aplicaciones [11, 15, 22, 24, 30, 35], incluyendo el objetivo de este trabajo de recuperar información perdida en el proceso de captura de las imágenes.

3.3. Solución adoptada

Este trabajo se ha basado en las técnicas que codifican la información con el uso de aperturas codificadas, utilizando modelos matemáticos que buscan minimizar la pérdida de información al capturar la imagen. La idea principal de estas técnicas consiste en entender que la información aparentemente perdida (zonas borrosas que impiden visualizar los objetos de la escena en esas zonas), en la imagen con blur, se puede recuperar (eliminar esas zonas borrosas permitiendo ver los objetos de la escena con claridad) si se sabe codificar durante el proceso de captura. A día de hoy sólo se ha conseguido llegar a una aproximación (con errores) de la escena real, siendo imposible reproducir dicha escena de manera exacta. Estos errores son debidos, por un lado, a las características inherentes al proceso de captura de la imagen y, por otro lado, al ruido añadido en dicho proceso.

De entre los trabajos que han abordado el problema de *defocus deblurring* mediante aperturas codificadas algunos utilizan varias imágenes [11], otros varias aperturas [35], e incluso trabajan a la vez en recuperación de profundidad de la escena e imagen enfocada [15]. Recientemente se presentó un método para diseñar aperturas codificadas casi óptimas para el problema de *defocus deblurring* [36]. Este método obtiene una métrica objetiva para evaluar la calidad de una apertura codificada, teniendo en cuenta las estadísticas de las imágenes naturales. Este trabajo, que permite recuperar una aproximación a la escena enfocada a partir de una sola imagen, constituye la base del presente proyecto.

3. Defocus deblurring

4. Obtención de aperturas codificadas óptimas

Se ha visto, en la sección anterior, cómo la forma de la apertura circular es la principal causa del blur. Por ello, los métodos de deblurring basados en aperturas codificadas buscan obtener una apertura cuya respuesta sea óptima en el dominio de la frecuencia, de cara a eliminar ese blur.

C. Zhou y S. Nayar proponen una métrica que, dada una apertura codificada, indica la calidad esperada a la hora de recuperar la información perdida en una imagen con blur por desenfoque. Es decir, se espera que la apertura que obtenga el menor valor para dicha métrica tenga una mejor respuesta en frecuencias (con respecto a otras aperturas). La derivación matemática de esta métrica, que se muestra a continuación, puede encontrarse en el Apéndice A:

$$R(K) = \frac{\sigma^2}{|K|^2 + \sigma^2/A} , \quad (4.1)$$

donde:

- $R(K)$ da como resultado una medida de la calidad de una apertura (a menor R mayor calidad),
- K es la respuesta de la apertura utilizada en el dominio de la frecuencia,
- σ es la desviación estándar que sigue el ruido de la imagen,
- A es el promedio del espectro de potencia de un conjunto de imágenes naturales.

Así pues, teniendo una métrica cuyo valor indica (de manera inversamente proporcional) la calidad de una apertura, el objetivo es hallar la apertura K cuyo $R(K)$ sea el menor posible. Se plantea entonces el problema como uno de optimización.

4.1. Elección del método de optimización

Dada una función que indica la calidad de una apertura a la hora de recuperar información perdida en una imagen con blur por desenfoque, las aperturas codificadas se obtienen por medio de un método de optimización que utiliza la ecuación 4.1 como función objetivo. La mayoría de los métodos de optimización necesitan una función derivable para avanzar en la búsqueda de la solución con menor resultado, pero dadas las características de la ecuación 4.1 (i.e. no es derivable analíticamente) y el espacio de búsqueda en este caso (número de posibles soluciones)

hacen que se decidiese utilizar un *algoritmo genético* como método de búsqueda de la mejor solución.

Un algoritmo genético es un tipo de algoritmo basado en la evolución biológica y la selección natural. Este tipo de algoritmo fue desarrollado por John Henry Holland en 1975 [12], interesando en la adaptación natural como una de las principales herramientas para dar lugar a mejores individuos en una especie con el paso del tiempo. Basado en la teoría de la evolución establecida por Darwin en 1859, decidió imitar los procesos adaptativos de los sistemas naturales mediante el diseño de sistemas artificiales. De esta manera, nacen los algoritmos genéticos como herramientas de resolución de problemas de búsqueda y optimización. Los principales problemas de los algoritmos de búsqueda y optimización, en general, son la complejidad de la función objetivo y la costosa exploración del espacio de búsqueda puesto que, en algunos casos, no existe una solución única (i.e. mínimo en una función periódica). La aportación que hacen los algoritmos genéticos a ese problema es su robustez, pues resultan menos afectados a falsas soluciones (i.e. mínimos locales) que los métodos tradicionales, y su uso con funciones de evaluación con derivaciones complejas.

4.2. Implementación de un algoritmo genético

Como se ha avanzado en la sección anterior, los algoritmos genéticos son métodos basados en el proceso evolutivo de los organismos vivos. En el mundo real, los individuos de una especie van evolucionando dando lugar a poblaciones de acuerdo con el principio de la selección natural y la supervivencia de los más fuertes postulado por Darwin. Por imitación a este proceso, la implementación de un algoritmo genético exige establecer una descripción de los individuos a evolucionar, codificando los datos para poder tratar con ellos de una manera más o menos sencilla.

Para ello, a partir de ahora, se va a denominar al conjunto de individuos de la población, que serán los posibles soluciones al problema de optimización, *fenotipo* y a cada uno de los individuos *cromosoma*. Generalmente los cromosomas se codifican como cadenas de símbolos o *genes*. Para el caso concreto (como en este trabajo) en el que los genes son dígitos binarios, denominaremos a la población de cromosomas como *genotipo*. La evolución de los cromosomas va a realizarse a través de iteraciones o *generaciones* utilizando dos operadores genéticos, el operador *cruce* y el operador *mutación* durante la ejecución del algoritmo. Los cromosomas van evolucionando hasta que se llega a un número de cromosomas máximo, se utiliza una función de *evaluación* o función de selección (la función objetivo del problema de minimización) para determinar los mejores cromosomas en dicha generación; los cromosomas que sobreviven dan lugar a la siguiente generación, simulando el proceso de selección natural.

Una vez definidos todos los elementos de un algoritmo genético, se puede hacer una primera división del algoritmo en *inicialización* y *generaciones*. De esta manera, se define cada una de las partes por la consecución de una serie de pasos que se explican a continuación.

- **Inicialización.** Se establece un primer fenotipo aleatorio (población inicial) de N cromosomas y, a continuación, se otorga una puntuación para cada cromosoma generado mediante el uso de la función de evaluación $R(K)$ establecida.

4. Obtención de aperturas codificadas óptimas

- **Generaciones.** Se realizarán los siguientes pasos mientras no se cumpla la condición de parada:

- Selección: elección de los M mejores cromosomas en base a la puntuación obtenida con la función de evaluación $R(K)$.
- Cruce: operador genético que combina los genes de dos cromosomas para la creación de un cromosoma nuevo.
- Mutación: operador genético que modifica los genes de un cromosoma para la creación de un cromosoma nuevo.
- Evaluación: de nuevo, se otorga una puntuación para cada cromosoma del fenotipo actual mediante el uso de la función de evaluación $R(K)$ establecida.

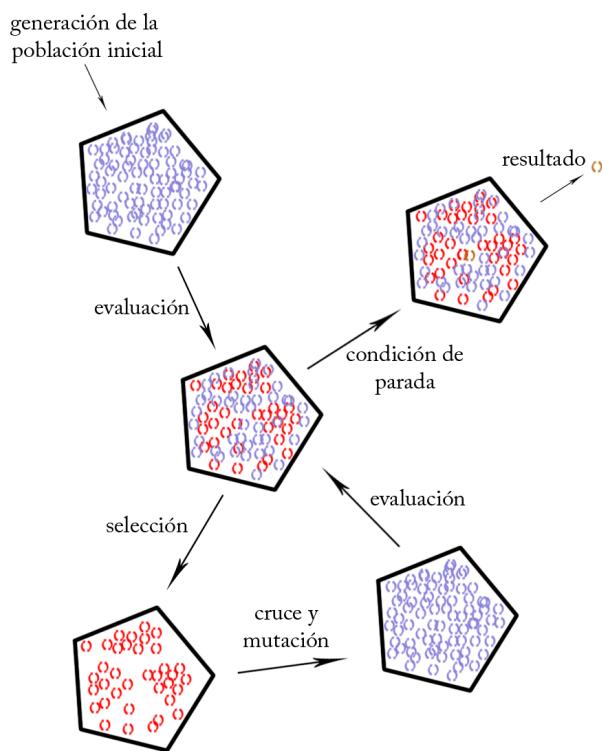


Figura 4.1: *Esquema, conceptual, de un algoritmo genético.*

4.2.1. Representación

La representación de los cromosomas se ha realizado mediante vectores de L elementos ($x_1, x_2, x_3, \dots, x_L$). Dependiendo de los valores posibles para esos L elementos (genes), se puede realizar una clasificación general de la siguiente manera:

- Representación binaria: genes con valor cero o uno.
- Representación entera: genes con valor entero.
- Representación real: genes con valor real.

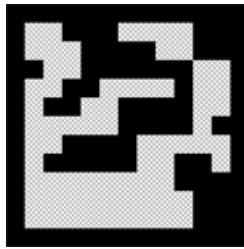


Figura 4.2: *Ejemplo de apertura codificada binaria. Dos tipos de píxeles, negros (opacos) y grises (transparentes).*

Para las aperturas codificadas se ha elegido una representación binaria. Teniendo en cuenta que el tamaño de los cromosomas condicionará el tiempo de convergencia del algoritmo a una solución válida, cada cromosoma se representa como un vector de n^2 elementos (genes), de manera que al mostrarlo en forma de matriz conforme una matriz $n \times n$, ya que se busca obtener una apertura codificada. De esta manera, el gen x en un cromosoma cualquiera será el píxel (x_1, y_1) de la apertura donde $x_1 = (x/n) + 1$ y $y_1 = \text{mod}(x/n)$. Como se puede ver en la Figura 4.2, el valor posible para un gen tiene el siguiente significado:

- Gen cero o gen negro: no deja pasar la luz a su través (como si de un objeto opaco se tratase).
- Gen uno o gen transparente: deja pasar perfectamente la luz a su través (como si de un objeto perfectamente transparente se tratase).

De esta manera, se puede hacer un pequeño análisis del número de posibles soluciones para la representación elegida. Teniendo en cuenta el tamaño de los cromosomas (n^2 genes) y los posibles valores para cada uno de ellos (cero o uno), se puede determinar que el número de soluciones posibles es 2^{n^2} .

4.2.2. Operadores genéticos: cruce y mutación

Los operadores genéticos son funciones empleadas en los algoritmos genéticos para mantener la diversidad genética en cada generación, siendo necesarios para obtener una evolución correcta de los individuos de la población. El operador de cruce equivale a la reproducción sexual y el operador de mutación equivale a la mutación biológica.

El operador de *cruce* es un operador binario; es el encargado de generar dos nuevos cromosomas a partir de una pareja de cromosomas base. Tal y como se ve en la Figura 4.3, el procedimiento es recorrer (gen a gen) los cromosomas base intercambiando, con una cierta probabilidad c_1 , los genes entre la pareja de cromosomas.

Cromosomas base de 17 genes	Cruce de genes marcados	Nuevos cromosomas generados
1 0 1 1 0 0 1 1 0 0 0 1 1 0 0 1 1	1 0 1 1 0 0 1 1 0 0 0 1 1 0 0 1 1 x x x x x x	1 0 1 1 0 1 1 0 1 1 0 1 1 0 0 1 1
0 0 0 1 0 1 1 1 0 1 0 1 1 0 0 1	0 0 0 1 0 1 1 1 0 1 0 1 1 0 0 1	0 0 0 1 0 0 1 1 0 0 0 1 0 0 1 1

Figura 4.3: *Operador binario de cruce.*

4. Obtención de aperturas codificadas óptimas

El operador de *mutación*, sin embargo, es un operador unario; es el encargado de añadir variabilidad al genotipo. En este caso, como se ve en la Figura 4.4, el procedimiento es recorrer (de nuevo, gen a gen) el cromosoma a mutar cambiando, con una cierta probabilidad c_2 , el valor de sus genes (de cero a uno, o viceversa).

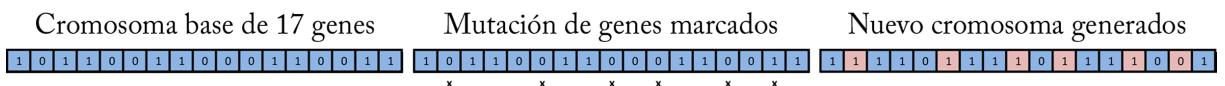


Figura 4.4: *Operador unario de mutación*.

4.2.3. Condición de parada

La alternativa que mejor determina la detención del algoritmo es establecer un número máximo de generaciones. Hay que entender que hay otro tipo de condiciones de parada, como la convergencia del algoritmo a un valor o la no evolución del cromosoma “mejor” a lo largo de varias generaciones. Así, se podrían añadir otras condiciones de parada, pero dada la naturaleza del problema que ocupa este trabajo es muy complicado determinar una condición de parada robusta (en términos de error relativo de la solución).

4.3. Convergencia del algoritmo

Una vez implementado el algoritmo genético, encargado de generar aperturas codificadas que después se iban a utilizar a la hora de recuperar información perdida en imágenes con blur por desenfoque, se creyó conveniente realizar un estudio acerca de la convergencia del algoritmo en base a sus parámetros (población, número de supervivientes, número de generaciones, probabilidad de cruzamiento y probabilidad de mutación).

Para ello se estableció un conjunto de ejecuciones entre las cuales la única variación era el valor de uno, y sólo uno, de los parámetros de llamada con el fin de interpretar la influencia de los mismos en los resultados finales (siempre con el mismo valor para el parámetro “ruido”, i.e. σ en la ecuación 4.1). Se propuso realizar cinco ejecuciones para diez valores distintos para cada parámetro, obteniendo 250 aperturas ($250 = (5 \text{ ejecuciones} * 10 \text{ valores}) * 5 \text{ parámetros}$).

De esta manera, se obtuvieron los valores alcanzados por la función de evaluación $R(K)$ durante la sucesión de generaciones, así como los tiempos empleados para cada ejecución. A la vista de los resultados, que se muestran en las Figuras 4.5 y 4.6, y en el Apéndice C, se decidieron utilizar los siguientes valores para los parámetros del algoritmo: 4000 individuos conformando el *genotipo*, 400 *supervivientes* en cada generación, 80 *generaciones*, probabilidad de *cruzamiento* de 0.2 y probabilidad de *mutación* de 0.05.

4. Obtención de aperturas codificadas óptimas

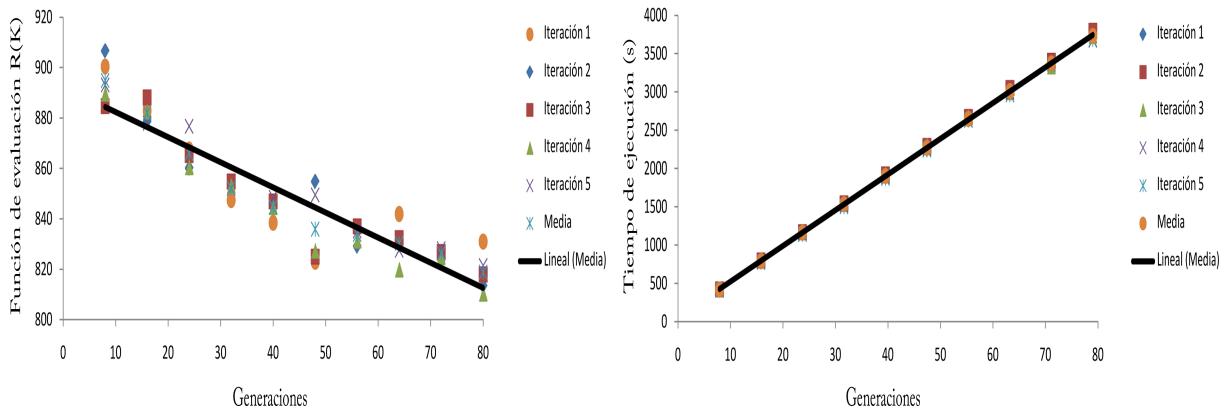


Figura 4.5: Izquierda, eje horizontal: *distintos valores para el parámetro generaciones*, eje vertical: *tendencia de la función de evaluación $R(K)$ con la variación del parámetro generaciones*. Derecha, eje horizontal: *distintos valores para el parámetro generaciones*, eje vertical: *tendencia del tiempo de ejecución del algoritmo con la variación del parámetro generaciones*.

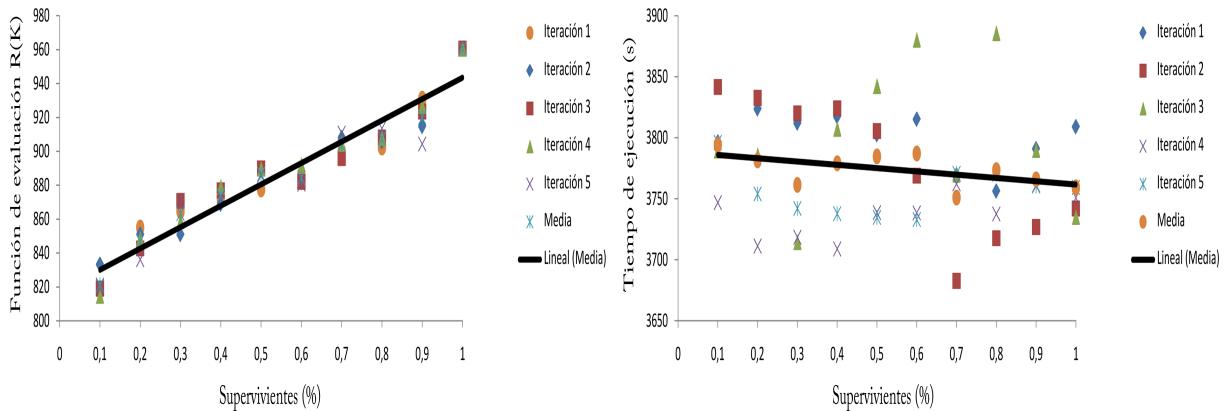


Figura 4.6: *Supervivientes (porcentaje con respecto al número de individuos de la población)*. Izquierda, eje horizontal: *distintos valores para el parámetro supervivientes*, eje vertical: *tendencia de la función de evaluación $R(K)$ con la variación del parámetro supervivientes*. Derecha, eje horizontal: *distintos valores para el parámetro supervivientes*, eje vertical: *tendencia del tiempo de ejecución del algoritmo con la variación del parámetro supervivientes*.

4.4. Aperturas obtenidas

A la hora de obtener una apertura codificada mediante la utilización del algoritmo genético, uno de los factores más importantes a tener en cuenta es el valor de *sigma* (es decir, la desviación estándar del ruido de la imagen) para el cual se quiere obtener la misma. Dada la ecuación 4.1, que define la función de evaluación utilizada, se puede ver que el valor de *sigma* determinará el tipo de apertura codificada obtenida. En este trabajo se eligieron distintos valores de *sigma*, cubriendo un amplio rango de soluciones, y para cada uno se realizaron varias ejecuciones del algoritmo genético (siete exactamente) con el objetivo de seleccionar la apertura con mejor puntuación en cada caso. En la Figura 4.7 se muestran las mejores aperturas obtenidas para distintos valores de *sigma*.

Además, como se anunciaba en la introducción de este documento, con el objetivo de explo-

4. Obtención de aperturas codificadas óptimas

rar el rendimiento obtenido para aperturas con distintas características, se obtuvieron aperturas con distinta resolución. Como se puede ver en la Figura 4.8, se obtuvieron aperturas de tamaño $7*7$ y de tamaño $20*20$ para dos valores representativos de σ .

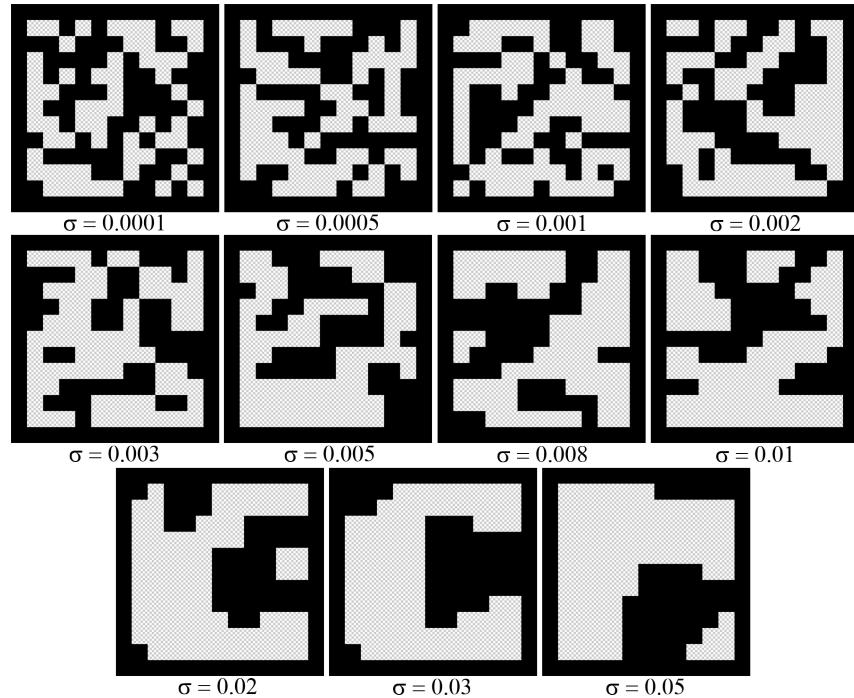


Figura 4.7: Aperturas $11*11$ obtenidas para distintos valores de σ .

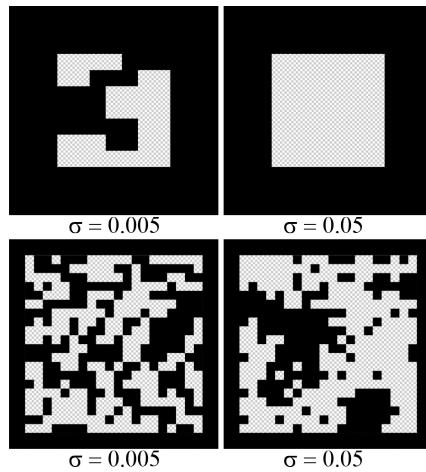


Figura 4.8: Arriba: aperturas $7*7$ obtenidas con el correspondiente valor de σ . Abajo: aperturas $20*20$ obtenidas con el correspondiente valor de σ .

4. Obtención de aperturas codificadas óptimas

5. Validación I: Simulaciones

Una vez obtenido un conjunto de aperturas para distintos niveles de ruido el siguiente paso es evaluarlas. El método explicado en este capítulo permite simular la captura de una imagen desenfocada a partir de una imagen enfocada para, a posteriori, re-enfocarla y poder comparar el resultado obtenido con la imagen inicial. Esta simulación supone el primer paso en la validación de las aperturas obtenidas, al confirmar que consiguen mejores resultados que una apertura circular.

5.1. Simulación del proceso de captura y recuperación de la información de la imagen

El procedimiento desarrollado se puede ver en la Figura 5.1. Partiendo de una imagen totalmente enfocada f_0 , se obtiene la imagen f simulando el proceso de captura de una fotografía desenfocada (con una apertura k). Como segunda parte del proceso se parte de la imagen desenfocada f , a partir de la cual se obtiene la imagen re-enfocada \hat{f}_0 (con la misma apertura k). Finalmente, como análisis del resultado obtenido para la apertura k , se mide el error mediante la utilización de una métrica objetiva de comparación entre f_0 y \hat{f}_0 . Los resultados finales de esta métrica se calcularon realizando la media entre los resultados obtenidos para un conjunto de diez imágenes distintas.

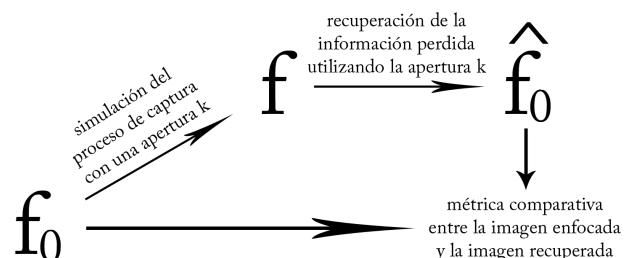


Figura 5.1: De izquierda a derecha: *proceso de desenfoque y re-enfoque de una imagen utilizando una apertura k , que permite validar dicha apertura*.

5.1.1. Proceso de captura o desenfoque de la imagen

Conviene recordar la ecuación 3.1 que define una imagen desenfocada:

$$f = f_0 * k_d + \eta, \quad (5.1)$$

donde f es la imagen desenfocada, f_0 es dicha imagen perfectamente enfocada (i.e. la escena real), k_d es la respuesta de la apertura utilizada a una profundidad d y η es el ruido de la imagen.

La ecuación 5.1 define el proceso real ocurrido en la toma de una fotografía con una cámara (analógica o digital) en el dominio espacial. Como se puede ver, interviene el operador *convolución* entre la imagen perfectamente enfocada y la apertura existente en la cámara fotográfica en el momento de la toma. Si todos los parámetros de la ecuación 5.1 son llevados al dominio de la frecuencia, haciendo uso de la transformada de Fourier, la convolución pasa a ser una multiplicación:

$$F = F_0 \cdot K_d + \zeta \quad (5.2)$$

En la imagen 5.2 se puede ver la simulación del desenfoque (derecha) de una imagen totalmente enfocada (izquierda), mediante la utilización de la ecuación 5.2 (es decir, la obtención de f). La apertura utilizada se normaliza con respecto a una apertura circular, simulando la pérdida de luz en la captura real, lo que produce el oscurecimiento en la imagen desenfocada.

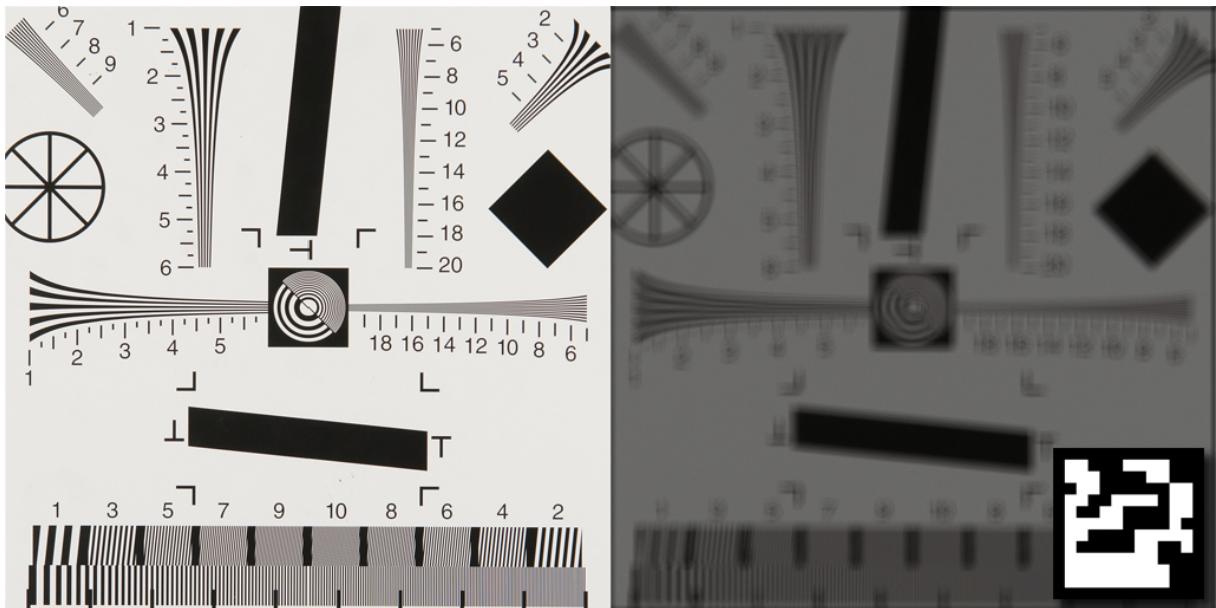


Figura 5.2: Izquierda: *imagen enfocada f_0* . Derecha: *imagen f obtenida mediante la ecuación 5.2, simulando una captura mediante el uso de la apertura mostrada en la esquina inferior derecha, apertura óptima obtenida para $\sigma = 0,005$* .

5.1.2. Recuperación de la información de la imagen

En este caso, partiendo de una imagen desenfocada f , se pretende conseguir una imagen \hat{f}_0 idéntica a la escena original capturada f_0 (en este caso conocida).

Haciendo uso de la formulación establecida para la deconvolución de Wiener [27]:

$$\hat{F}_0 = \frac{F \cdot K}{|K|^2 + |C|^2} \quad (5.3)$$

donde \hat{F}_0 es la imagen recuperada; F es la imagen desenfocada y K es la apertura utilizada (todas ellas en el dominio frecuencial). C es la relación de potencia señal-ruido y se puede obtener como σ^2/A siendo A la media del espectro de potencia de un conjunto de imágenes naturales (ver Apéndice A). Notar que el tamaño (en píxeles) de las variables en la ecuación 5.3 son, por norma general, distintos ya que las aperturas tienen tamaño 11*11 y, en la mayoría de los casos, las imágenes a tratar poseen un tamaño mayor. Por ello, es necesario realizar un *padding* o *relleno* a la variable de menor tamaño, en este caso a la apertura K utilizada. Esta operación de relleno consiste en insertar la apertura en el centro de una matriz de ceros, del tamaño de la imagen a tratar.

En la imagen 5.3 se puede ver la recuperación de la información (derecha) de una imagen desenfocada (izquierda), mediante la utilización de la ecuación 5.3 (es decir, la obtención de \hat{f}_0).

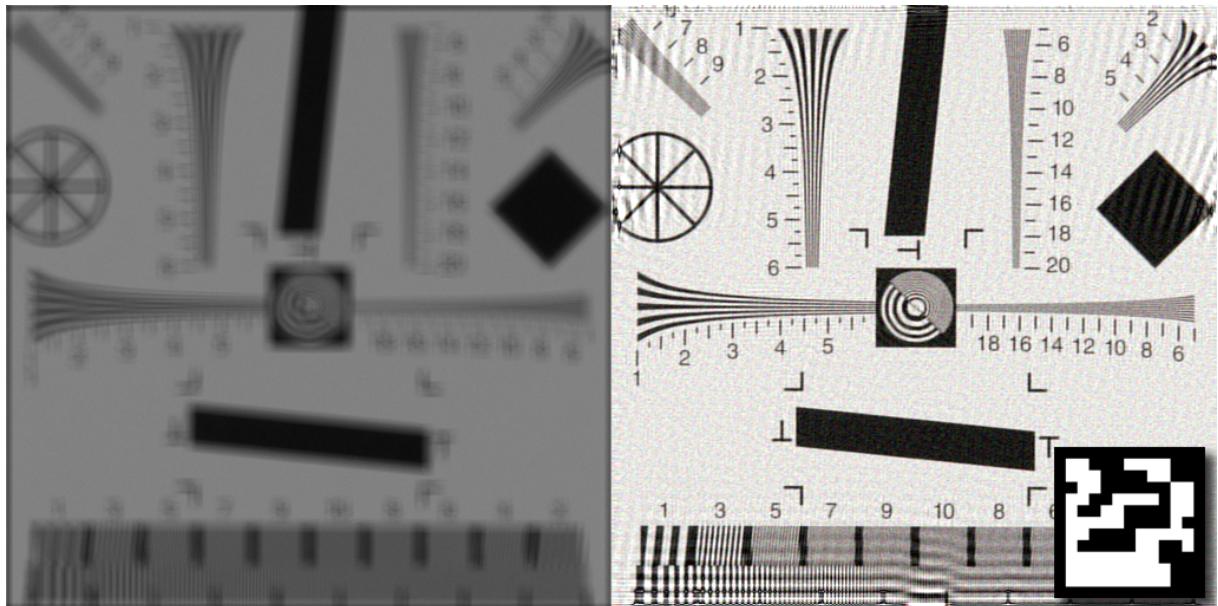


Figura 5.3: Izquierda: *imagen desenfocada* f . Derecha: *imagen* \hat{f}_0 *obtenida tras la recuperación de la información, mediante la ecuación 5.3, usando la apertura mostrada en la esquina inferior derecha, apertura óptima obtenida para* $\sigma = 0,005$.

Como se ve en la Figura 5.3 la imagen recuperada \hat{f}_0 no está perfectamente enfocada pero los detalles (i.e. dígitos), que en la imagen desenfocada f no son apreciables, son perfectamente legibles. Además se aprecia la existencia de artefactos en la imagen recuperada \hat{f}_0 en forma de bandas negras, esto es debido a la operación de deconvolución.

5.2. Análisis del error obtenido: Norma L_2 e índice SSIM

Con el objetivo de establecer una cuantificación para el error cometido al recuperar la información perdida en la captura de las imágenes es necesario utilizar una métrica objetiva. Se han evaluado dos métricas distintas, que se detallan a continuación.

5.2.1. Norma L_2

Se decidió utilizar, en primera instancia, la norma (o distancia euclídea) L_2 entre la imagen recuperada \hat{f}_0 tras la deconvolución y la escena real (imagen totalmente enfocada) f_0 , que para imágenes de tamaño $n \cdot m$ se define como:

$$d(\hat{f}_0, f_0) = \sum_{i=1, j=1}^{n \cdot m} \|\hat{f}_0(i, j) - f_0(i, j)\|^2 \quad (5.4)$$

Cada píxel en una imagen posee un valor entre 0 (totalmente negro) y 1 (totalmente blanco), de tal manera que la distancia máxima entre dos píxeles es 1. Como se puede ver en la Figura 5.4, la distancia L_2 entre dos imágenes opuestas (una imagen cuyos píxeles tienen, todos, valor 0 y otra imagen cuyos píxeles tienen, todos, valor 1), obtiene como resultado la distancia máxima posible o error máximo posible (tamaño en píxeles de las imágenes). En cambio, calcular la distancia L_2 entre dos imágenes iguales obtiene como resultado la distancia mínima posible o error mínimo posible (cero).

$$d(\boxed{}, \boxed{\begin{smallmatrix} m \\ n \end{smallmatrix}}) = m \cdot n \quad d(\boxed{}, \boxed{}) = 0$$

Figura 5.4: Izquierda: *norma (o distancia) L_2 entre dos imágenes totalmente opuestas*. Derecha: *norma L_2 entre dos imágenes totalmente iguales*.

Se calculó, de este modo, el resultado para la norma (o distancia) L_2 obtenido para cada apertura (en función del valor de σ). En la Figura 5.5 se puede ver el porcentaje del error obtenido, con respecto al error máximo posible, para cada una de las aperturas.

σ	0.0001	0.0005	0.001	0.002
L_2	0,658	0,636	1,344	0,684
σ	0.003	0.005	0.008	0.01
L_2	0,777	0,691	0,814	0,836
σ	0.02	0.03	0.05	circular
L_2	0,925	1,284	2,553	1,076
σ	0.005 (7x7)	0.05 (7x7)	0.005 (20x20)	0.05 (20x20)
L_2	0,620	2,221	1,222	2,875

Figura 5.5: *Resultados de la norma L_2 para aperturas calculadas para diferentes valores de σ . Se muestran los porcentajes con respecto al error máximo posible. Las mayoría de las aperturas obtenidas dan lugar a un menor error con respecto al máximo error posible en comparación con el error obtenido por una apertura circular.*

En la Figura 5.5 se puede ver que la mayoría de las aperturas obtenidas obtienen un porcentaje menor de error (norma L_2), en el proceso de recuperación de información, que una apertura circular. Las que tienen un porcentaje de error mayor son las calculadas para valores de σ elevados, muy superiores al nivel de ruido de las imágenes utilizadas en la validación.

5.2.2. Índice $SSIM$

Con el objetivo de verificar la validez de los resultados obtenidos con la norma L_2 se decidió utilizar otra métrica, eligiendo para ello la medida $SSIM$ por su aceptación entre los investigadores en el campo de la informática gráfica. La medida $SSIM$ [31] se desarrolló como mejora de métodos tradicionales de comparación de imágenes ya que, en teoría, tiene en cuenta la percepción humana. Normalmente se calcula entre conjuntos de píxeles de las imágenes a comparar (ventanas de tamaño $n \cdot n$). Para dos ventanas (x e y), se define como:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (5.5)$$

donde, μ_x, μ_y son la media, σ_x^2, σ_y^2 es la varianza, σ_{xy} es la covarianza de x y y respectivamente y c_1, c_2 son constantes obtenidas por el rango dinámico L de las imágenes que se van a comparar (donde $L = 2^{\#bits \text{ por pixel}} - 1$). En este caso, como se puede ver en la Figura 5.6, el índice $SSIM$ entre dos imágenes opuestas, obtiene como resultado el menor valor posible (-1). En cambio, el índice $SSIM$ entre dos imágenes iguales obtiene como resultado el mayor valor posible (uno).

$$d(\blacksquare, \square) = -1 \quad d(\square, \square) = 1$$

Figura 5.6: Izquierda: índice $SSIM$ entre dos imágenes totalmente opuestas. Derecha: índice $SSIM$ entre dos imágenes totalmente iguales.

Se calculó, de este modo, el resultado para la medida $SSIM$ obtenido para cada apertura (en función del valor de σ). En la Figura 5.7 se pueden ver los índices obtenidos para cada una de las aperturas.

σ	0.0001	0.0005	0.001	0.002
SSIM	0,9424	0,9448	0,8865	0,9415
σ	0.003	0.005	0.008	0.01
SSIM	0,9301	0,9385	0,9221	0,9207
σ	0.02	0.03	0.05	circular
SSIM	0,8833	0,8171	0,6506	0,9020
σ	0.005 (7x7)	0.05 (7x7)	0.005 (20x20)	0.05 (20x20)
SSIM	0,9423	0,6819	0,9294	0,6487

Figura 5.7: Resultados de la medida $SSIM$ para aperturas calculadas para diferentes valores de σ . Se muestran los valores obtenidos para dicha medida (índice $SSIM$). La mayoría de las aperturas obtenidas dan lugar a un índice $SSIM$ mayor en comparación con el índice obtenido por una apertura circular.

En la Figura 5.7 se puede ver que la mayoría de las aperturas obtenidas obtienen un índice mayor (medida *SSIM*), en el proceso de recuperación de información, que una apertura circular. Estos resultados, coherentes con los obtenidos al aplicar la norma L_2 , indican que las aperturas que mejor van a funcionar son las aperturas para valores de sigma de 0.0001, 0.0005, 0.002 y 0.005 entre las aperturas de tamaño 11×11 y la apertura para valor de sigma 0.005 con tamaño 7×7 . Los resultados, con fotografías reales, se verán en el capítulo siguiente.

5.3. Espectro de potencia

En la Sección 3.1 se explica que la información de detalle de una imagen se corresponde con las altas frecuencias de la misma. De este modo, la pérdida de energía en dichas frecuencias es uno de los motivos por los que una imagen aparece desenfocada. La respuesta de una apertura circular al capturar una imagen produce gran pérdida de energía en estas frecuencias (empeorando el SNR), además de numerosos pasos por cero en distintas frecuencias (dificultando o, incluso, impidiendo la recuperación de la información mediante la deconvolución en esas frecuencias). Para analizar esta respuesta en frecuencias se propone otro análisis de las aperturas obtenidas. Dada una señal $f(x, y)$ en el dominio espacial, se puede trasladar al dominio de la frecuencia mediante el uso de la transformada de Fourier. En el dominio de la frecuencia, dicha señal se puede representar en términos de su magnitud y fase:

$$F(u, v) = |F(u, v)| e^{j\phi(u, v)}, \quad (5.6)$$

donde $|F(u, v)|$ define la amplitud de la señal y ϕ define la fase; siendo $|F(u, v)|$ calculado de la siguiente manera:

$$|F(u, v)| = [R^2(u, v) + I^2(u, v)]^{1/2} \quad (5.7)$$

Una manera de cuantificar la energía que posee una señal para cada una de sus frecuencias es calcular su espectro de potencia. El espectro de potencia de una señal se define como:

$$P(u, v) = |F(u, v)|^2 = R^2(u, v) + I^2(u, v) \quad (5.8)$$

Aplicando la ecuación 5.8 a una apertura cualquiera se puede conocer su respuesta en el dominio de la frecuencia, como se muestra en la Figura 5.8 en escala logarítmica (debido a las grandes diferencias entre los órdenes de magnitud a representar posibilitando la comparación entre las aperturas a analizar).

Como se puede ver en la Figura 5.8, las aperturas obtenidas tienen una mejor respuesta en frecuencias que una apertura circular. Se puede observar que la amplitud de la señal es mayor, a lo largo de todas las frecuencias (eje X), en las aperturas codificadas obtenidas que en una apertura circular. Además, si se observa la respuesta en las altas frecuencias, la diferencia de amplitud entre las aperturas codificadas y una apertura circular es lo suficientemente grande como para permitir recuperar los detalles (i.e. dígitos) en una imagen con las aperturas codificadas, algo impensable utilizando una apertura circular. Finalmente, las señal del espectro de potencia de las

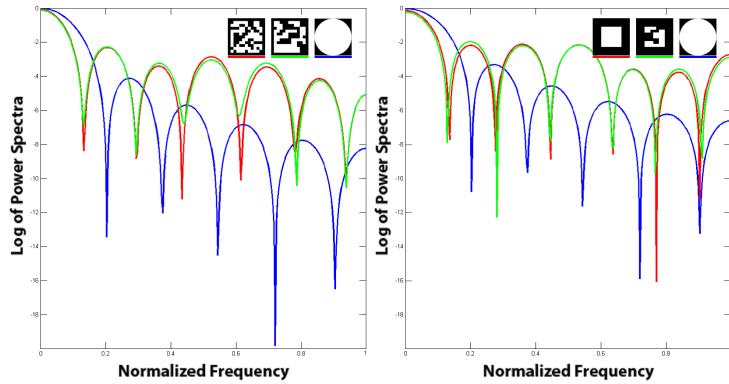


Figura 5.8: Espectro de potencia para cuatro aperturas codificadas y una apertura circular. Las aperturas codificadas obtenidas poseen mejor respuesta, es decir mayor amplitud (energía) a lo largo de todas las frecuencias, que una apertura circular. El eje X muestra frecuencia normalizada, nótense que la escala del eje Y es logarítmica.

aperturas obtenidas es más suave y posee un número menor de pasos por cero, lo que posibilita recuperar mayor parte de la información perdida en una imagen que una apertura circular.

6. Validación II: Experimentos reales

Una vez realizada la validación de las aperturas mediante simulación, conforme a los explicado en la sección anterior, se trasladó el problema al mundo real. Para ello las fotografías son capturadas con una cámara fotográfica a la que se le introducen, físicamente, las aperturas codificadas.

6.1. Configuración del sistema

El primer paso del proceso es imprimir las aperturas en un material fotolitográfico, lo que se denomina un *photomask* (similar a una transparencia). En este punto se tenían varias aperturas codificadas obtenidas mediante el algoritmo genético, pero además se decidió añadir aperturas propuestas en otros trabajos previos como [15], [30] y [36]; así como un patrón MURA, una imagen femenina (la fotografía de Lena, bien conocida en el campo del procesamiento de imagen), figuras como círculos concéntricos o símbolos cinematográficos como los símbolos de Batman y Mickey Mouse. La impresión de todas las aperturas seleccionadas se muestra en la Figura 6.1.

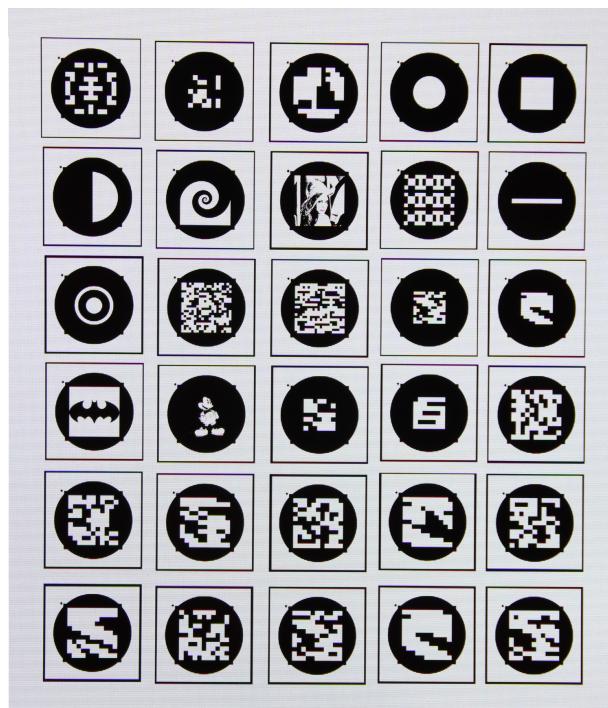


Figura 6.1: *Aperturas impresas en material fotolitográfico.*

Una vez impresas las aperturas, un punto crítico era tener que abrir el objetivo adquirido para insertar una apertura en su interior. Por ello, antes de adquirir el objetivo, se estuvo buscando información acerca de los objetivos existentes para la cámara de que se disponía; tras encontrar documentación sobre el desmontaje de uno de los objetivos en concreto, se decidió adquirirlo y así poder empezar los experimentos. En la Figura 6.2 se puede ver, a la izquierda, el equipo utilizado para los experimentos reales y, a la derecha, el resultado de insertar una apertura codificada en el objetivo. La cámara utilizada es una *Canon EOS 550D* y el objetivo un *Canon EF 50mm f/1.8*. En el Apéndice B se muestra, mediante fotografías, todo el proceso de apertura del objetivo e inserción de una apertura codificada, así como varias imágenes del objetivo con distintas aperturas insertadas en el mismo.



Figura 6.2: Izquierda: *equipo utilizado para los experimentos reales*. Derecha: *apertura codificada insertada en el objetivo*.

6.2. Calibración del sistema

Como se explica en la Sección 3.1, la PSF de una apertura codificada es la función de dispersión del punto o respuesta del sistema óptico completo (cámara y objetivo con una apertura codificada), cuando la entrada al sistema es un impulso (i.e. un punto). Calibrar el sistema implica conocer la PSF del sistema óptico completo a distintas profundidades y en distintas posiciones del plano imagen.

6.2.1. Metodología de obtención de la PSF del sistema

La PSF de una apertura codificada determinará los resultados a la hora de realizar la deconvolución de las imágenes, existiendo distintos métodos para la obtención de la PSF de una apertura:

- Siguiendo la ecuación 5.2 se obtiene que la imagen con blur, F , es igual a $F_0 \cdot K$ más el ruido de la imagen. De esta manera, si se es capaz de obtener una imagen con blur F y su pareja enfocada F_0 , se puede deducir la PSF como $K = F/F_0$. Esto implica no mover la cámara entre la toma de la imagen con blur y su pareja totalmente enfocada o, si hay movimiento entre las tomas, alinear a nivel de píxel ambas imágenes.
- Encontrar la PSF por medio de un método de optimización tomando una función de minimización apropiada (i.e. la norma L_2 entre la imagen totalmente enfocada f_0 y la

imagen recuperada \hat{f}_0 a partir de una imagen desenfocada f). Esto implica partir de una PSF lo más próxima a la PSF a obtener para que el algoritmo converja lo antes posible.

- Dado que la PSF es la respuesta del sistema óptico a un punto, otra opción era crear una luz puntual lo suficientemente intensa, para lo cual una de las mejores opciones es el uso de un diodo LED. El uso de un diodo LED, y no un láser, tiene como objetivo el evitar dañar el sensor de la cámara fotográfica al impactar el rayo de luz en el mismo.

En este trabajo se probaron los tres métodos anteriormente expuestos:

- Para el primero de ellos era imposible no mover la cámara entre las dos capturas necesarias puesto que el funcionamiento físico de la cámara lo impedía. Se procedió a tomar ambas fotografías intentando mover el sistema lo menos posible y alineando, a posteriori, ambas imágenes manualmente pero, finalmente, no se obtuvieron buenos resultados. El principal problema era la presencia de valores igual a cero, lo que producía divisiones por cero a la hora de obtener la PSF, impidiendo obtener la misma de manera correcta.
- En el caso de la utilización de un método iterativo no se consiguieron, tampoco, resultados satisfactorios; esto, añadido al gran tiempo de convergencia del algoritmo provocó desechar este método y continuar con el siguiente.
- Con el tercero de ellos, sin embargo, se obtuvo la PSF para varias aperturas codificadas, como se puede ver en la Figura 6.3 para la apertura con $\sigma = 0,005$ a diferentes profundidades ¹. Estas PSFs permitirán deconvolucionar las imágenes con blur, dando lugar a imágenes próximas a la escena real capturada, como se va mostrar en la Sección 6.4. Además, hay que notar que para que las PSFs funcionasen de manera correcta había que realizar un proceso de umbralización, eliminando parte del halo que las envuelve al capturarlas y la luz residual del ambiente (*stray light*).

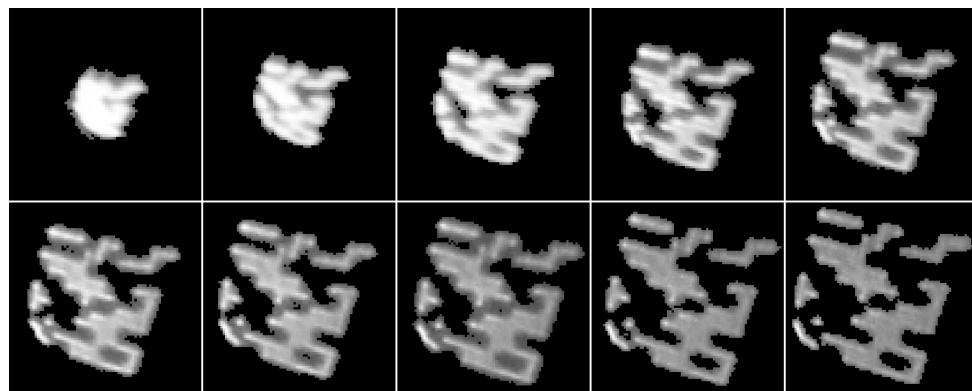


Figura 6.3: Obtención de PSFs para una apertura codificada, en intervalos de 10 centímetros, para un rango (0..1) metros de profundidad de desenfoque estando el plano focal a un metro de la cámara.

6.2.2. Construcción de un array de diodos LED

Como se ha visto anteriormente, la PSF varía con la profundidad y la posición en el plano imagen donde es capturada, siendo posible obtener la PSF de una apertura codificada para una

¹distancia entre el plano focal y la escena.

6. Validación II: Experimentos reales

distancia cualquiera y para cualesquier coordenadas de la imagen. Si se está fotografiando un objeto que ocupa, en el eje X, una gran parte de la imagen capturada, el obtener la PSF de la apertura codificada en las coordenadas correspondientes a uno de los extremos del objeto daría una recuperación distinta de la imagen en esa posición que en el extremo contrario. Por ello, idealmente, se necesitaría un “tablero” en el que se pudiesen colocar emisores puntuales de luz idénticos en cuanto a luminosidad y a distintas alturas (cubriendo todo el plano imagen a cada profundidad para poder elegir con cual de las PSFs se quiere desenfocar qué parte de la imagen desenfocada, como se muestra en la Figura 6.4. Así, dada una imagen desenfocada, se podrá elegir con cual de las PSFs se quiere recuperar qué parte de la imagen desenfocada.

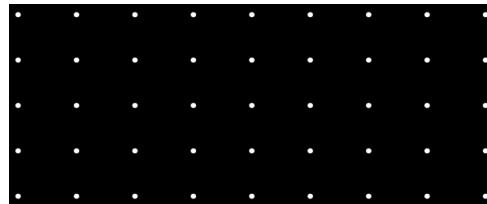


Figura 6.4: *Tablero ideal para obtener un conjunto de PSFs de una apertura codificada a una distancia cualquiera.*

Esta idea se va a denominar *array de diodos LED*. Se construyó un array de diodos LED, de una sola fila con 8 diodos LED separados entre sí 10 centímetros usando materiales “caseros” (cartón, cable de cobre y un soldador de estaño). Este array 1D fue utilizado para obtener las PSF que intervinieron en la recuperación de información de las fotografías con blur, obtenidas con la cámara. En la Figura 6.5 se muestra el array de diodos LED una vez construido. En la imagen superior se muestra la respuesta (como se puede ver es el propio punto de luz) de la apertura cuando el array de diodos LED está totalmente enfocado. En las imágenes central e inferior se muestra la respuesta de la apertura a medida que se va desenfocando la escena capturada. La metodología utilizada para las capturas se explica en la siguiente sección.

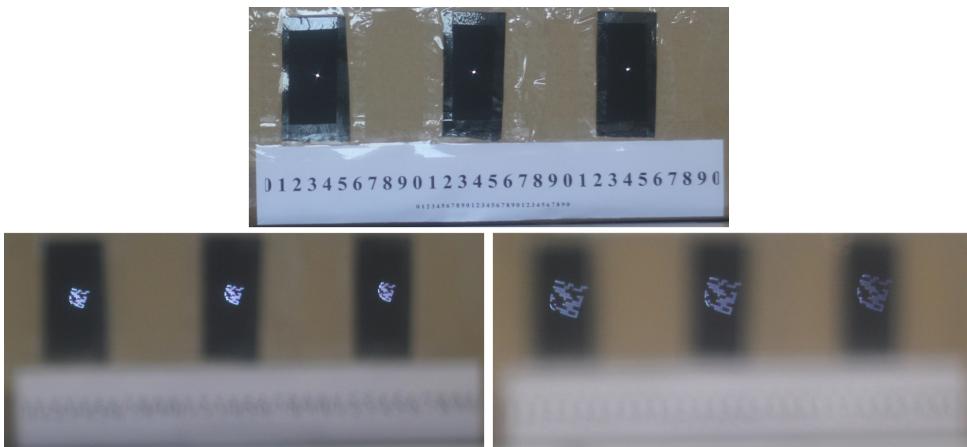


Figura 6.5: *Array de diodos LED construido. Arriba: PSFs con el array de diodos LED totalmente enfocado (puntos). Abajo: PSFs para dos desenfoques distintos del array de diodos LED, a profundidades de desenfoque de 50 centímetros y un metro.*

6.2.3. Captura de las imágenes de calibración

Una vez introducida una apertura en el objetivo de la cámara fotográfica, mediante el uso del array de diodos LED explicado, se procedió a obtener sus PSFs a distintas profundidades de desenfoque. En este trabajo, con el objetivo de obtener datos para un intervalo amplio de trabajo, se obtuvieron las PSFs para 10 distancias distintas. Esto se consiguió alejando el array de diodos LED, desde la distancia D (plano focal, escena enfocada), en intervalos de diez centímetros hasta llegar a una distancia (con respecto a la cámara) de dos metros, como se muestra en la Figura 6.6. Notar que, para hacer el proceso más preciso, se utilizó un disparador a distancia para no tener que tocar físicamente la cámara entre las distintas capturas.

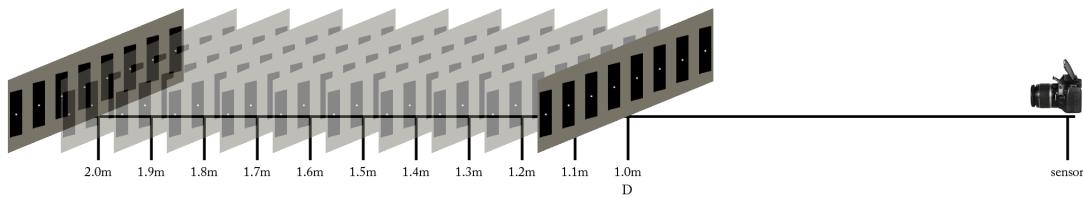


Figura 6.6: Procedimiento utilizado para obtener las PSFs de una apertura insertada en el objetivo de la cámara.

Los pasos para la obtención de las PSFs son:

- Se fija la cámara en una posición en la que pueda permanecer inmóvil durante todo el proceso.
- Se enfoca el array de diodos LED a una distancia D en el eje Z desde la posición de la cámara (en este caso $D = 1m$). Los objetos en otras posiciones en el eje Z, como se explicó en el Capítulo 3, aparecerán desenfocados a medida que se incremente su distancia con respecto a D .
- Manteniendo el enfoque a la distancia D se mueve el array de diodos LED en el eje Z para obtener la PSF de la apertura a distintas profundidades de desenfoque.

De esta manera, para cada apertura, se obtiene un conjunto de PSFs como el mostrado en la Figura 6.3.

6.3. Captura y recuperación de las imágenes con blur

Una vez calibrado el sistema, la metodología aplicada a la hora de capturar imágenes con blur fue la misma que la explicada para la obtención de las PSFs de una apertura, es decir, fijar el enfoque a una distancia $D = 1m$ y, a continuación, mover la escena (variando la profundidad) sin modificar ningún otro parámetro obteniendo, finalmente, imágenes desenfocadas. La diferencia en el caso de la obtención de imágenes con blur es que, una vez finalizado el proceso de captura para las distintas distancias, se procedió a capturar las mismas escenas totalmente enfocadas. Esto se decidió para que, una vez se procediese con la recuperación de la información perdida, se pudiera comparar la imagen resultante con la escena real (totalmente enfocada). Se tomaron imágenes con las aperturas obtenidas y con la apertura circular (inherente al objetivo)

para comparar los resultados, mostrando los mismos en la siguiente sección.

Finalmente, se procedió a recuperar la información de las imágenes capturadas de forma análoga a la explicación de la Subsección 5.1.2 para los resultados simulados. Se utiliza la deconvolución de Wiener, con una estimación de SNR de 0.005.

6.4. Resultados

Después de explicar la metodología para los experimentos reales se muestran, en las Figuras 6.7 y 6.8, dos ejemplos de los resultados obtenidos en la recuperación de imágenes con blur. En el Apéndice D se muestran un conjunto de resultados obtenidos para el resto de aperturas evaluadas. Las aperturas evaluadas son las generadas para valores de σ pequeños (menores de 0.005), por ser las que daban mejores resultados en las simulaciones (véanse las Figuras 5.5 y 5.7 de la Sección 5.2). Los valores de ruido mayores son inusuales en fotografías convencionales, aunque no se ha comprobado para capturas de escenas con poca iluminación (i.e. escenas nocturnas) en las que el ruido podría alcanzar valores mayores.

En las Figuras 6.7 y 6.8 se puede ver la recuperación de información totalmente perdida como las letras de los libros en las escenas fotografiadas. En las figuras se pretende enfocar los objetos que están posicionados en el centro de la imagen con blur (a una profundidad de desenfoque de 40 cm y 60 cm, respectivamente). Para ello se utilizó la respuesta más cercana, del array de diodos LED, al centro de la imagen. Al utilizar la respuesta, cuyas coordenadas se aproximan al centro de la imagen con blur, se consigue recuperar la información de los objetos que, espacialmente, están alrededor de esas coordenadas (siempre que su profundidad sea la misma que la de la respuesta utilizada). Sin embargo, en la Figura 6.9 se puede apreciar la existencia de objetos a distintas profundidades, cuya información no se recupera. En dichas zonas de la imagen aparecen artefactos puesto que la PSF con la que se deconvoluciona no es la adecuada para esas profundidades. No obstante, en el centro de la figura, se puede comprobar la calidad de nuestras aperturas codificadas a la hora de usar su respuesta para recuperar la información perdida en una captura.

En la Figura 6.10 se puede ver una comparativa de los resultados obtenidos con aperturas óptimas para tres valores distintos de σ con los resultados obtenidos al utilizar la PSF de una apertura circular en igualdad de condiciones (misma profundidad de desenfoque). Como se puede apreciar, la recuperación de información con una apertura circular es mínima, en comparación con la recuperación obtenida con aperturas codificadas, apareciendo gran cantidad de artefactos (esta vez debido a la mala respuesta de una apertura circular).

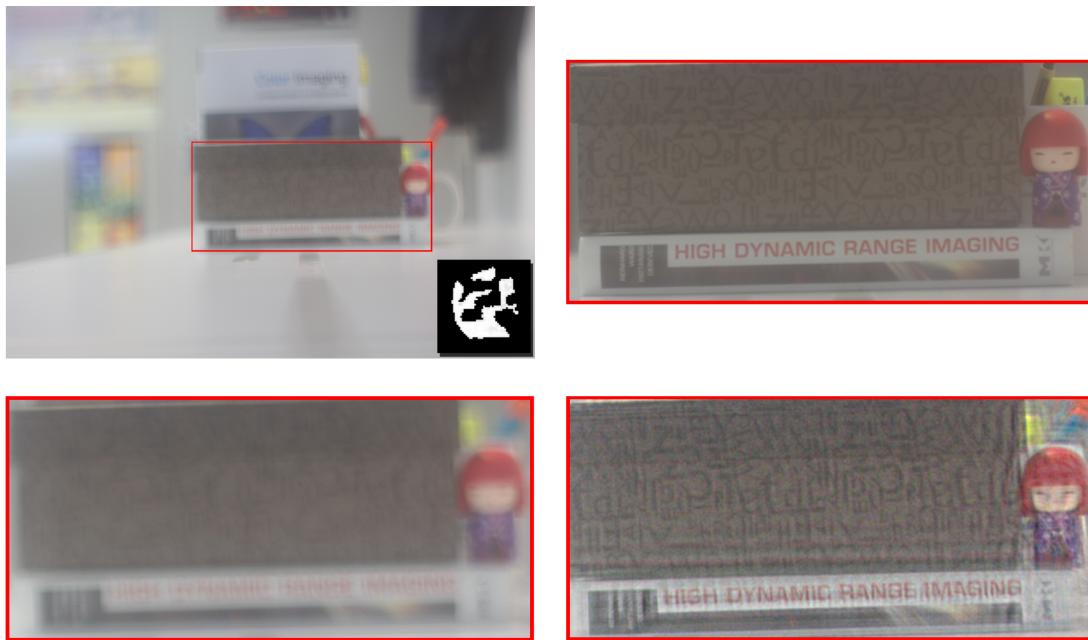


Figura 6.7: Resultado de la apertura óptima para $\sigma = 0.005$ para profundidad de desenfoque de 40 centímetros. Arriba a la izquierda: foto capturada con aparición de blur y PSF de la apertura utilizada, a la derecha: zoom de detalle de la escena totalmente enfocada. Abajo a la izquierda: zoom de detalle de la foto capturada, a la derecha: zoom de detalle de la foto obtenida tras recuperar la información perdida en la captura.

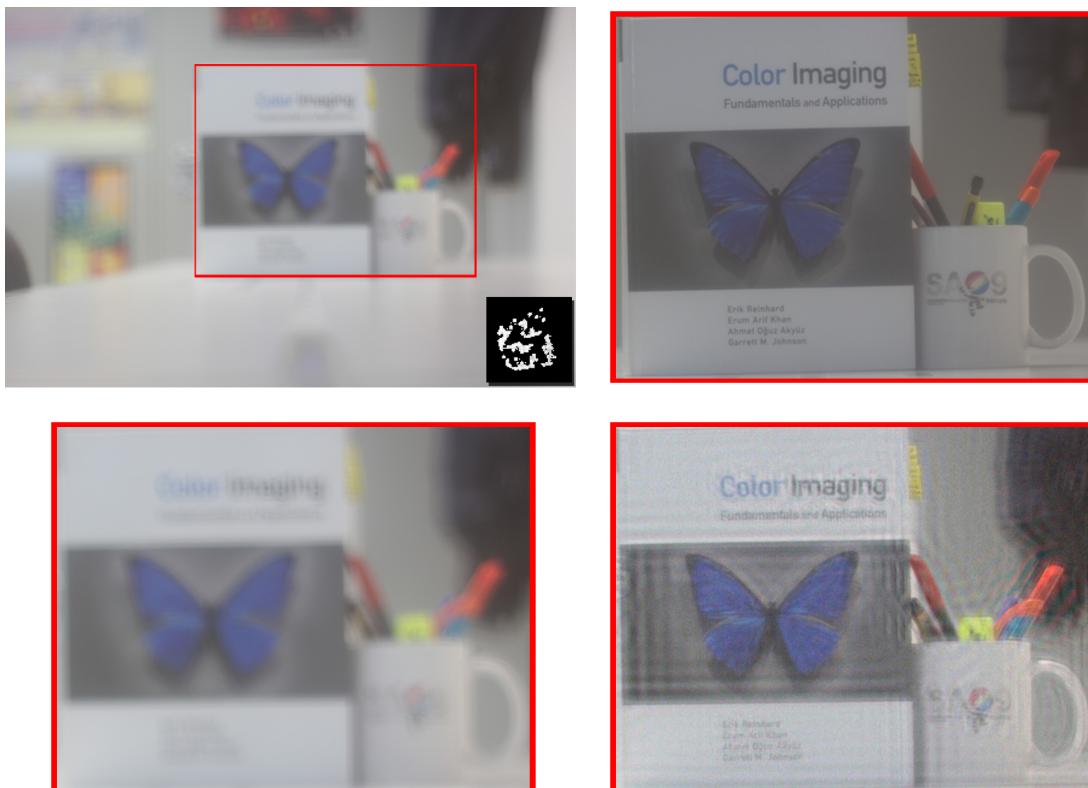


Figura 6.8: Resultado de la apertura óptima para $\sigma = 0.0001$ para profundidad de desenfoque de 60 centímetros. Arriba a la izquierda: foto capturada con aparición de blur y PSF de la apertura utilizada, a la derecha: zoom de detalle de la escena totalmente enfocada. Abajo a la izquierda: zoom de detalle de la foto capturada, a la derecha: zoom de detalle de la foto obtenida tras recuperar la información perdida en la captura.

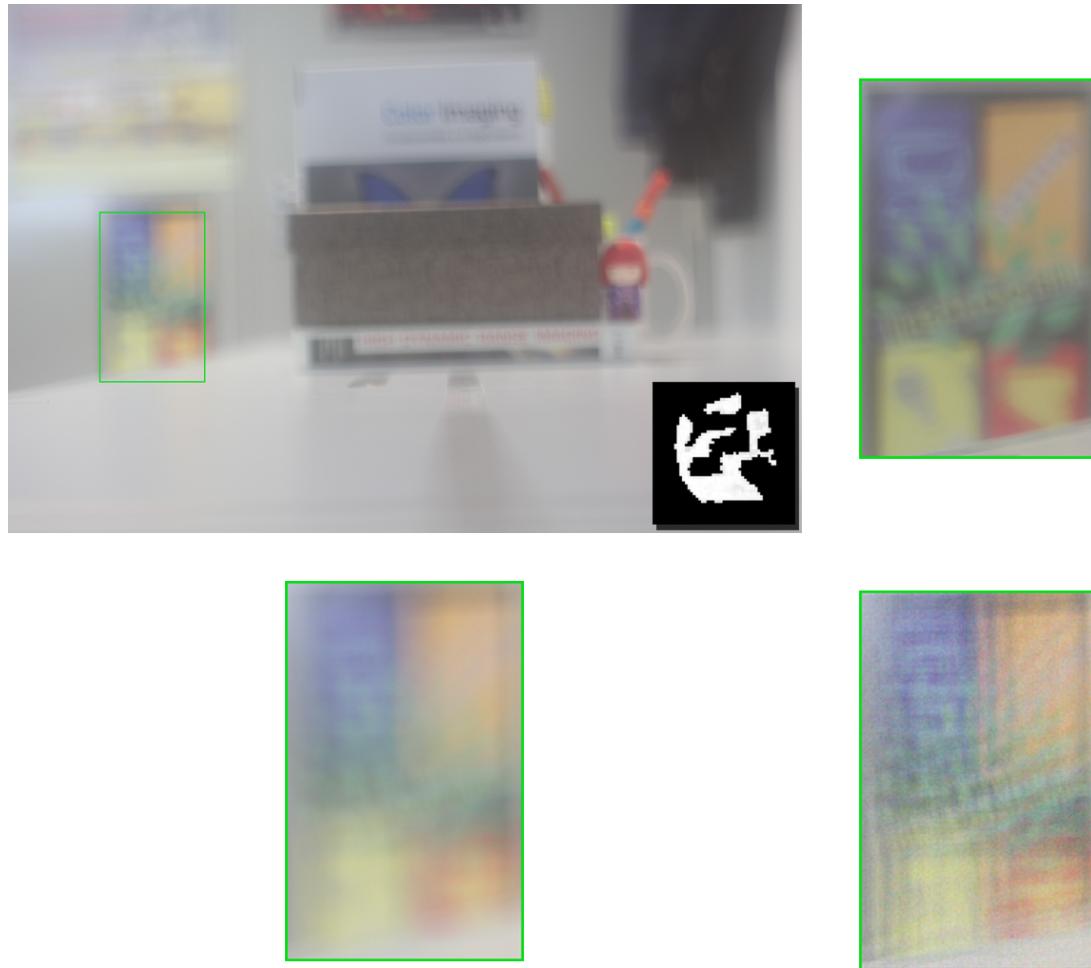


Figura 6.9: Importancia de la profundidad de desenfoque para la que se obtiene la PSF utilizada en el proceso. La información de los objetos que están a distinta profundidad de desenfoque (como el marcado en la figura) no podrá ser recuperada. Arriba a la izquierda: foto capturada con aparición de blur y PSF de la apertura utilizada, a la derecha: zoom de detalle de la escena totalmente enfocada. Abajo a la izquierda: zoom de detalle de la foto capturada, a la derecha: zoom de detalle de la foto obtenida tras recuperar la información perdida en la captura.

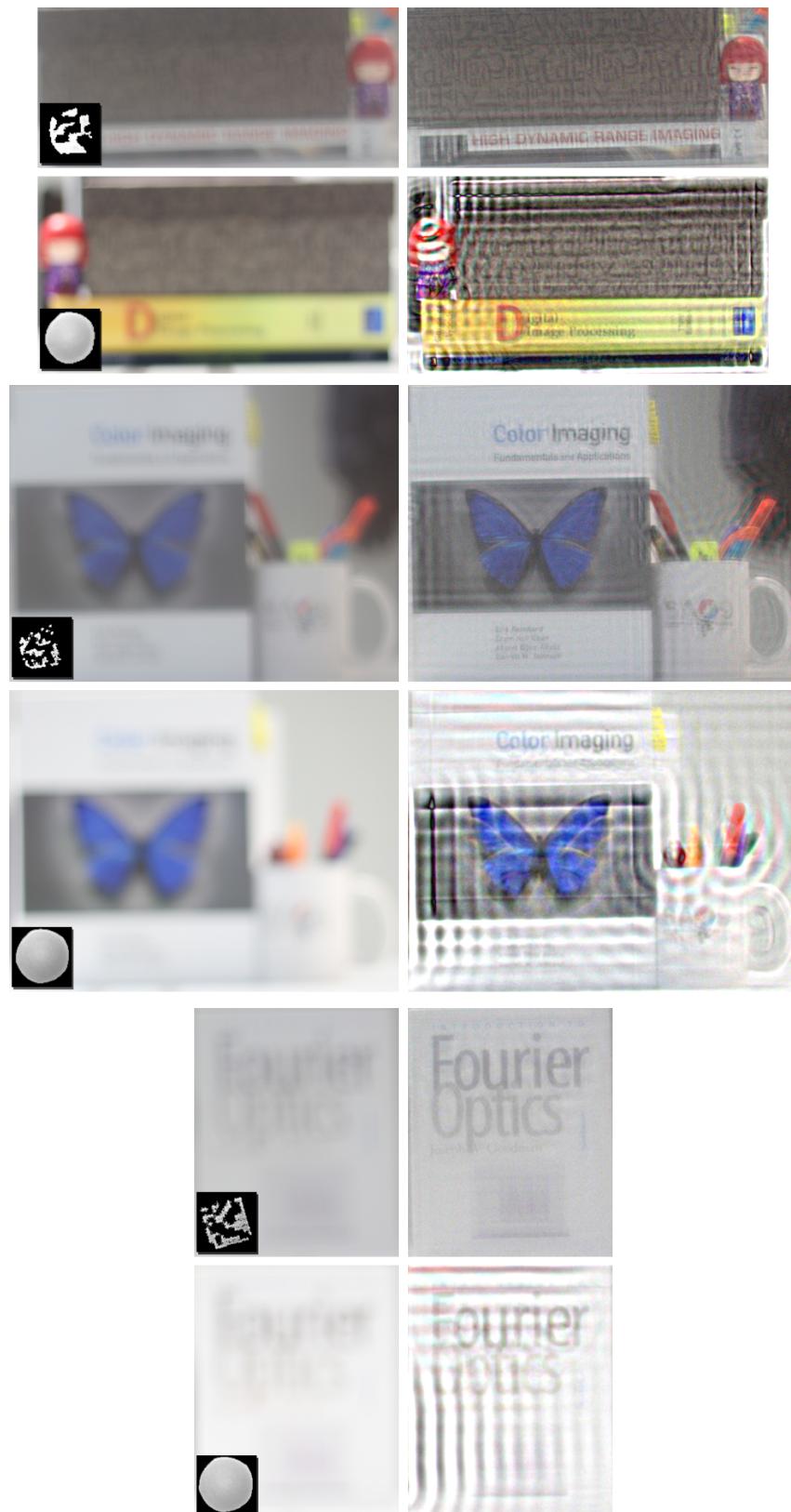


Figura 6.10: Comparativa entre resultados obtenidos con aperturas codificadas y una apertura circular en igualdad de condiciones. Arriba: resultados de la apertura óptima para $\sigma = 0.005$ y una apertura circular para profundidad de desenfoque de 40 centímetros. Centro: resultados de la apertura óptima para $\sigma = 0.0001$ y una apertura circular para profundidad de desenfoque de 60 centímetros. Abajo: resultados de la apertura óptima para $\sigma = 0.002$ y una apertura circular para profundidad de desenfoque de 90 centímetros.

6. Validación II: Experimentos reales

7. Conclusiones y trabajo futuro

Una vez finalizado el proyecto, se van a poner de manifiesto las aportaciones y limitaciones encontradas en el trabajo realizado, a modo de resumen final sobre el mismo. Además se comentarán los posibles puntos que se podrían extender en un trabajo futuro y las conclusiones personales obtenidas.

7.1. Conclusiones

La primera parte del proyecto generó un amplio trabajo de investigación en un campo emergente, como lo es la fotografía computacional, recopilando un estudio del arte. Dicho estudio ofrece una visión sucinta, pero representativa y estructurada, de un campo de muy reciente aparición y con muy poca literatura establecida.

El proyecto ofrece una solución de blur por desenfoque en imágenes. Para ello, el trabajo se basó en una reciente aproximación para dicho problema, utilizando aperturas codificadas. De esta manera, se implementó un algoritmo genético que obtendría un conjunto de aperturas codificadas. Hay que tener en cuenta que los múltiples parámetros de este tipo de algoritmos deben ser estudiados, con el objetivo de encontrar una solución válida (en este caso una apertura codificada con mejores resultados que una apertura circular), para lo cual se realizó un estudio de la convergencia del algoritmo.

Se analizaron las aperturas codificadas obtenidas por el algoritmo genético, comparando sus resultados con los resultados obtenidos por una apertura circular, en un entorno controlado de simulaciones. Una vez validadas estas aperturas, se procedió a realizar pruebas con experimentos reales concluyendo los resultados obtenidos en la Sección 6.4 y en el Apéndice D.

El trabajo desarrollado permite recuperar información perdida en una imagen, siendo las únicas entradas una fotografía con blur y la PSF con la que se desea trabajar. Con el método desarrollado se ha conseguido recuperar gran parte de la información perdida en la captura de una escena, demostrándose un rendimiento de las aperturas obtenidas claramente superior al de una apertura circular. Sin embargo, existen algunas limitaciones que no permiten recuperar la “totalidad” de la información perdida. La limitación principal es la utilización de la misma PSF en toda la imagen capturada, en la que puede haber objetos a distintas profundidades. El método desarrollado permite recuperar la información perdida en la zona de la imagen en la que los objetos están a una profundidad conocida, apareciendo artefactos en el resto de la imagen.

7.2. Desarrollo del proyecto

El tiempo dedicado al desarrollo de este trabajo ha sido de 10 meses en dos intervalos temporales. La primera parte del proyecto, dedicada completamente al estudio del estado del arte de la fotografía computacional e investigación de trabajos realizados por otros autores, se realizó durante abril y junio de 2010 durante tres horas diarias (a la vez, realizando asignaturas pendientes de la titulación).

El proyecto se retomó en septiembre de 2010 con una jornada laboral de seis horas diarias hasta el mes de noviembre; a partir de entonces y hasta febrero de 2011 la jornada laboral ha sido de ocho horas diarias. Siendo un proyecto de investigación, han ido surgiendo nuevos problemas a resolver dados los resultados obtenidos en cada momento, teniendo que incluir objetivos intermedios hasta llegar a resolver el objetivo principal del mismo.

Además, en el 2011 se decidió enviar una publicación del trabajo realizado (extendida por un pequeño estudio de aperturas no binarias) al Congreso Ibero-American de Informática Gráfica (SIACG 2011), cuya fecha límite de aceptación de publicaciones fue el 7 de febrero de 2011. Esto supuso dedicar una parte del desarrollo del proyecto a la creación de la publicación, así como de una batería de resultados finales; por lo que el trabajo debía estar funcionando con tiempo suficiente para satisfacer la entrega a tiempo de la publicación.

7.3. Trabajo futuro

Una primera mejora del trabajo realizado es conseguir una *ejecución menos costosa* del algoritmo genético, paralelizando la ejecución del mismo para conseguir acelerar la sucesión de generaciones. Por otro lado, ya que la transmitancia de luz de una apertura codificada es menor que la de una apertura circular, *optimizar la función de evaluación* consiguiendo tener en cuenta esa característica sería otra posible extensión del trabajo realizado. Otra principal mejora, sería *trabajar con varias PSFs en la imagen*, aplicando una PSF distinta a cada zona de la imagen, permitiendo recuperar casi la totalidad de la información perdida para cualquier fotografía con blur por desenfoque. Además, encontrar otro *método de obtención de las PSFs* de una apertura a distintas profundidades de desenfoque podría mejorar los resultados obtenidos. Finalmente, realizar un *estudio más extenso acerca de la convergencia* del algoritmo a aperturas óptimas en relación a los valores de los parámetros podría determinar unos mejores valores de los mismos a la hora de buscar aperturas codificadas que funcionen de mejor manera que las obtenidas.

7.4. Conclusiones personales

De manera subjetiva, el desarrollo de mi proyecto fin de carrera en el seno del GIGA, junto a Belén y Diego, ha generado una satisfacción enorme en varios aspectos.

Para empezar, este proyecto ha sido un proyecto de investigación sobre un campo muy emergente y su desarrollo no se limita a “construir una muro con los ladrillos que mis directores de proyecto me faciliten”; de esta manera he aprendido a trabajar bajo otro enfoque que el apren-

7. Conclusiones y trabajo futuro

dido en la carrera muchas veces limitado a pensar en “como hacer que mi programa funcione”.

Para continuar, el haber realizado experimentos reales sobre todo lo que, en otras ocasiones queda en el estudio teórico de un problema (o, a lo sumo, en simulaciones), es algo que tampoco me esperaba llegar a conseguir, siendo una parte muy gratificante.

Por último, ser parte de una publicación en un congreso me parece algo que no se consigue, en la mayoría de las ocasiones, antes de acabar la carrera universitaria; de esta manera, que los acontecimientos me hayan permitido formar parte de una publicación es muy satisfactorio.

7. Conclusiones y trabajo futuro

Bibliografía

- [1] M. Aggarwal and N. Ahuja. Split aperture imaging for high dynamic range. *International Journal of Computer Vision*, 58:7–17, June 2004.
- [2] M. Ben-Ezra and S. K. Nayar. Motion deblurring using hybrid imaging. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 657–664, Jun 2003.
- [3] V. Brajovic and T. Kanade. A sorting image sensor: an example of massively parallel intensity-to-time processing for low-latency computational sensors. In *IEEE International Conference on Robotics and Automation (Proc.)*, pages 1638–1643, 1996.
- [4] P.J. Burt and R.J. Kolczynski. Enhanced image capture through fusion. *Fourth International Conference on Computer Vision*, pages 173–182, 11-14 May 1993.
- [5] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs.
- [6] J. M. Dicarlo, F. Xiao, and A. W Brian. Illuminating illumination. In *Ninth Color Imaging Conference*, pages 27–34, 2001.
- [7] E. Eisemann and F. Durand. Flash photography enhancement via intrinsic relighting. *ACM Transactions on Graphics*, 23:673–678, August 2004.
- [8] S. Gottesman and E. Fenimore. New family of binary arrays for coded aperture imaging. *Applied Optics*, 1989.
- [9] E. H. Adelson and John Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:99–106, 1992.
- [10] R. J. Handy. High dynamic range ccd detector image. U.S. Patent 4623928, November 1986.
- [11] S. Hiura and T. Matsuyama. Depth measurement by the multi-focus camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1998.
- [12] John H. Holland. *Adaptation in natural and artificial systems*. MIT Press, Cambridge, MA, USA, 1992.
- [13] T. Jack, A. Agrawal, and R. Raskar. Why I want a gradient camera. In *IEEE Computer Vision and Pattern Recognition (Proc.)*, CVPR '05, pages 103–110, Washington, DC, USA, 2005. IEEE Computer Society.
- [14] D. Krishnan and R. Fergus. Dark flash photography. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, page To appear, 2009.

- [15] A. Levin, R. Fergus, F. Durand, and W. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, 26(3), 2007.
- [16] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen. Programmable aperture photography: multiplexed light field acquisition. *ACM Transactions on Graphics*, 27(3), 2008.
- [17] B. C. Madden. Extended intensity range imaging. Technical report, 1993.
- [18] S. Mann and R. W. Picard. Being undigital with digital cameras: extending dynamic range by combining differently exposed pictures. Technical Report 323, M.I.T. Media Lab Perceptual Computing Section, Boston, Massachusetts, 1994.
- [19] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 374–380, Jun 1999.
- [20] H. Nagahara, S. Kuthirummal, C. Zhou, and S. K. Nayar. Flexible depth of field photography. In *European Conference on Computer Vision*, Oct 2008.
- [21] S. K. Nayar, V. Branzoi, and T. Boult. Programmable imaging using a digital micromirror array. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 436–443, Jun 2004.
- [22] S. K. Nayar and T. Mitsunaga. High dynamic range imaging: spatially varying pixel exposures. In *IEEE Computer Vision and Pattern Recognition (Proc.)*, pages 472–479, 2000.
- [23] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics*, 23:664–672, August 2004.
- [24] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. *ACM Transactions on Graphics*, 25:795–804, July 2006.
- [25] R. Raskar and J. Tumblin. Computational photography. <http://web.media.mit.edu/~raskar/>.
- [26] R. Raskar, J. Yu, and A. Ilie. A non-photorealistic camera: detecting silhouettes with multi-flash. In *SIGGRAPH 2003 Technical Sketch: Conference Abstracts and Applications*, 2003.
- [27] S. Reeves. Image deblurring - wiener filter. <http://blogs.mathworks.com/steve/2007/11/02/imagedeblurring-wiener-filter/>, 2007.
- [28] Ng. Ren. Fourier slice photography. *ACM Transactions on Graphics*, 24:735–744, July 2005.
- [29] R. A. Street. High dynamic range segmented pixel sensor array. U.S. Patent 5789737, August 1998.

- [30] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics*, 26, July 2007.
- [31] A. C. Sheikh H. R. Wang, Z. Bovik and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004.
- [32] D. D. Wen. High dynamic range charge coupled device. U.S. Patent 4873561, October 1989.
- [33] Y. Yitzhaky, I. Mor, A. Lantzman, and N.S. Kopeika. A direct method for restoration of motion blurred images. *Journal of the Optical Society of America A*, page 1998.
- [34] J. Zand. Coded aperture imaging in high energy astronomy. <http://lheawww.gsfc.nasa.gov/docs/cai/coded.html>, 1996.
- [35] C. Zhou, S. Lin, and S. K. Nayar. Coded aperture pairs for depth from defocus. In *IEEE International Conference on Computer Vision*, Oct 2009.
- [36] C. Zhou and S. K. Nayar. What are good apertures for defocus deblurring? In *IEEE International Conference on Computational Photography*, Apr 2009.
- [37] S. Zhuo and T. Sim. On the recovery of depth from a single defocused image. In *Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns*, 2009.
- [38] A. Zomet and S. K. Nayar. Lensless imaging with a controllable aperture. In *IEEE Conference on Computer Vision and Pattern Recognition*, Jun 2006.

Apéndice A. Derivación matemática de la función de evaluación

Como se explica en el Capítulo 3, se busca diseñar una apertura codificada cuya respuesta en frecuencias sea mejor que la respuesta de una apertura circular, con el objetivo de recuperar la información perdida en la imagen capturada. Se sigue la aproximación de Zhou y Nayar [36] en la que la calidad de una apertura viene dada por el error entre la imagen enfocada y el resultado de la deconvolución de la imagen desenfocada con la apertura. Para comenzar, conviene definir la formulación, en el dominio de la frecuencia, establecida para la deconvolución de Wiener:

$$\hat{F}_0 = \frac{F \cdot \bar{K}}{|K|^2 + |C|^2}, \quad (\text{A.1})$$

donde:

- \hat{F}_0 es la imagen re-enfocada, recuperada.
- F es dicha imagen desenfocada.
- K es la respuesta de la apertura utilizada.
- C es la relación de potencia señal-ruido (SNR).

De esta manera, se establece la medición de la calidad de una apertura como el resultado obtenido al calcular la esperanza de la norma (o distancia) L_2 entre la imagen re-enfocada \hat{F}_0 y la misma imagen enfocada F_0 , en espacio frecuencial:

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

donde ζ es el ruido de la imagen.

Sustituyendo la ecuación A.1 en la ecuación A.2, se obtiene:

$$R(K, F_0, C) = \mathbb{E}_{\zeta}[\left\| \frac{\zeta \cdot \bar{K} - F_0 \cdot |C|^2}{|K|^2 + |C|^2} \right\|^2] \quad (\text{A.3})$$

A. Derivación matemática de la función de evaluación

Asumiendo que el ruido sigue una distribución gaussiana de media 0 y una desviación típica dada por σ , $\zeta \sim N(0, \sigma^2)$, la ecuación A.3 queda:

$$R(K, F_0, C) = \left\| \frac{\sigma \cdot \bar{K}}{|K|^2 + |C|^2} \right\|^2 + \left\| \frac{F_0 \cdot |C|^2}{|K|^2 + |C|^2} \right\|^2 \quad (\text{A.4})$$

Para poder eliminar de la ecuación A.4 la incógnita F_0 , que es desconocida, se utiliza un modelo natural de imágenes. La esperanza de $|F_0|^2$ es calculada como:

$$A(\xi) = \int_{F_0} |F_0(\xi)|^2 d\mu(F_0), \quad (\text{A.5})$$

donde ξ representa frecuencia y A se calcula promediando el espectro de potencia de un conjunto de imágenes naturales. Sustituyendo dicha esperanza A.5 en la ecuación A.4, se obtiene:

$$R(K, C) = \left\| \frac{\sigma \cdot \bar{K}}{|K|^2 + |C|^2} \right\|^2 + \left\| \frac{A^{1/2} \cdot |C|^2}{|K|^2 + |C|^2} \right\|^2 \quad (\text{A.6})$$

El valor de $|C|^2$ que, para un K dado, minimiza el valor de R es $|C|^2 = \sigma^2/A$. Sustituyendo dicho valor en la ecuación A.6 se elimina otra incógnita. Si se reorganiza la ecuación A.6, se obtiene la función de evaluación empleada:

$$R(K) = \frac{\sigma^2}{|K|^2 + \sigma^2/A} \quad (\text{A.7})$$

Apéndice B. Desmontaje del objetivo

En este apéndice se va a mostrar el proceso de desmontaje del objetivo a la hora de introducir una apertura codificada en el interior del sistema de lentes.



Figura B.1: *Objetivo utilizado (EF 50mm f 1.8 II) sin abrir.*



Figura B.2: *Paso 01: voltear el objetivo y extraer la tapa posterior del mismo. La tapa se extrae cuando se coloca el objetivo en la cámara fotográfica.*

B. Desmontaje del objetivo



Figura B.3: *Paso 02: extracción de dos micro-tornillos para, a posteriori, extraer una tapa interior (dejando visible la microelectrónica del objetivo).*



Figura B.4: *Paso 03: extracción de la tapa que permite cambiar el modo de enfoque del objetivo entre modo manual y modo automático. Esto permitirá extraer la carcasa exterior del objetivo.*



Figura B.5: *Paso 04: extracción de la carcasa exterior del objetivo.*



Figura B.6: *Paso 05: una vez se tiene abierto el objetivo se separa el sistema de lentes del aro de enfoque (pieza circular que permite enfocar la escena en modo de enfoque manual).*



Figura B.7: *Paso 06: sistema de lentes, en cuyo interior se aloja el diafragma del objetivo.*



Figura B.8: *Paso 07: colocación de una apertura en el sistema de lentes.*

B. Desmontaje del objetivo

Apéndice C. Estudio de la convergencia del algoritmo genético

En este apéndice aparecen los resultados del estudio explicado en la Sección 4.3. Se incluye la evolución de $R(K)$ y el tiempo de ejecución para diferentes parámetros.

C.1. Supervivientes

El parámetro *supervivientes* indica el número de cromosomas que son seleccionados al final de cada generación para crear la siguiente generación, es decir los cromosomas que actúan como padres de la siguiente generación. En los resultados obtenidos se observa que cuanto menor es el valor para dicho parámetro (lo que implica un mayor número de cromosomas nuevos en cada generación) el valor de $R(K)$ es menor, siendo el tiempo de ejecución muy similar para cualquier valor del parámetro.

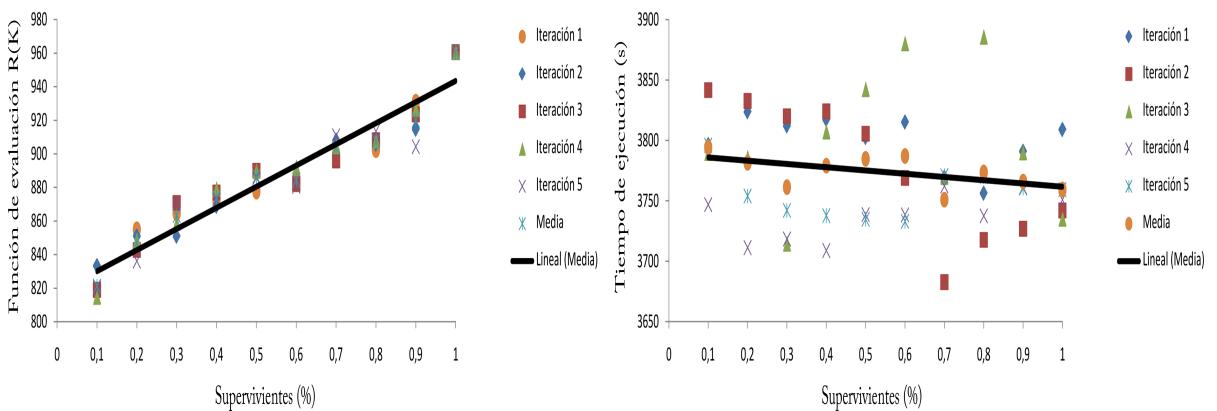


Figura C.1: *Supervivientes (porcentaje con respecto al número de individuos de la población)*. Izquierda, eje horizontal: distintos valores para el parámetro supervivientes, eje vertical: tendencia de la función de evaluación $R(K)$ con la variación del parámetro supervivientes. Derecha, eje horizontal: distintos valores para el parámetro supervivientes, eje vertical: tendencia del tiempo de ejecución del algoritmo con la variación del parámetro supervivientes.

C.2. Generaciones

El parámetro *generaciones* indica el número de generaciones del algoritmo, es decir determina el período de evolución del fenotipo. En los resultados obtenidos se observa que cuanto mayor es el valor para dicho parámetro (lo que implica un mayor número generaciones) el valor de $R(K)$ es menor, siendo el tiempo de ejecución proporcional al incremento del valor del parámetro.

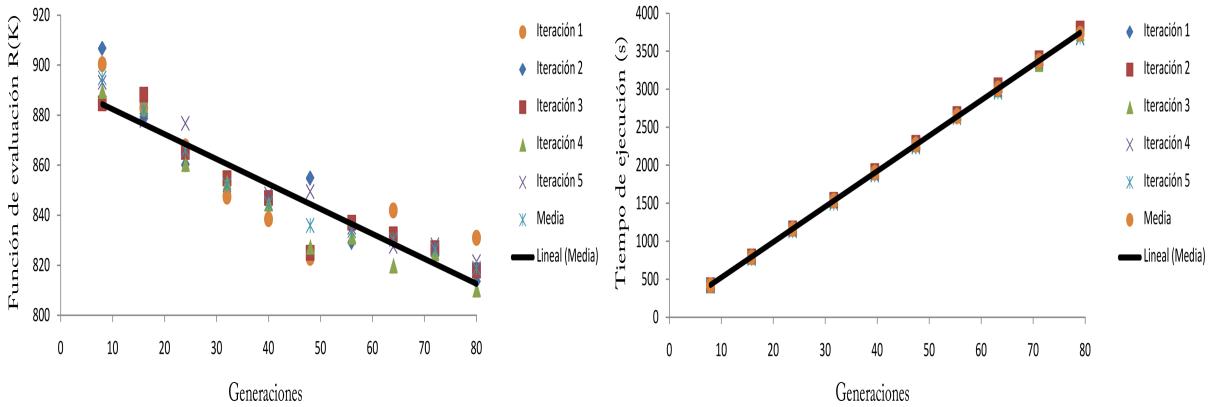


Figura C.2: Izquierdo, eje horizontal: *distintos valores para el parámetro generaciones*, eje vertical: *tendencia de la función de evaluación $R(K)$ con la variación del parámetro generaciones*. Derecha, eje horizontal: *distintos valores para el parámetro generaciones*, eje vertical: *tendencia del tiempo de ejecución del algoritmo con la variación del parámetro generaciones*.

C.3. Población

El parámetro *población* indica el número de cromosomas que conforman cada generación. En los resultados obtenidos se observa que el valor de $R(K)$ es muy similar para cualquier valor del parámetro, siendo el tiempo de ejecución proporcional al incremento del valor del parámetro.

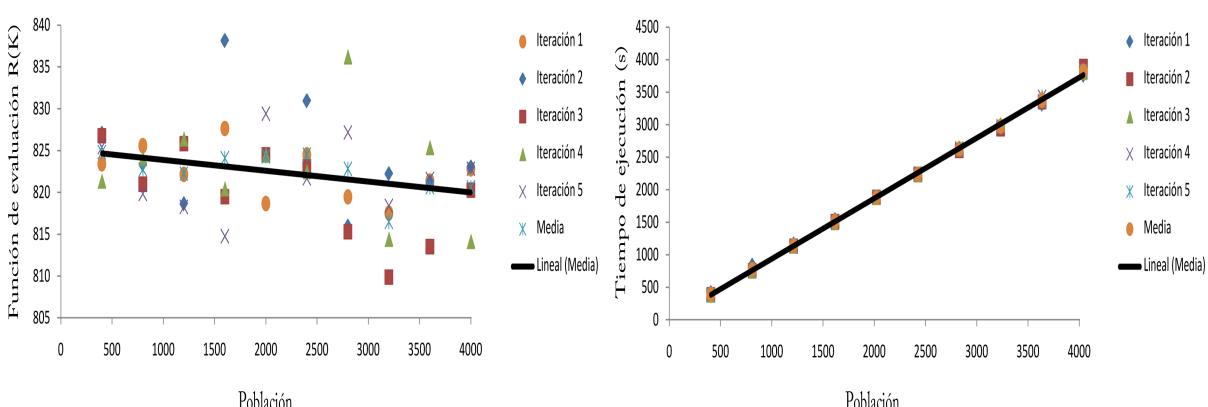


Figura C.3: Izquierdo, eje horizontal: *distintos valores para el parámetro población*, eje vertical: *tendencia de la función de evaluación $R(K)$ con la variación del parámetro población*. Derecha, eje horizontal: *distintos valores para el parámetro población*, eje vertical: *tendencia del tiempo de ejecución del algoritmo con la variación del parámetro población*.

C.4. Cruzamiento

El parámetro *cruzamiento* indica la probabilidad de que un gen sea intercambiado entre dos cromosomas en la operación de cruzamiento. En los resultados obtenidos se observa que el valor de $R(K)$ así como el tiempo de ejecución es muy similar para cualquier valor del parámetro.

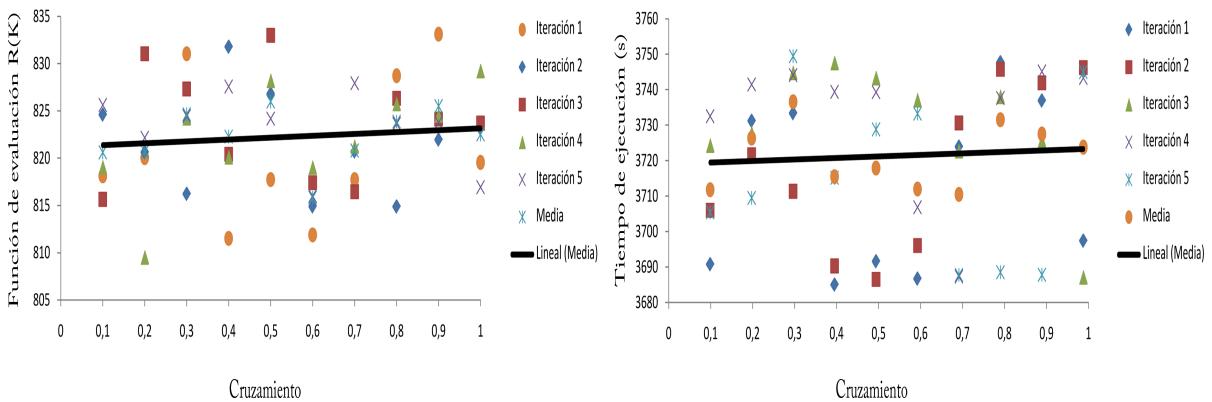


Figura C.4: Izquierda, eje horizontal: *distintos valores para el parámetro cruzamiento*, eje vertical: *tendencia de la función de evaluación $R(K)$ con la variación del parámetro cruzamiento*. Derecha, eje horizontal: *distintos valores para el parámetro cruzamiento*, eje vertical: *tendencia del tiempo de ejecución del algoritmo con la variación del parámetro cruzamiento*.

C.5. Mutación

El parámetro *mutación* indica la probabilidad de que un gen sea mutado (es decir, cambiado de valor) en un cromosoma en la operación de mutación. En los resultados obtenidos se observa que el valor de $R(K)$ así como el tiempo de ejecución es muy similar para cualquier valor del parámetro.

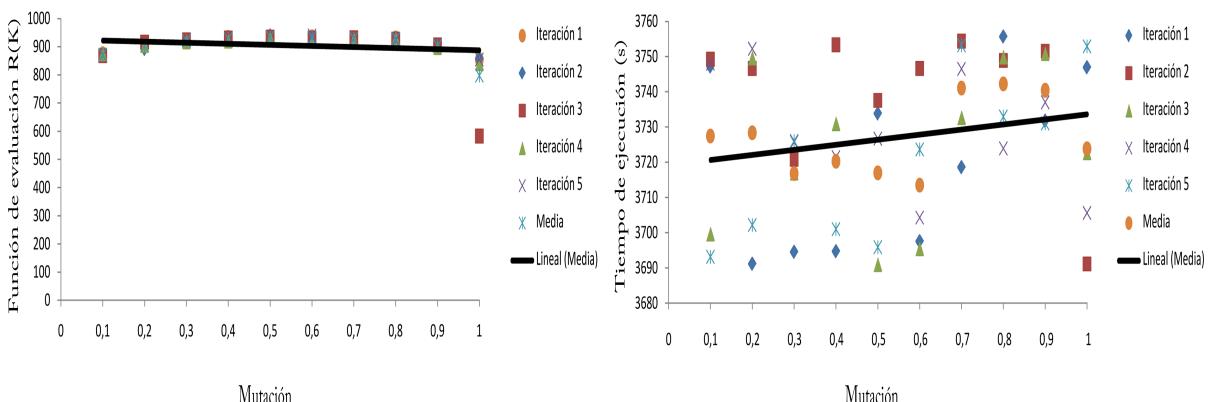


Figura C.5: Izquierda, eje horizontal: *distintos valores para el parámetro mutación*, eje vertical: *tendencia de la función de evaluación $R(K)$ con la variación del parámetro mutación*. Derecha, eje horizontal: *distintos valores para el parámetro mutación*, eje vertical: *tendencia del tiempo de ejecución del algoritmo con la variación del parámetro mutación*.

C. Estudio de la convergencia del algoritmo genético

Apéndice D. Resultados de las pruebas reales



Figura D.1: *Resultados de la apertura óptima para $\sigma = 0.0001$. Se muestran, de arriba a abajo, resultados para tres profundidades de desenfoque distintas (40, 60 y 90 centímetros respectivamente). Para cada profundidad, izquierda: imagen capturada con blur por desenfoque, derecha: imagen obtenida tras recuperar la información perdida en la captura, abajo: PSF de la apertura utilizada.*



Figura D.2: *Resultados de la apertura óptima para $\sigma = 0.001$. Se muestran, de arriba a abajo, resultados para tres profundidades de desenfoque distintas (40, 60 y 90 centímetros respectivamente). Para cada profundidad, izquierda: imagen capturada con blur por desenfoque, derecha: imagen obtenida tras recuperar la información perdida en la captura, abajo: PSF de la apertura utilizada.*



Figura D.3: *Resultados de la apertura óptima para $\sigma = 0.002$. Se muestran, de arriba a abajo, resultados para tres profundidades de desenfoque distintas (40, 60 y 90 centímetros respectivamente). Para cada profundidad, izquierdo: imagen capturada con blur por desenfoque, derecha: imagen obtenida tras recuperar la información perdida en la captura, abajo: PSF de la apertura utilizada.*



Figura D.4: *Resultados de la apertura óptima para $\sigma = 0.0005$. Se muestran, de arriba a abajo, resultados para tres profundidades de desenfoque distintas (40, 60 y 90 centímetros respectivamente). Para cada profundidad, izquierda: imagen capturada con blur por desenfoque, derecha: imagen obtenida tras recuperar la información perdida en la captura, abajo: PSF de la apertura utilizada.*



Figura D.5: *Resultados de la apertura óptima para $\sigma = 0.005$. Se muestran, de arriba a abajo, resultados para tres profundidades de desenfoque distintas (40, 60 y 90 centímetros respectivamente). Para cada profundidad, izquierda: imagen capturada con blur por desenfoque, derecha: imagen obtenida tras recuperar la información perdida en la captura, abajo: PSF de la apertura utilizada.*



Figura D.6: *Resultados de la apertura óptima para $\sigma = 0.005$ y tamaño $7*7$. Se muestran, de arriba a abajo, resultados para tres profundidades de desenfoque distintas (40, 60 y 90 centímetros respectivamente). Para cada profundidad, izquierda: imagen capturada con blur por desenfoque, derecha: imagen obtenida tras recuperar la información perdida en la captura, abajo: PSF de la apertura utilizada.*

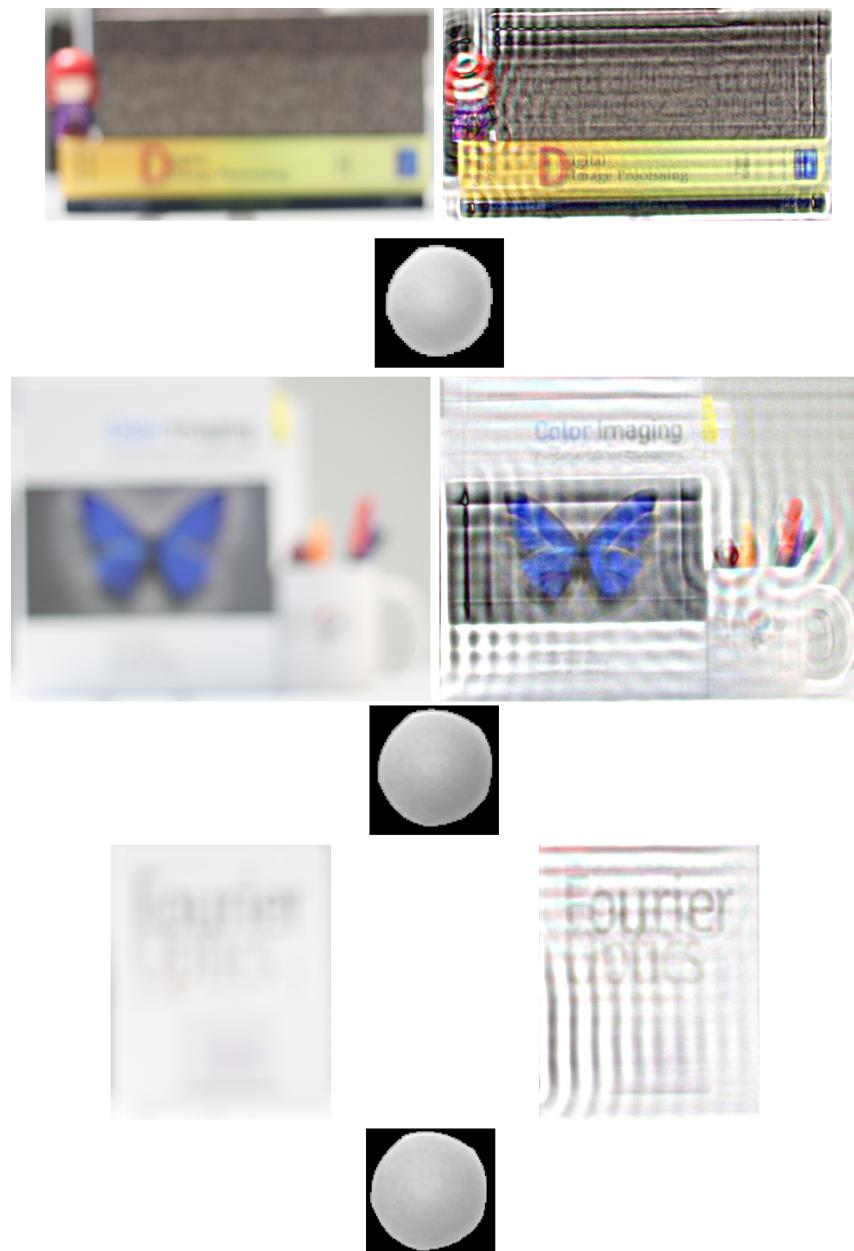


Figura D.7: *Resultados de la apertura circular.* Se muestran, de arriba a abajo, resultados para tres profundidades de desenfoque distintas (40, 60 y 90 centímetros respectivamente). Para cada profundidad, izquierda: imagen capturada con blur por desenfoque, derecha: imagen obtenida tras recuperar la información perdida en la captura, abajo: PSF de la apertura utilizada.

D. Resultados de las pruebas reales

Apéndice E. Artículo: Coded Apertures for Defocus Deblurring

En febrero del 2011 se envió un artículo basado en el trabajo realizado en este proyecto (extendido por un pequeño estudio de aperturas no binarias) al Congreso Ibero-Americanico de Informática Gráfica (SIACG 2011). El resultado no sólo fue la aceptación del artículo enviado si no que, además, fue reconocido como uno de los cinco mejores artículos del congreso. En junio del mismo año, en la celebración del mismo, el artículo se eligió como uno de los tres mejores del congreso, siendo propuesto para aparecer (de manera extendida) en la revista “Computer Graphics Forum, The International Journal of the Eurographics”, revista líder en artículos técnicos sobre gráficos por ordenador con un factor de impacto de 1.681 y posición 22/93 en el ranking JRC.

La versión final del artículo, enviada al congreso, se muestra en este anexo.

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]. Veeraghavan et al. [VRA*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*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 η 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 σ , $N(0, \sigma^2)$. 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.

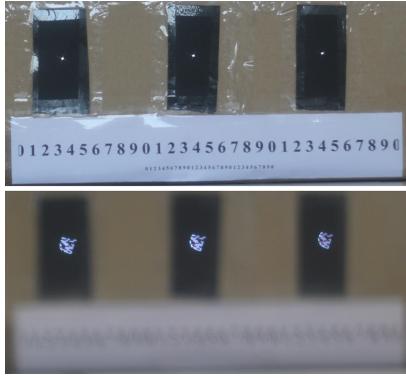


Figure 2: Our poor man’s LED array used to calibrate the PSFs of the apertures. Top: Focused image. Bottom: Image taken with one of our coded apertures at a defocus depth of 70 cm.

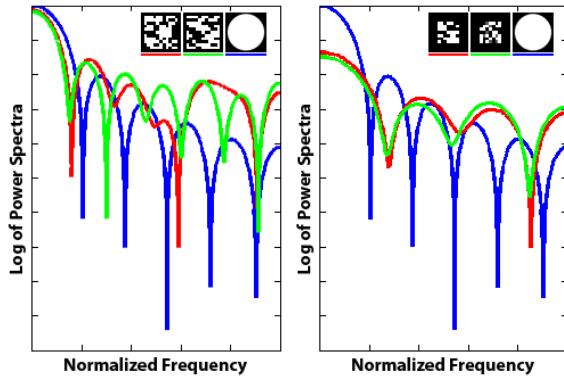


Figure 3: Power spectra comparison of different apertures with respect to a circular aperture (blue). Left: Our apertures for resolution 11×11 and noise levels $\sigma = 0.001$ (red) and $\sigma = 0.005$ (green). Right: Our apertures for resolution 7×7 , binary (red) and non-binary (green).

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 ζ , 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 ζ 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 ξ 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 σ 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, 11×11 and 7×7 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 σ 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 11×11 apertures designed for noise levels of $\sigma = 0.001$ and $\sigma = 0.005$, and the third one is a 7×7 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 7×7 aperture and the

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

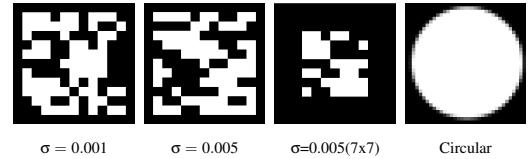


Figure 4: Apertures used in our experiments.

	0.001	0.005	0.005 (7x7)	circular
L2 norm	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 11×11 apertures eliminate less frequencies than the circular aperture, and the 7×7 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).

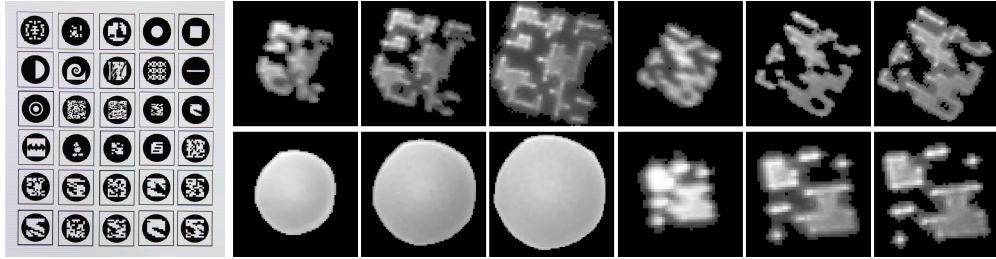


Figure 6: Left: Photomask sheet showing some of the apertures used. Right: PSFs at three different defocus depths (40, 70 and 90 cm) for the four apertures depicted in Figure 4.



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 7×7 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 7×7 . The reason of this is two-fold; first, computational cost of the algorithm is significantly reduced, and second, our previous experiments have shown that 7×7 apertures yield results on par with (or better than) 11×11 apertures. The value of σ (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-

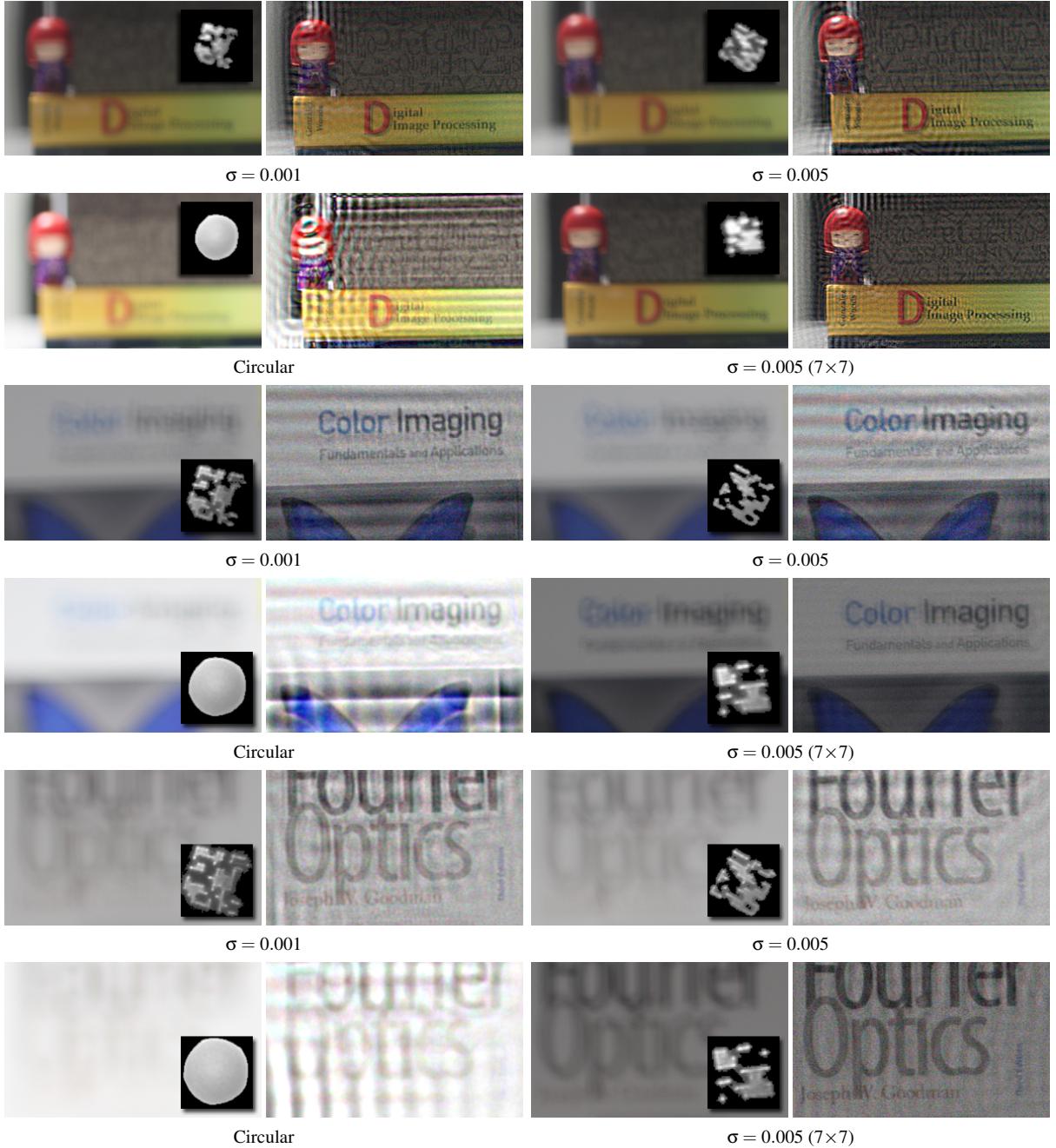


Figure 8: From top to bottom, each of the three scenes have been captured at a defocus depth of 40, 70 and 90 cm, respectively. For each pair of images, the left image shows the captured defocused image and the right image the recovered one. Insets depict the PSF of the aperture used in each case.

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

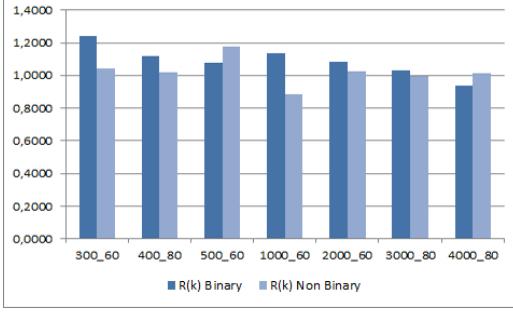


Figure 9: Average value of the quality metric $R(K)$ for binary and non-binary apertures and for different initial parameters of the genetic algorithm.

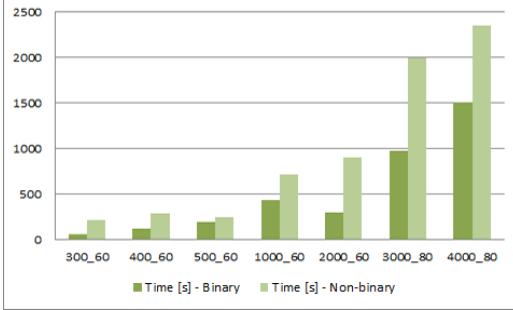


Figure 10: Average value of the time until convergence (in seconds) for binary and non-binary apertures and for different initial parameters of the genetic algorithm.

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

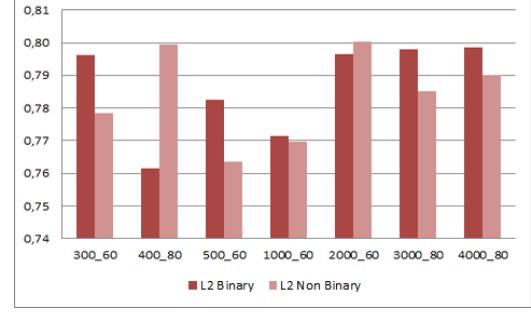


Figure 11: Average value of the L_2 norm for binary and non-binary apertures and for different initial parameters of the genetic algorithm. Values show percentage with respect to the maximum error.

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.

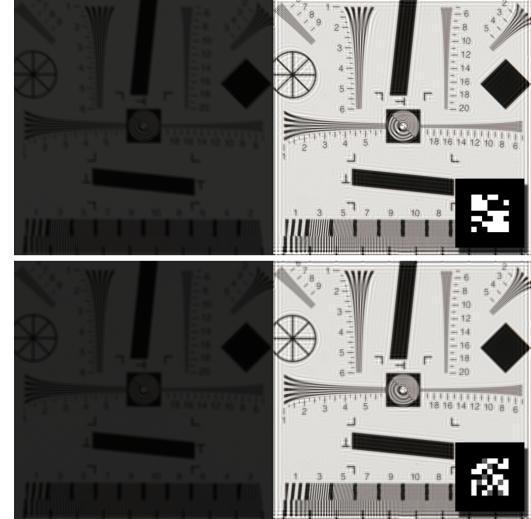


Figure 12: Top left: Defocused image of the ISO 12233 chart obtained using Equation 1 with the aperture shown in the inset in the right as convolution kernel. The aperture is the optimal binary aperture we obtained. Top right: Image recovered using Wiener deconvolution. Bottom row: Same for the optimal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

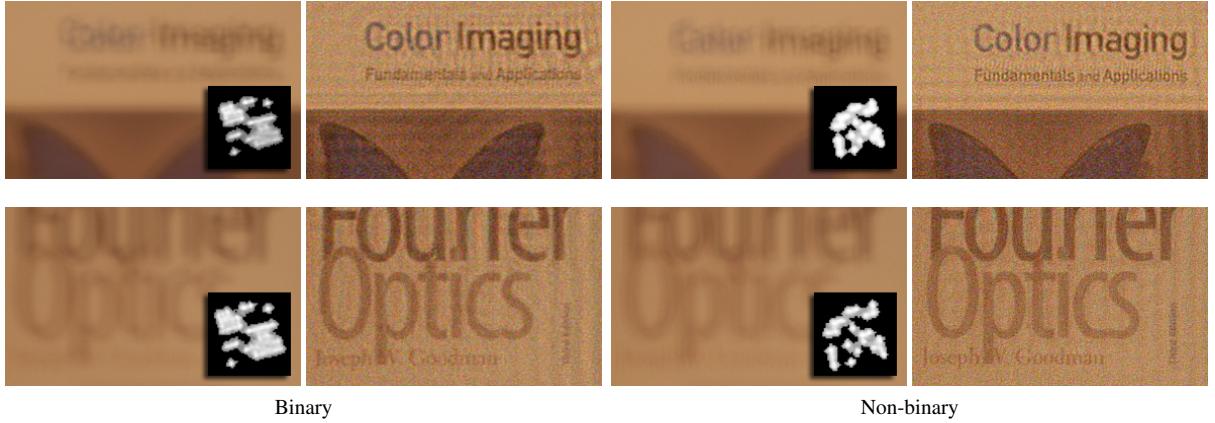


Figure 13: For each pair of images, the left image shows the captured defocused image and the right image the deblurred image. Insets depict the PSF used in each case. Defocus depths are 70 cm (top scene) and 90 cm (bottom scene). The color of the images differs from those in Figure 8; please note that 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 7×7 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](#)