

Proyecto Fin de Carrera

DESCOMPOSICIÓN DE UNA ESCENA EN ILUMINACIÓN GLOBAL Y DIRECTA

Autor

Fernando Martín Domenech

Directora

Elena Garcés García

PONENTE

Dr. D. Diego Gutiérrez Pérez

Escuela de ingeniería y Arquitectura
Curso 2011-2012

Proyecto fin de carrera

DESCOMPOSICIÓN DE UNA ESCENA EN ILUMINACIÓN GLOBAL Y DIRECTA



AUTOR: FERNANDO MARTÍN DOMENECH

DIRECTORA: ELENA GARCÉS GARCÍA
PONENTE: DR. D. DIEGO GUTIÉRREZ PÉREZ

INGENIERIA INFORMÁTICA. CURSO 2011-2012



Universidad de
Zaragoza



Escuela
ingeniería y
arquitectura



Departamento de
Informática e Ingeniería
de Sistemas



Graphics and
Imaging Lab

Agradecimientos

A mis padres Luis y Ana y a toda mi familia y amigos que no dudaron de mí ni cuando yo mismo lo hacía.

También quiero agradecer a todos y cada uno de los miembros del laboratorio Graphics and Imaging Lab el apoyo brindado y la simpatía y amabilidad con la que me acogieron desde el primer día, especialmente a Elena y Diego.

Resumen

Las últimas investigaciones en fotografía computacional han permitido que problemas previamente irresolubles con la fotografía convencional, sean ahora factibles al introducir nuevos dispositivos y técnicas de procesado. Un ejemplo de este tipo de problemas es la obtención de las componentes de iluminación de una escena iluminada por una fuente de luz. Estas componentes son: *iluminación directa*, producida por la incidencia de la luz en un punto, e *iluminación global*, resultante de los rebotes de la luz en otros puntos de la escena como consecuencia de fenómenos físicos tales como reflexión, dispersión o difusión.

Así se tiene que, la componente directa nos ofrece una medida pura acerca de cómo las propiedades materiales de un objeto interactúan con la fuente emisora de luz y la cámara, mientras que la componente global muestra interacciones más complejas entre los distintos objetos que conforman la escena. Por lo tanto, esta descomposición resulta muy útil en numerosas tareas relacionadas con la fotografía computacional tales como la edición o captura de materiales, la re-iluminación de escenas o extracción de la geometría.

Obtener esta descomposición a partir de una única fotografía de la escena es un proceso inviable debido al elevado número de incógnitas y la poca información de la que se dispone. Por tanto, para lograr este objetivo, el trabajo aquí realizado se basa en la publicación de Nayar et. al. [NR06] en la cual se emplea luz estructurada para, mediante la proyección de un patrón de alta frecuencia, obtener información adicional acerca de la escena y así hacer el problema soluble. Este proyecto tiene dos partes claramente diferenciadas. En la primera parte, que denominamos *descomposición en entorno estático*, se reproducen los dos algoritmos expuestos en dicha publicación para obtener la descomposición de la imagen: primero, a partir de varias capturas de la escena y segundo, a partir de una única captura de la misma. Para esta primera parte se requiere una habitación oscura, un proyector LCD fijo que proyecta un patrón de alta frecuencia, una cámara digital para realizar las capturas y la herramienta software Matlab para procesar la información. En la segunda parte del proyecto, denominada *descomposición en entorno móvil*, se implementan los mismos algoritmos sobre el dispositivo móvil Android de Nvidia Tegra 3. Este proceso, más complejo por tratarse de un entorno no controlado, únicamente requerirá la luz del sol como foco y un objeto opaco alargado cuya sombra proyectada servirá como patrón de alta frecuencia.

Índice general

1. Introducción	2
1.1. Motivación	2
1.2. Trabajo relacionado	5
1.3. Objetivos	6
1.4. Contexto	6
1.5. Estructura de la memoria	6
2. Estudio del problema	8
2.1. Definición de iluminación directa y global	8
2.2. Luz estructurada	12
3. Implementación	14
3.1. Algoritmo de descomposición	14
3.2. Implementación en entorno estático	18
3.2.1. Descomposición a partir de varias capturas	19
3.2.2. Descomposición a partir de una captura	22
3.3. Implementación en entorno móvil	27
4. Resultados	30
5. Conclusiones	39

ÍNDICE GENERAL

5.1.	Trabajo realizado	39
5.2.	Resumen temporal del proyecto	40
5.3.	Conclusiones y trabajo futuro	41
5.4.	Conclusiones personales	42
A.	Desarrollo de la aplicación para el tablet NVIDIA Tegra	45
A.1.	Android	46
A.2.	Java Native Interface (JNI)	47
A.3.	Librerías de la FrankenCamera (FCam)	47
A.4.	Estructura de la aplicación móvil	49
A.5.	Arquitectura NVIDIA Tegra 3	50

Capítulo 1

Introducción

En este primer capítulo de la memoria del Proyecto Fin de Carrera titulado *Descomposición de una escena en iluminación global y directa* se abordan los aspectos más esenciales del mismo como son la motivación, trabajo relacionado, el contexto y los objetivos marcados.

1.1. Motivación

El campo de la fotografía computacional reúne conceptos de informática gráfica, óptica y visión por computador; y engloba técnicas de captura, procesado y manipulación de imágenes que permiten mejorar y extender las funcionalidades provistas por las cámaras digitales. En los últimos años, dicho campo ha evolucionado considerablemente con la aparición de nuevos métodos y dispositivos. Esta evolución hace posible que problemas hasta entonces irresolubles puedan ser abordados. Un ejemplo de este tipo de problemas es la descomposición de una escena iluminada por una fuente de luz en sus componentes de iluminación: directa y global.

Una fuente de luz, ya sea natural (el Sol) o artificial, genera una serie de rayos que son propagados por un medio físico (generalmente el aire). Tal y como muestra la figura 1.1, los rayos de luz generados por la fuente viajan a través del medio hasta que alcanzan una superficie. En ese momento, parte de la energía lumínica de dichos rayos es absorbida por la superficie impactada. La energía restante es re-emitida en una nueva dirección y se propaga hasta que alcanza otra superficie, y así sucesivamente, provocando lo que se conoce como rebotes de la luz. Este comportamiento se repite hasta que la luz se

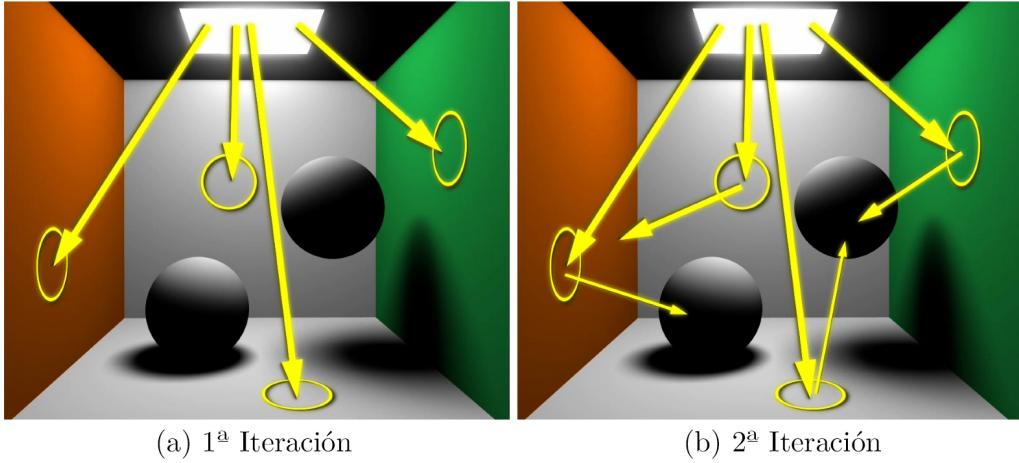


Figura 1.1: La figura de la izquierda (a) muestra el primer viaje de los rayos de luz, y la figura de la derecha (b) muestra los primeros rebotes de luz como consecuencia de los rayos emitidos anteriormente. Nótese que las flechas que representan los rayos de luz son cada vez más cortas denotando una pérdida de energía progresiva. Imágenes obtenidas de www.digitaltutors.com

expande por completo. Por lo tanto, en entornos reales, los objetos de una escena no son únicamente iluminados por la(s) fuente(s) de luz (*iluminación directa*), sino también por aquellos rayos de luz que son reflejados por otros objetos (*iluminación global*). El objetivo principal de este proyecto es obtener la descomposición de una escena en estos dos componentes: una imagen que representa la iluminación directa (sin contemplar los rebotes de la luz) y otra que representa la iluminación global (que contempla la iluminación indirecta de la escena). Un ejemplo de dicha descomposición se muestra en la figura 1.2.

A la hora de abordar el problema de la descomposición que aquí se plantea, el principal obstáculo que aparece es la falta de información sobre la escena. Obtener dicha descomposición a partir de una única fotografía digital de la escena es un proceso inviable dado el elevado número de incógnitas y la poca información de la que se dispone. Esto es así puesto que, a priori, lo único que se conoce acerca de una fotografía digital es el valor discreto de cada uno de los elementos que componen la imagen, comúnmente conocidos como *píxeles*.

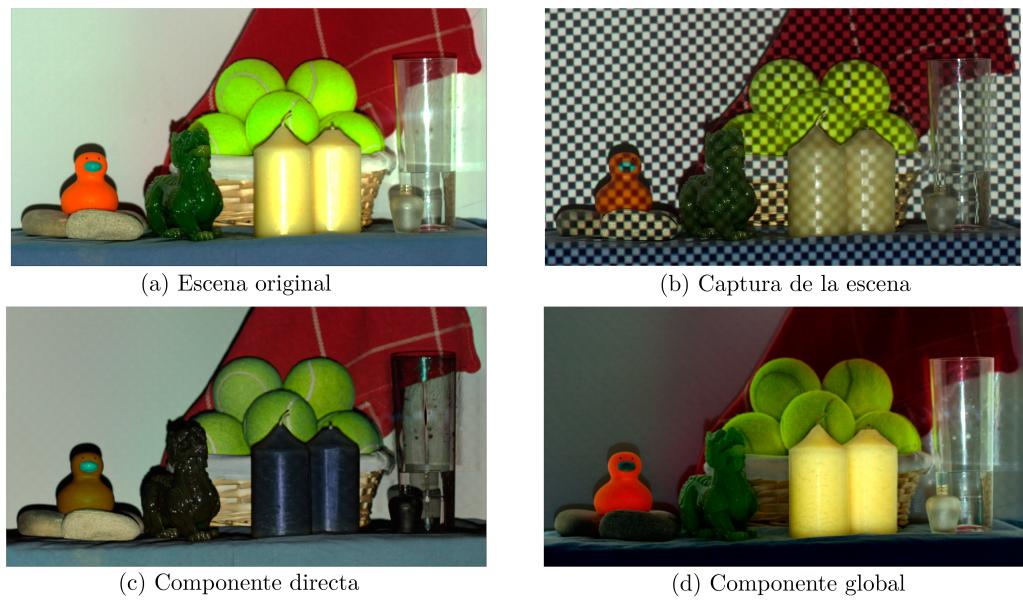


Figura 1.2: *Ejemplo de descomposición de una escena en iluminación directa y global empleando luz estructurada. La componente directa (c) mide como un material interactúa con la fuente de luz y con la cámara. Así puede verse en el dragón de la escena cómo los destellos de luz quedan capturados en dicha componente. Por su parte, la componente global (d) mide interacciones más complejas entre los distintos objetos de la escena. Si se miran detenidamente las pelotas de tenis, se observa como la parte superior de las mismas se encuentra coloreada de rojo como consecuencia de los rayos de luz reflejados por la servilleta de dicho color.*

Para obtener información adicional acerca de la escena y así poder realizar la descomposición, se empleará la técnica de la luz estructurada propuesta por Nayar et. al [NR06]. Esta técnica consiste en proyectar un patrón binario de alta frecuencia sobre la escena. Gracias a la proyección de dicho patrón sobre la escena, parte de los píxeles capturados contendrán información correspondiente a la componente directa y a la componente global simultáneamente y la otra parte de los píxeles, aquellos sobre los cuales hayan sido proyectadas las zonas oscuras (inactivas) del patrón, contendrán información debida únicamente y exclusivamente a la componente global.

1.2. Trabajo relacionado

A pesar del reto que supone la resolución de este problema, no han sido muchos los trabajos que en él se han visto involucrados. Fue a partir de 1980 cuando, con la aparición de múltiples algoritmos como los propuestos por Brooks y Horn [BB89] y Woodham [WOO80] para obtener la geometría de la escena a partir del análisis de las sombras (*Shape-from-shading*) y la iluminación (*Shape-from-brightness*) respectivamente, se observó que dichos algoritmos no tenían en cuenta la iluminación global debida a los múltiples rebotes de los rayos de luz y que, por lo tanto, no producían resultados correctos en escenas complejas con occlusiones, huecos y concavidades. Fue considerada por primera vez la iluminación global para este tipo de algoritmos en 1991 por Nayar et. al. [NK91], sin embargo, tan sólo funcionaba con superficies lambertianas¹ y este enfoque era difícil de generalizar dado que la gran mayoría de objetos presentan superficies complejas con reflejos especulares o dispersión de la luz dentro del material (*subsurface scattering*). Más adelante, en 2005, Seitz et. al. [MK05] propusieron un interesante método para resolver el problema de la descomposición en componentes global y directa. Dicha aproximación consistía en iluminar cada punto de la escena por separado, realizar una captura y determinar en qué medida dicho punto contribuía a la iluminación del resto de la escena. Aunque la base teórica es acertada, el proceso resulta completamente ineficiente. Finalmente, en 2006 Nayar et. al. [NR06] propusieron el algoritmo que da pie a este proyecto y sobre el cual se ha trabajado.

Asimismo, los conceptos de iluminación directa e iluminación global están íntimamente relacionados con el renderizado de imágenes por ordenador. Cuando se renderiza una imagen sintética a partir de un modelo, primero se calcula la iluminación directa y posteriormente se calculan varios niveles de rebotes de la luz (iluminación global) para que la escena final luzca lo más realista posible. Este proyecto plantea el problema inverso: dada una escena real se pretende descomponer en sus dos componentes de iluminación. Sloan et. al. [SS02] y Arikan et. al [AO05], probaron que al calcular por separado estos componentes se consigue aumentar la velocidad de renderizado de imágenes sintéticas. De este modo, la descomposición que aquí se trata puede dar pie a la creación de nuevos algoritmos de renderizado más eficientes en el futuro.

¹Una superficie lambertiana refleja la luz de manera uniforme, es decir, su apariencia no varía independientemente del punto de vista del observador

1.3. Objetivos

Los objetivos planteados para este proyecto y que se cumplieron satisfactoriamente fueron los siguientes:

1. Estudio del problema de descomposición de una escena en sus componentes de iluminación global y directa.
2. Reproducción e implementación de los métodos propuestos en la publicación de Nayar et. al. [NR06] basados en la proyección de luz estructurada con un proyector fijo y la posterior captura de la escena con una cámara fotográfica. Esta primera configuración será denominada como entorno estático a lo largo de esta memoria.
3. Adaptación del método a un entorno móvil no controlado: implementación en el tablet prototipo de NVIDIA Tegra 3 sobre la plataforma Android.

1.4. Contexto

El proyecto ha sido desarrollado dentro del Graphics and Imaging Lab, laboratorio del grupo de informática gráfica avanzada (GIGA) perteneciente al Departamento de Informática e Ingeniería de Sistemas (DIIS) de la Universidad de Zaragoza. Ha sido dirigido por Elena Garcés García y supervisado por el Dr. Diego Gutiérrez Pérez. Asimismo, el tablet prototipo de Nvidia Tegra 3 ha sido una donación de la empresa Nvidia Corp. como resultado de una colaboración concedida en concurrencia competitiva con otras universidades.

1.5. Estructura de la memoria

Los siguientes capítulos de la memoria están estructurados de la siguiente forma: En el capítulo 2 se definen los conceptos claves de iluminación directa e iluminación global. También se analiza en detalle la técnica de la luz estructurada y sus posibles aplicaciones. En el capítulo 3 se detalla el algoritmo de descomposición empleado, se exponen dos métodos distintos para realizar la descomposición en un entorno estático y un tercer método para realizar dicha descomposición en un entorno móvil mediante el uso del tablet NVIDIA Tegra. En el capítulo 4 se analizan los resultados obtenidos por todos

Introducción

los métodos implementados. Finalmente, en el capítulo 5 se valora el trabajo realizado en este proyecto. Además, este documento cuenta con un anexo en el cual se explica el desarrollo de la aplicación para el tablet NVIDIA Tegra.

Capítulo 2

Estudio del problema

Este capítulo define los conceptos claves de iluminación directa e iluminación global. Asimismo, analiza en detalle la técnica de la luz estructurada.

2.1. Definición de iluminación directa y global

Considérese un escenario en el cual se tiene una superficie cualquiera iluminada por una fuente de luz y una cámara digital enfocándola, tal y como muestra la figura 2.1. La fuente de luz genera una serie de rayos que inciden directamente sobre la superficie, sin interactuar con ningún otro medio físico ni material por el camino (iluminación directa). A su vez, tras impactar la superficie, dichos rayos pueden, o bien salir rebotados en otras direcciones, o bien ser absorbidos por la superficie, contribuyendo también a la iluminación de la escena. Este tipo de interacciones más complejas de la luz contribuyen a formar la denominada iluminación global. La cantidad de luz retenida dependerá del tipo de material (p.ej: un material opaco, como puede ser la madera, absorbe mayor cantidad de luz que un cuerpo translucido, como puede ser un folio de papel).

Cuando un rayo de luz impacta contra una superficie, la nueva dirección de los rayos reflejados se define mediante la función BRDF (Bidirectional reflectance distribution function). Esta función modela el comportamiento de los rayos de luz reflejados de modo que, para el caso de la iluminación directa, estos rayos son inmediatamente despedidos nada mas alcanzar la superficie coincidiendo el punto de entrada con el punto de salida. Sin embargo, aquellos rayos de luz que hayan sido absorbidos y que penetran en

la superficie impactada, pueden viajar a través del interior de la misma y emerger al exterior por un punto distinto al punto de entrada. Este comportamiento es conocido como Dispersión dentro del medio (o Subsurface Scattering). Prácticamente todos los materiales del mundo real presentan un cierto grado de dispersión dentro del medio, pero este fenómeno contribuye de forma especialmente significativa al aspecto que presentan ciertos materiales translúcidos como pueden ser el mármol, la cera o la piel humana.

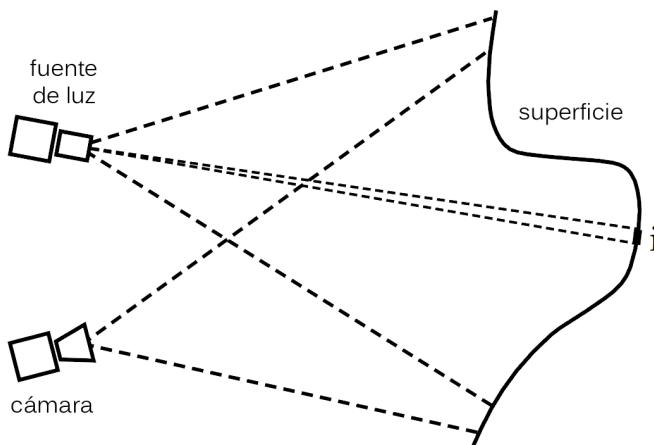


Figura 2.1: *Escenario simple donde una superficie es iluminada por una fuente de luz y capturada mediante una cámara. Imagen obtenida de [NR06].*

Se tiene por lo tanto que la iluminación medida para cada punto de la escena es el resultado de la suma de dos componentes distintas: la *componente directa* y la *componente global*. La componente directa se debe a la incidencia directa de la luz en un punto, mientras que la componente global es la suma resultante de los rebotes de la luz en otros puntos de la escena como consecuencia de varios fenómenos físicos tales como reflexión, dispersión dentro del medio, dispersión volumétrica y transparencia tal y como muestra la figura 2.2¹. La figura 2.4 muestra varias imágenes que presentan dichos fenómenos.

¹Un medio participativo es un material cuyas propiedades afectan al transporte de la luz a través de su volumen. Algunos ejemplos son el agua, la niebla o el humo.

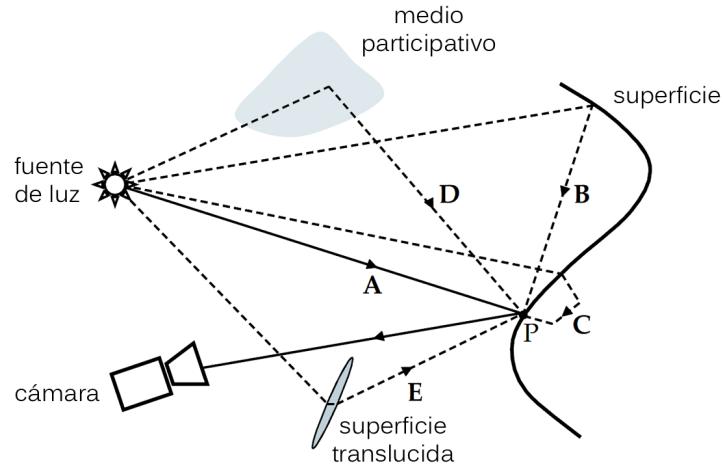


Figura 2.2: La iluminación total de un punto P de la escena se debe a la componente directa de dicho punto (A) y a la componente global que incluye la luz reflejada (B), la luz que viaja por el interior de la superficie (C), la luz que atraviesa diferentes medios físicos (D) y la luz que atraviesa superficies translúcidas (E). Imagen obtenida de [NR06].

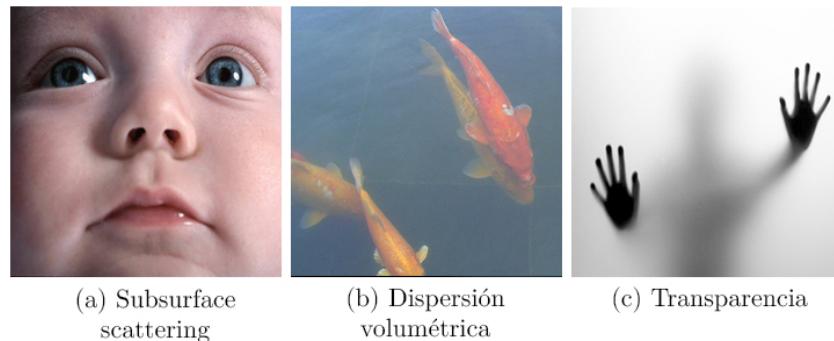


Figura 2.3: La piel humana (a) se compone de varias capas. Estas capas son especialmente delicadas en la zona del rostro, facilitando que los rayos de luz atraviesen la piel y se dispersen en su interior. Al cambiar de medio (del aire al agua) (b), los rayos de luz varían la velocidad y la dirección de propagación. Por último, las superficies translúcidas (c) contribuyen en gran medida a la iluminación global de una escena dado que dejan pasar gran parte de la energía lumínica de los rayos. Imágenes obtenidas de [NR06].

El sensor de una cámara digital no puede desambiguar los datos capturados y separar la información correspondiente a cada componente. Por tanto, resulta interesante realizar dicha descomposición puesto que aporta información adicional acerca de la escena: mientras que la componente directa nos ofrece una medida más pura acerca de cómo las propiedades de un material interactúan con la fuente emisora de luz y la cámara, la componente global muestra interacciones más complejas entre los distintos objetos que conforman la escena. En la figura 2.4 se observa una escena con dos esferas, estando la misma iluminada teniendo en cuenta sólo la componente directa primero y considerando tanto la componente directa como la componente global después. En este segundo caso, la escena presenta un aspecto mucho más realista, resultando especialmente llamativo cómo los colores de las paredes son reflejados en las caras laterales de las esferas.

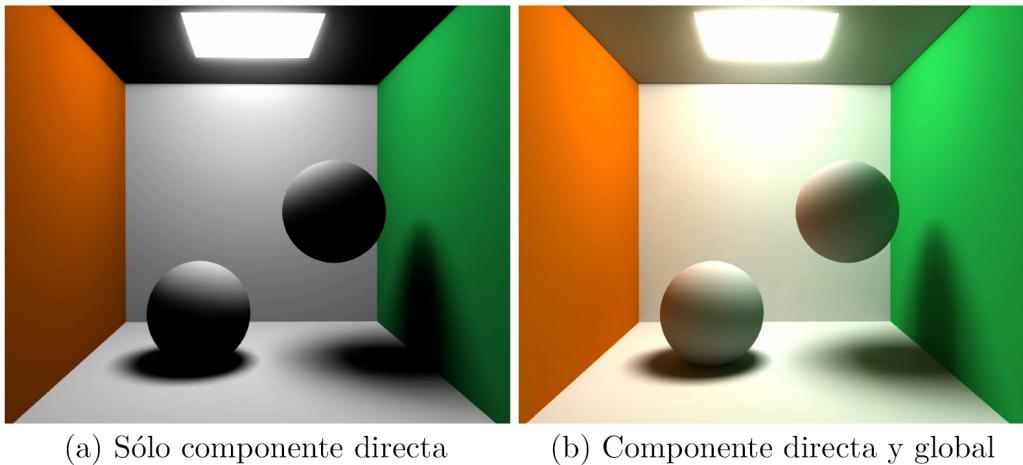


Figura 2.4: *Escena sintética iluminada tan sólo teniendo en cuenta la componente directa en la imagen de la izquierda (a) y teniendo en cuenta tanto la componente directa como la componente global en la imagen de la derecha (b). Imágenes obtenidas de www.digitaltutors.com*

2.2. Luz estructurada

Cuando se habla de emplear luz estructurada se refiere al proceso de proyectar un patrón conocido de píxeles sobre una escena, generalmente barras verticales y/o horizontales formando algún tipo de mallado regular. La manera en la que dicho patrón se deforma cuando se aplica sobre una superficie concreta permite a los sistemas de visión por computador obtener cierta información sobre los objetos que componen la escena, como la geometría de los mismos o su profundidad.

La figura 2.5 muestra un ejemplo de un patrón de rayas verticales proyectado sobre una cara humana. Se observa claramente cómo dichas rayas no sufren deformación alguna cuando inciden directamente sobre el fondo, pero sí lo hacen cuando alcanzan el rostro expuesto delante, revelando así información acerca de la geometría del mismo.

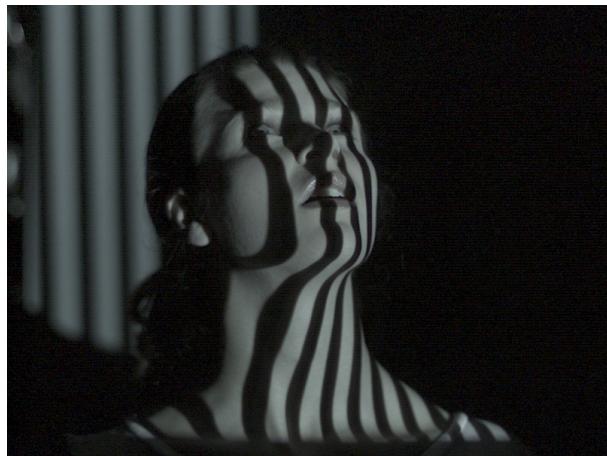


Figura 2.5: *Un patrón de rayas verticales es proyectado sobre una cara humana. Imagen obtenida de [JD06].*

De igual manera que en este trabajo, otros muchos estudios han empleado la técnica de la luz estructurada para obtener información adicional de una escena dada. Así pues, en Scharstein et. al. [SS03] se emplea también un proyector LCD para generar un patrón de alta frecuencia binario y de este modo obtener un mapa de profundidad de la escena, tal y como se observa en la figura 2.6. En Lanman et. al. [LT07] y en Aliaga et. al. [AX08] por contra, el objetivo es obtener una reconstrucción 3D de la escena y en Silberman et. al. [SF11] el empleo de la luz estructurada permite realizar una segmentación por objetos de la escena. Como puede observarse, se trata de una técnica muy popular y utilizada dentro del campo de la fotografía computacional.

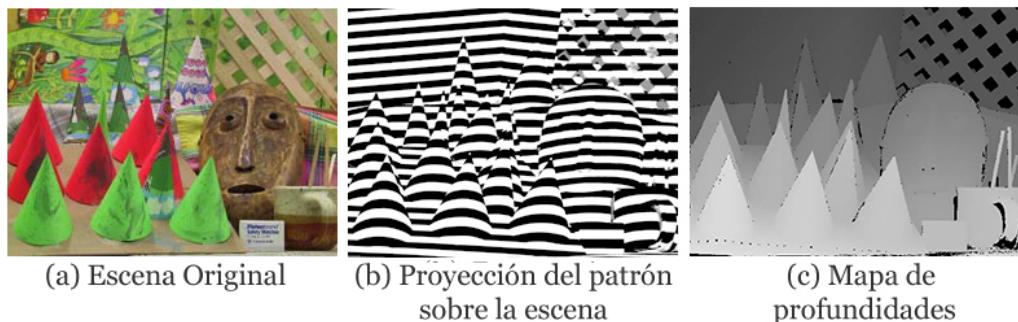


Figura 2.6: Técnica empleada en Scharstein et. al. [SS03] para obtener el Mapa de profundidades de una escena. La figura (a) muestra la escena original. En la figura (b) se muestra un patrón de alta frecuencia horizontal proyectado sobre la escena original. Finalmente, la figura (c) muestra el mapa de profundidades obtenido. En dicho mapa se representan con tonos claros los objetos cercanos y con tonos más oscuros los objetos lejanos.

Capítulo 3

Implementación

En este capítulo se detalla el algoritmo de descomposición empleado. Se exponen dos métodos distintos para realizar la descomposición mediante el uso de luz estructurada en un entorno estático, y se muestra cómo también es posible obtener dicha descomposición en un entorno móvil no controlado gracias al uso de el tablet NVIDIA Tegra.

3.1. Algoritmo de descomposición

Una posible forma de realizar la descomposición de la escena en sus dos componentes de iluminación consiste en medir la componente global de la escena por separado y calcular la componente directa después a partir de los datos obtenidos. De acuerdo con Seitz et. al. [MK05] esto es posible si se ilumina cada pequeño punto de la escena individualmente y se captura una imagen para determinar la contribución de dicho punto al resto de la escena. De esta forma, punto por punto es posible construir la componente global. Sin embargo, a pesar de que el método es teóricamente válido, resulta prohibitivamente costoso dada la gran cantidad de capturas requeridas.

En este proyecto se reproduce el algoritmo planteado por Nayar et. al [NR06] que permite, gracias a la técnica de la luz estructurada, realizar la descomposición mediante el uso de patrones de alta frecuencia de forma eficiente, ya que no es necesario iluminar la escena punto por punto.

Para una escena cualquiera, se describe la iluminación de la misma mediante el término $L[x, \vec{w}]$, que mide la cantidad de luz emitida por el punto x

en la dirección \vec{w} . Si se denota que el vector \vec{w} viaja desde un punto x hasta otro punto y se puede rescribir el término anterior como $L(w_x^y)$. Parte de la luz emitida por dicho punto x impacta contra otros objetos de la escena y es reflejada. A su vez, esta luz reflejada es re-emitida en una nueva dirección pudiendo alcanzar de nuevo otra superficie. Estos rebotes de la luz se suceden hasta que la luz se expande por completo. Por tanto, $L(w_x^y)$ se puede dividir en dos componentes: la luz que ha sido emitida una única vez directamente desde la fuente (iluminación directa) y la luz que ha sido rebotada dos o más veces (iluminación global).

$$L(w_x^y) = L^1(w_x^y) + L^{2,3,\dots}(w_x^y) \quad (3.1)$$

La primera componente, $L^1(w_x^y)$, está determinada por la función BRDF del punto x . La segunda componente, $L^{2,3,\dots}(w_x^y)$ depende de la luz reflejada que impacta el punto x desde otros puntos de la escena. La ecuación 3.1 se puede expresar como una ecuación integral, conocida en el ámbito de la informática gráfica como la *ecuación de renderizado* [KAJ86].

$$L(w_x^y) = L^1(w_x^y) + \int_{x'} A(w_x^y, w_{x'}^x) L(w_{x'}^x) dx' \quad (3.2)$$

La función $A(w_x^y, w_{x'}^x)$ define la cantidad de luz que el punto x refleja hacia el punto y cuando la recibe desde el punto x' . Si $x = x'$ se tiene que $A(w_x^y, w_{x'}^x) = 0$, es decir, la contribución lumínica en dicho punto se debe únicamente a la iluminación directa. Supóngase ahora que la superficie de la escena que se desea descomponer se divide en un numero finito de pequeñas porciones. La ecuación de renderizado 3.2 se puede ahora escribir de modo discreto como:

$$L[c, i] = L_d[i] + \sum_j A[i, j] L[i, j] \quad (3.3)$$

Donde $L[c, i]$ denota la iluminación de cada porción concreta de la escena i medida por la cámara c y $L_d[i]$ corresponde a la componente directa anteriormente denominada como $L^1(w_x^y)$. El segundo término de la ecuación corresponde, por tanto, a la componente global de la escena, definida como:

$$L_g[c, i] = \sum_{j \in P} A[i, j] L[i, j] \quad (3.4)$$

donde, $P = \{j | 1 \leq j \leq N, j \neq i\}$. $L[i, j]$ mide la iluminación de la porción i, j y la matriz $A[i, j]$ incorpora tanto la función BRDF de i como la configuración geométrica relativa de i, j . Es posible además separar la componente global, $L_g[c, i]$, en dos componentes de modo que $L_g[c, i] = L_{gd}[c, i] + L_{gg}[c, i]$, donde $L_{gd}[c, i]$ se debe a los rayos de luz rebotados por primera vez tras provenir directamente de la fuente de luz y $L_{gg}[c, i]$ se debe a los rayos de luz rebotados por segunda vez o sucesivas.

$$\begin{aligned} L_{gd}[c, i] &= \sum_{j \in P} A[i, j] L_d[i, j], \\ L_{gg}[c, i] &= \sum_{j \in P} A[i, j] L_g[i, j]. \end{aligned} \quad (3.5)$$

Considérese ahora que tan sólo una fracción α de los píxeles de la fuente de luz se proyectan sobre la escena y que dichos píxeles están uniformemente distribuidos sobre la escena completa para producir un patrón de alta frecuencia tal y como muestra la figura 3.1. Se dice que dichos píxeles se encuentran *activos*, mientras que aquellos píxeles que no proyectan rayos de luz se denominan *inactivos*.

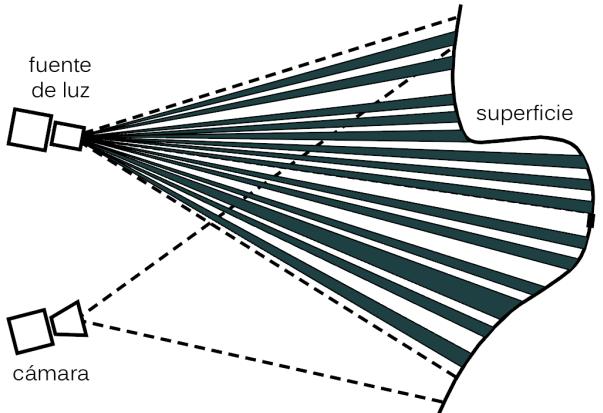


Figura 3.1: Cuando la fuente de luz proyecta un patrón de alta frecuencia sobre la escena, aquellas porciones que estén iluminadas directamente incluirán tanto la componente directa como la global, mientras que las porciones no iluminadas directamente contienen únicamente componente global.

El conjunto de aquellas porciones directamente iluminadas se denota como $Q = \{k | k \in N \text{ y } lit(k) = 1\}$, donde la función *lit* indica si una porción

se encuentra directamente iluminada o no. Así pues, las componentes establecidas previamente quedan definidas de la siguiente forma:

$$\begin{aligned} L_{gd}^+[c, i] &= \sum_{j \in Q} A[i, j] L_d[i, j], \\ L_{gg}^+[c, i] &= \sum_{j \in P} A[i, j] L_g^+[i, j]. \end{aligned} \quad (3.6)$$

Nótese que $L_{gd}^+[c, i]$ difiere de $L_{gd}[c, i]$ en que tan sólo las αM porciones iluminadas contienen componente directa. Por lo tanto, si el término $A[i, j]$ y el término $L_d[i, j]$ (componente directa) son infinitamente diferenciables con respecto de la frecuencia del patrón de iluminación, se tiene que:

$$\begin{aligned} L_{gd}^+[c, i] &= \alpha L_{gd}[c, i], \\ L_{gg}^+[c, i] &= \alpha L_{gg}[c, i]. \end{aligned} \quad (3.7)$$

Se consideran dos capturas diferentes de la escena donde, en la primera de ellas L^+ , la escena está iluminada con un patrón de alta frecuencia en el cual una fracción α de píxeles se encuentran activos y una segunda L^- , iluminada con el patrón de alta frecuencia inverso, es decir, $1 - \alpha$ píxeles activos. Cuando un punto de la escena está iluminado, es decir, sobre él se ha proyectado uno de los α píxeles activos, dicho punto contiene información correspondiente a la iluminación directa y a la iluminación global. Mientras que si dicho punto no está iluminado, éste sólo contiene información relativa a la iluminación global.

Si la porción de la escena i queda iluminada en la primera imagen capturada L^+ entonces no lo estará en la segunda imagen L^- , y viceversa. Así se tiene:

$$\begin{aligned} L^+[c, i] &= L_d[c, i] + \alpha L_g[c, i], \\ L^-[c, i] &= (1 - \alpha) L_g[c, i]. \end{aligned} \quad (3.8)$$

Por consiguiente, si α es conocido, es posible calcular la componente directa y la componente global para cada uno de los píxeles con tan sólo dos capturas. Hasta ahora, se ha supuesto que cuando un píxel de la fuente de luz no se encuentra activo no produce luz en absoluto. Sin embargo, esto no es del todo cierto en el caso de un proyector LCD puesto que no es capaz de bloquear ciertos rayos de luz por completo cuando otros sí están activos (existen pequeñas fugas). Se asume entonces que cada píxel inactivo tiene un

cierto brillo b , donde $0 \leq b \leq 1$. Si se modifican las expresiones anteriores para contemplar este nuevo parámetro quedan de la siguiente forma:

$$\begin{aligned} L^+[c, i] &= L_d[c, i] + \alpha L_g[c, i] + b(1 - \alpha)L_g[c, i], \\ L^-[c, i] &= bL_d[c, i] + (1 - \alpha)L_g[c, i] + \alpha bL_g[c, i]. \end{aligned} \quad (3.9)$$

De nuevo, si α y b son conocidos, la descomposición puede llevarse a cabo con tan sólo dos capturas. Nótese que si el valor de α es o muy próximo a 1 o muy cercano a 0 una de las dos imágenes apenas quedará iluminada. Con la intención de maximizar la frecuencia de iluminación en ambas imágenes, la mejor elección es $\alpha = \frac{1}{2}$, es decir, la mitad de la escena quedará iluminada y la otra mitad no. Finalmente se tiene:

$$\begin{aligned} L^+[c, i] &= L_d[c, i] + (1 + b)\frac{L_g[c, i]}{2}, \\ L^-[c, i] &= bL_d[c, i] + (1 + b)\frac{L_g[c, i]}{2}. \end{aligned} \quad (3.10)$$

Para todos los métodos desarrollados en este trabajo las expresiones anteriores serán válidas para calcular la componente directa y la componente global. Dichos métodos diferirán entre sí en la forma de obtener las matrices de iluminación máxima $L_{max} = L^+$ e iluminación mínima $L_{min} = L^-$.

3.2. Implementación en entorno estático

Para poder aplicar el algoritmo de descomposición previamente expuesto y obtener una descomposición correcta y precisa es requisito indispensable que la escena capturada se encuentre iluminada por una única fuente de luz. Por lo tanto, para que ningún otro rayo de luz interfiera con la fuente emisora, la habitación debe estar acondicionada para permanecer a oscuras durante el proceso de captura. Todas las imágenes empleadas por los sucesivos métodos correspondientes a este primer escenario fueron tomadas a oscuras en el laboratorio 2.06 del edificio Ada Byron de la Universidad de Zaragoza. El proyector empleado fue el modelo JVC DLA-SX21, con un brillo de 1.500 ANSI lumens y la resolución fue fijada a 1024x768 píxeles para proyectar los patrones de alta frecuencia sobre las escenas. Para la obtención de las imágenes fue usada la cámara réflex Canon EOS 500D situada fija sobre la base de un trípode de fotografía para que el objetivo de la cámara quedase alineado con la fuente de luz. Una correcta alineación entre dichos objetos

minimiza la aparición de sombras molestas en los resultados. Las imágenes capturadas se almacenaron en formato RAW, ya que otros formatos de almacenamiento, al aplicar cierta compresión sobre los datos de las imágenes, producían resultados imprecisos (ver sección 4).

3.2.1. Descomposición a partir de varias capturas

De acuerdo con el algoritmo de descomposición presentado al comienzo de este capítulo, dos imágenes deberían ser suficientes para poder obtener la separación buscada: una primera captura con el patrón de alta frecuencia proyectado sobre la escena y una segunda empleando el patrón inverso o *complementario*. Este patrón complementario contiene casillas blancas donde antes había casillas negras y viceversa.

Fusionando estas dos capturas se tiene que, teóricamente, cada uno de los píxeles que conforman la escena ha sido capturado tanto bajo la influencia de las casillas negras del patrón (presentando sólo información correspondiente a la componente global) como bajo la influencia de las casillas blancas (presentando información correspondiente a la componente directa y a la componente global simultáneamente). Sin embargo, en la práctica existen dos problemas relacionados con el proyector LCD. En primer lugar, dado que la escena capturada puede presentar varios objetos, unos más alejados que otros, el patrón de alta frecuencia no puede quedar perfectamente enfocado sobre todos ellos. En segundo lugar, existen pequeñas fugas de luz entre las casillas del patrón causando que las casillas negras no sean suficientemente oscuras en toda su dimensión. Como se observa en la figura 3.2, las casillas negras del patrón se debilitan en las zonas fronterizas y sólo una pequeña parte de ellas resulta realmente útil para realizar la descomposición.

Para superar estos inconvenientes, se tomaron un mayor número de capturas de las que la teoría requiere. En esta fase del trabajo, se usó un patrón de ajedrez cuyos cuadrados eran de un tamaño de 8x8 píxeles. Para un reparto óptimo de las zonas útiles de las casillas del patrón, éste fue desplazado cinco veces (con un desplazamiento de 3 píxeles) en cada una de las dos dimensiones espaciales para capturar un total de veinticinco imágenes, tal y como muestra la figura 3.3. Si no se realiza correctamente este barrido a lo largo de la escena, o si se realizan un menor número de capturas, no todos los píxeles cuentan con la información adecuada y se obtienen resultados de poca calidad como muestra la figura 3.4.

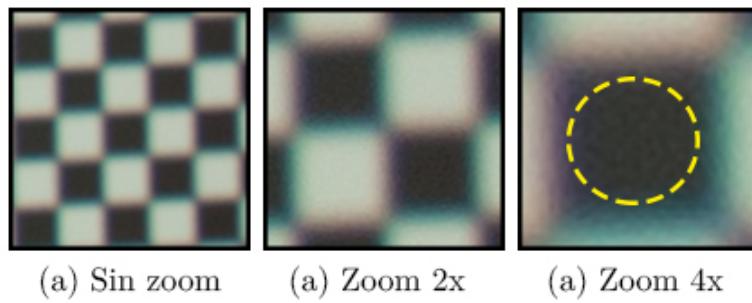


Figura 3.2: Patrón de ajedrez desenfocado sobre el fondo de la escena, sin zoom(a), con zoom 2x (b) y con zoom 4x (c). El círculo amarillo marcado con trazo discontinuo representa la región útil de cada casilla negra del patrón.

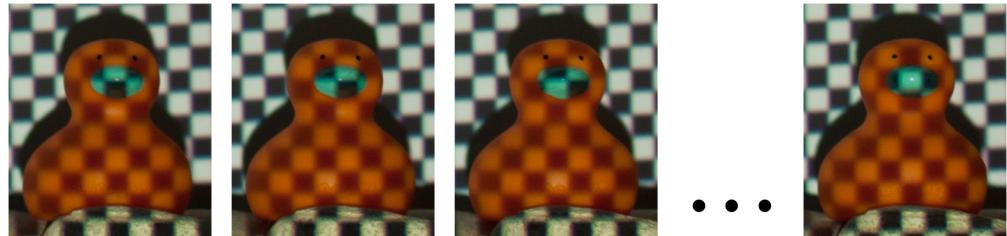


Figura 3.3: Región aumentada de la escena capturada con el patrón de ajedrez proyectado. Se observa cómo el patrón se desplaza en cada captura con respecto de su predecesora hasta conseguir veinticinco imágenes distintas puesto que en ningún momento se repite la distribución del patrón.

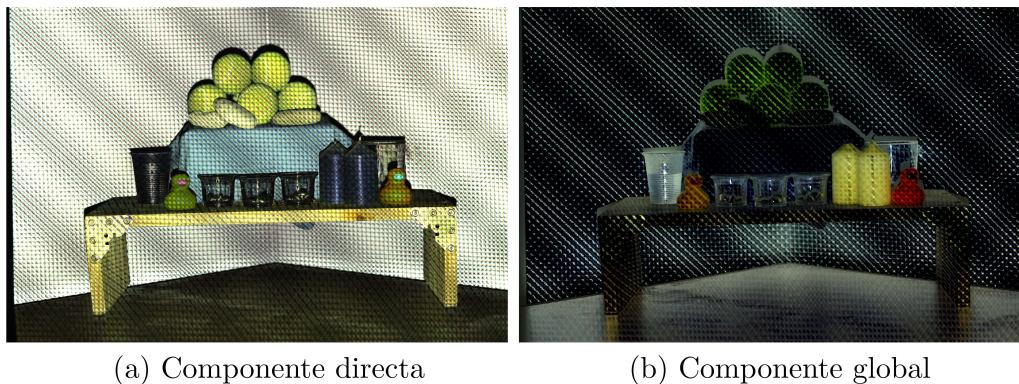


Figura 3.4: El patrón de ajedrez ha sido incorrectamente desplazado a lo largo de la escena y por consiguiente los resultados obtenidos son de baja calidad.

Una vez fueron correctamente capturadas las veinticinco imágenes, se procesaron y se construyeron las matrices de iluminación máxima L_{max} e iluminación mínima L_{min} . Para ello, teniendo en cuenta que todas las capturas cubrían exactamente el mismo área de la escena (la única variación entre ellas era el desplazamiento del patrón de alta frecuencia), se seleccionó píxel a píxel, de entre todas las imágenes disponibles aquél que tuviera el valor más bajo y aquél que tuviese el valor más alto. Una vez se tuvieron L_{max} y L_{min} se obtuvo las componentes directa L_d y global L_g mediante el uso de las ecuaciones 3.10, tal y como muestra la figura 3.5

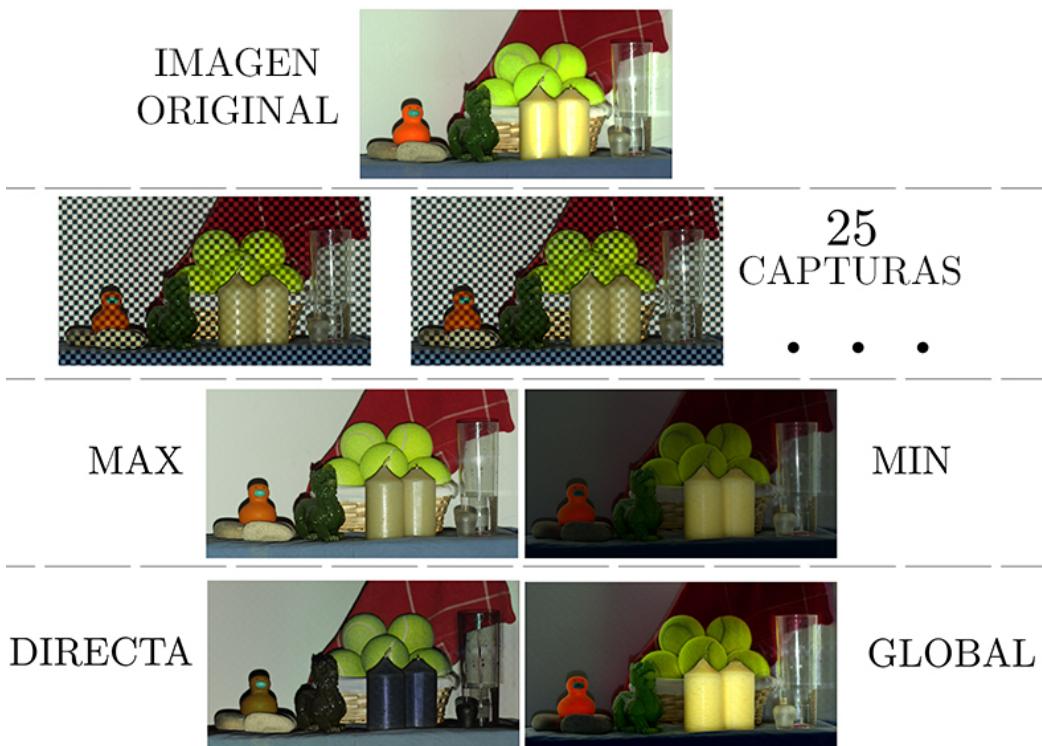


Figura 3.5: *Pasos necesarios para realizar la descomposición con veinticinco capturas y un patrón de alta frecuencia de ajedrez. Los resultados correspondientes a este método se exponen en el capítulo 4.*

3.2.2. Descomposición a partir de una captura

El método anterior era capaz de realizar la descomposición sin perder resolución respecto de las imágenes capturadas; sin embargo, era necesario realizar un elevado número de capturas. Por lo tanto a pesar de permitir obtener resultados de gran calidad, dicho método resulta un tanto tedioso y poco práctico.

No obstante, si se reduce la resolución de las imágenes finales, es posible realizar la descomposición con una única captura de la escena. La idea clave de este nuevo método reside en que, una vez capturada la única imagen de la escena que contiene el patrón de alta frecuencia proyectado, ésta se subdivide en varios fragmentos cuadrangulares. Cada uno de estos fragmentos es lo suficientemente grande como para englobar varios ciclos del patrón de alta frecuencia, y así seleccionar dentro de cada fragmento el píxel con mayor valor y el píxel con menor valor para construir las matrices de menor resolución L_{max} y L_{min} , tal y como ilustra la figura 3.6.

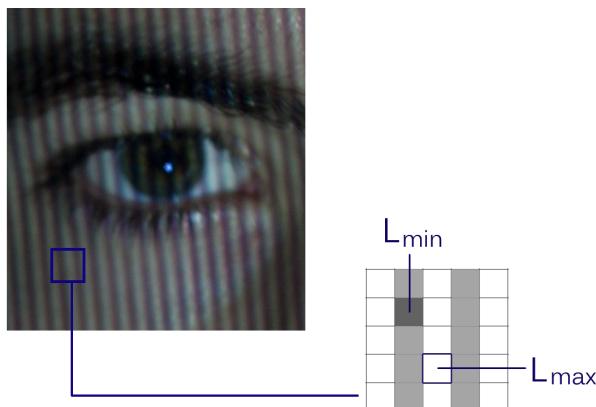


Figura 3.6: *Descomposición usando tan sólo una captura, la cual ha sido dividida en fragmentos de donde se obtienen los valores máximos y mínimos para formar las matrices L_{max} y L_{min} .*

Sin embargo, para que esta idea funcionase adecuadamente sería necesaria una proyección del patrón muy precisa y que la escena a ser capturada no contuviese excesiva cantidad de detalle. Por consiguiente, se cambió el patrón de ajedrez anterior por un patrón que alternase rayas verticales activas e inactivas de 1 píxel de grosor cada una. A pesar del cambio de patrón, esta primera aproximación del método no aporta resultados suficientemente buenos dada la fuerte pérdida de resolución sufrida en las imágenes finales.

Para solventar el problema de la pérdida resolución, se trabajó con la imagen capturada en sus dimensiones completas y a partir de ella se formaron las matrices de iluminación máxima y mínima correspondientes. Para la formación de dichas matrices L_{max} y L_{min} se recorrió la única captura de la escena píxel a píxel y en función del valor de cada uno de los mismos, éste se asignó bien a la matriz de máximos o bien a la matriz de mínimos. Obviamente, como resultado se tuvo que dichas matrices resultantes contenían numerosos píxeles con valor indefinido dado que faltaban alrededor de la mitad de los datos en cada una de ellas, tal y como muestra la figura 3.7.

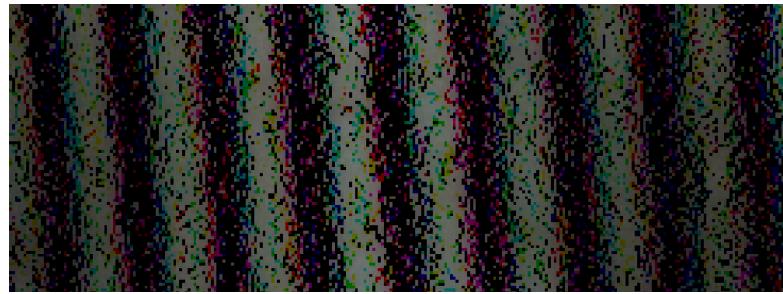


Figura 3.7: *Región aumentada correspondiente a la matriz de iluminación máxima L_{max} obtenida a partir de una única captura. Se observa que antes de aplicar el algoritmo de interpolación multitud de píxeles carecen de información.*

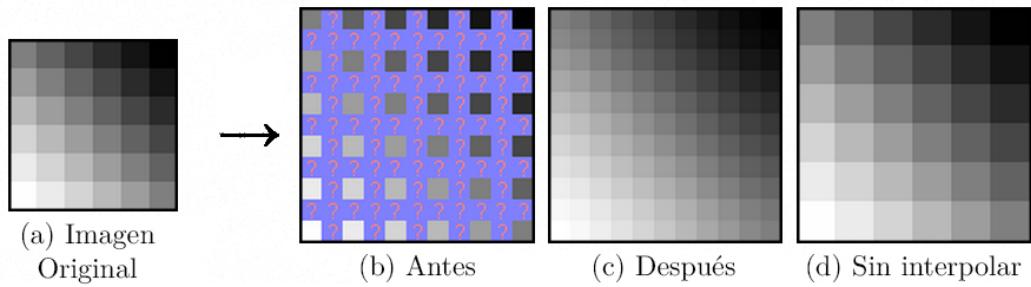


Figura 3.8: *Al ampliar la imagen original (a) se observa cómo aparecen ciertos píxeles cuyo valor se desconoce (b). Las imágenes (c) y (d) muestran cómo queda la imagen redimensionada aplicando y sin aplicar el algoritmo de interpolación respectivamente. Imágenes obtenidas de www.cambridgeincolour.com*

Para poder obtener una descomposición de calidad, fue necesario completar dichos píxeles mediante un algoritmo de interpolación. La interpolación es una técnica que permite estimar valores desconocidos a partir de valores cercanos de los cuales sí se conoce su valor. Lógicamente, cuantos más valores conocidos se tengan en cuenta, más acertada será la estimación. Una aplicación típica de esta técnica es la ampliación de una fotografía, como muestra la figura 3.8.

Las imágenes que se desean interpolar en este método carecen de una gran cantidad de datos (aproximadamente la mitad de sus píxeles tienen valor desconocido). Además, como consecuencia del patrón de alta frecuencia vertical, se tiene que algunos de los píxeles desconocidos encontrarán la información necesaria en sus inmediaciones mientras que otros necesitarán viajar más lejos para obtener información. Estas características hacen que ninguno de los algoritmos de interpolación existentes se comporten adecuadamente para el caso que aquí se contempla, exigiendo por tanto la elaboración de un algoritmo de interpolación propio.

El algoritmo de interpolación desarrollado en este proyecto funciona de la siguiente forma: para cada píxel con valor desconocido se consideran dieciocho de sus píxeles vecinos formando tres niveles distintos tal y como muestra la figura 3.9. A la hora de formar dichos conjuntos, se escogen siempre píxeles con valor definido. En caso de que el píxel elegido no tenga valor, se prueba con el siguiente (p.ej: si el píxel situado inmediatamente a la derecha no tiene valor, se busca el siguiente píxel a la derecha). No obstante, los píxeles que han tenido que ser buscados más allá de su posición original influyen en menor medida. Finalmente, una vez se han formado los tres conjuntos se procede a interpolar el valor desconocido de acuerdo con los coeficientes *coef1*, *coef2* y *coef3* que ponderan cada nivel.

$$\text{interp}(i, j) = \text{nivel1} * \text{coef1} + \text{nivel2} * \text{coef2} + \text{nivel3} * \text{coef3}$$

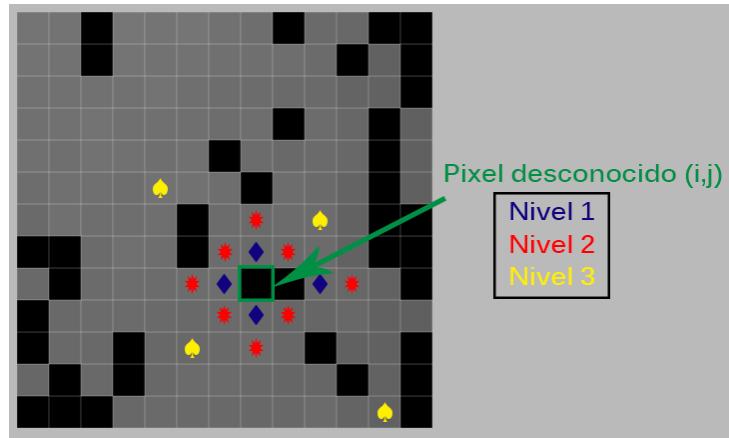


Figura 3.9: *Formación de los niveles con píxeles vecinos con valor conocido usados por el algoritmo de interpolación para estimar el valor del píxel desconocido.*

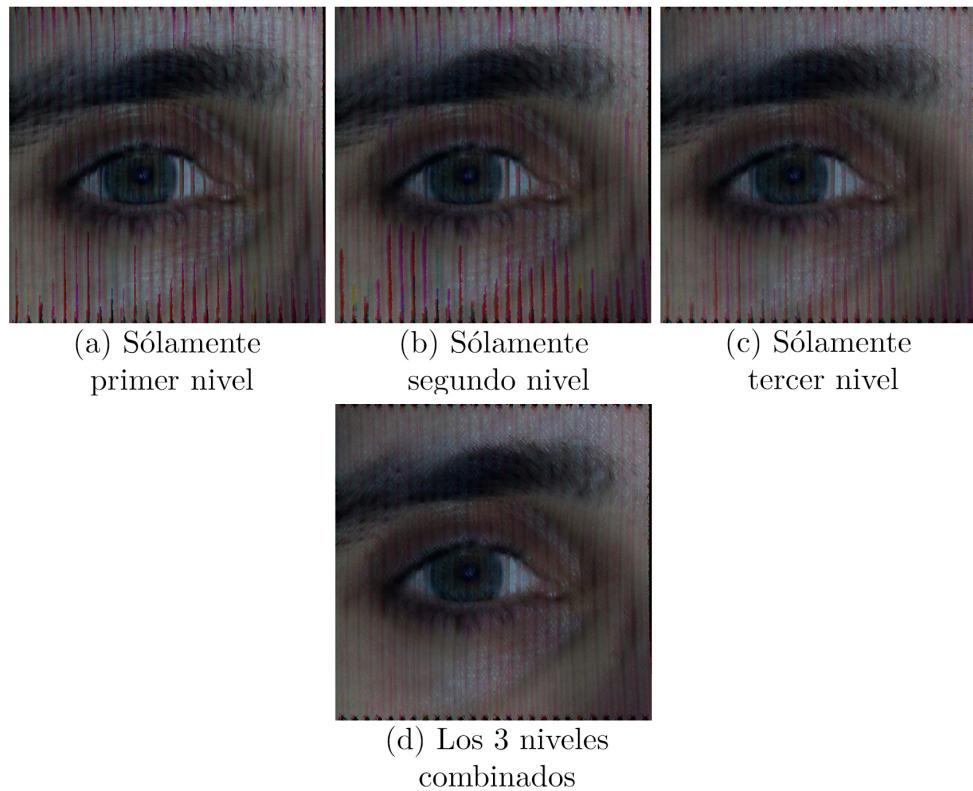


Figura 3.10: *Ningún nivel por sí sólo consigue mejores resultados que una adecuada combinación entre todos ellos.*

Gracias a los coeficientes asociados a cada nivel de píxeles vecinos, el algoritmo de interpolación permite ajustar el peso que recibe cada uno de ellos a la hora de estimar un valor desconocido, pudiendo así decidir si los píxeles más cercanos son más importantes que los píxeles más alejados y viceversa. Una adecuada colaboración entre los tres niveles consigue mejores resultados de los que cualquier nivel podría obtener por separado, tal y como muestra la figura 3.10.

Una vez completado el proceso de interpolación, se obtuvieron las matrices de iluminación máxima L_{max} e iluminación mínima L_{min} correctamente rellenas. A partir de las mismas fue posible calcular las componentes directa L_d y global L_g mediante el uso de la ecuación 3.10, tal y como muestra la figura 3.11

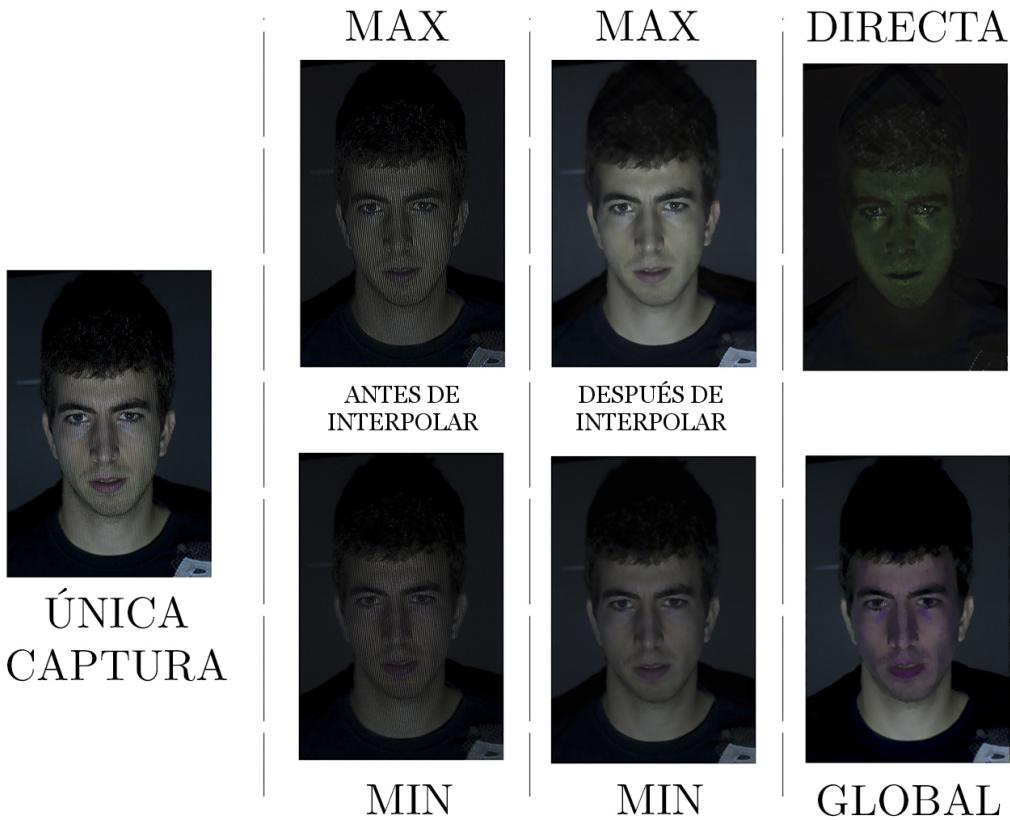


Figura 3.11: *Pasos necesarios para realizar la descomposición con una única captura y un patrón de alta frecuencia rayado. Los resultados correspondientes a este método se exponen en el capítulo 4.*

3.3. Implementación en entorno móvil

Los dos métodos anteriores presentan la clara desventaja de necesitar un entorno controlado, donde el proyector actúa como fuente de luz a la par que proyecta el patrón de alta frecuencia sobre la escena para realizar las capturas con la cámara digital. Este tercer método permite realizar la descomposición de cualquier escena a plena luz del día, mediante el uso del tablet NVIDIA y un objeto alargado opaco como puede ser una simple varilla de madera tal y como se puede ver en la figura 3.12. Los detalles técnicos acerca de la implementación de la aplicación móvil desarrollada en Android para realizar la descomposición en el tablet NVIDIA se encuentran en el anexo A.



Figura 3.12: *Configuración del entorno móvil. El tablet NVIDIA realiza la grabación de la escena sobre la cual se proyecta y se desplaza de lado a lado la sombra de la varilla de madera.*

El Sol es una fuente de luz difusa, es decir, la luz que emite incide sobre los objetos desde múltiples vértices, por lo que proporciona una oscuridad más heterogénea y hace que las sombras sean menos nítidas cuanto más lejos esté un objeto de la superficie que oscurece. La *umbra* es la parte de la sombra donde el Sol queda completamente bloqueado y la *penumbra* es la parte de la sombra donde el Sol está parcialmente ocluido. En nuestro caso, el tamaño relativo de *umbra* y *penumbra* de la sombra proyectada depende de lo alejado que esté la varilla de la escena: cuanto más se aleja la varilla de la escena, la parte de *umbra* es más estrecha mientras que la *penumbra* aumenta.

Para simular el patrón de alta frecuencia correctamente y realizar una buena descomposición de la escena, la varilla debe estar lo suficientemente cerca para que la penumbra no sea demasiado ancha y cada punto sea registrado en umbra en alguna de las imágenes capturadas.

El Sol es por tanto nuestra fuente de luz ahora, y con la ayuda de la varilla de madera se proyecta su sombra sobre la escena de modo que mientras el tablet (en estático) captura una secuencia de imágenes de la escena, la sombra de la varilla se desplaza poco a poco de principio a fin realizando un barrido completo. Para realizar dicho barrido de la escena de forma adecuada, se puede ajustar el parámetro que define el número de fotogramas que contendrá la grabación. También es posible ajustar el intervalo de tiempo existente entre cada fotograma capturado en la grabación. Dicho intervalo de tiempo está establecido en 0.2 segundos, es decir 5fps¹, que es la máxima velocidad de captura que el tablet NVIDIA permite en modo RAW (no vídeo). Por defecto la aplicación realiza un total de 45 capturas, necesitando por lo tanto un total de 10 segundos, tiempo suficiente para hacer dos barridos completos (ida y vuelta) de forma suave y uniforme. Sin embargo, si la sombra proyectada no se desplaza uniformemente sobre la escena capturada puede dar lugar a la aparición de discontinuidades entre las sombras. En aquellos píxeles pertenecientes a las zonas discontinuas no se cuenta con toda la información necesaria para poder realizar la descomposición y por lo tanto los resultados obtenidos serán parcialmente erróneos, tal y como muestra la figura 3.13

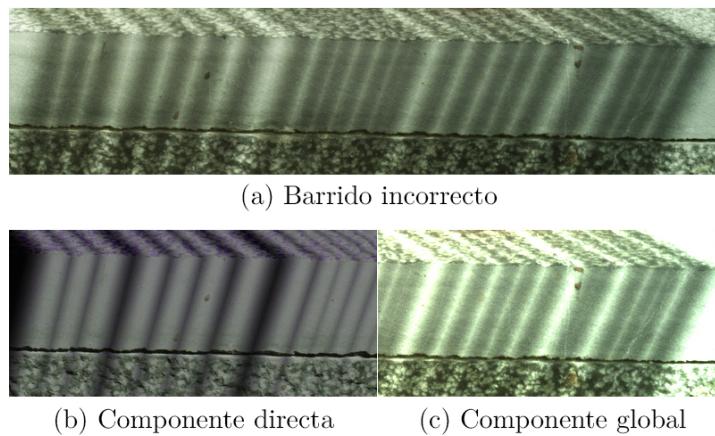


Figura 3.13: *Desplazamiento incorrecto de la varilla sobre la escena provocando la aparición de discontinuidades entre las sombras capturadas.*

¹fps = fotogramas por segundo.

Otro aspecto importante a tener en cuenta es el grosor de la sombra proyectada sobre la escena. Idealmente, cuanto más fina sea la sombra mayor será la frecuencia del patrón simulado mediante el desplazamiento de las sombras de la varilla y por lo tanto más precisos serán los resultados obtenidos. Obviamente, para poder registrar una sombra más fina a lo largo de toda la escena será necesario aumentar el tiempo de grabación conllevando una pérdida de rendimiento dado el crecimiento de datos a procesar.

Con una sincronización adecuada y con la ayuda de todas las imágenes capturadas se pudo, para cada píxel de la escena, obtener su valor máximo (cuando la sombra no le afectaba) y su valor mínimo (cuando estaba bajo la sombra de la varilla). De este modo, se formaron las matrices de iluminación máxima L_{max} e iluminación mínima L_{min} para finalmente, como sucedía en los métodos anteriores obtener las componentes directa L_d y global L_g mediante el uso de la ecuación 3.10, tal y como muestra la figura 3.14.

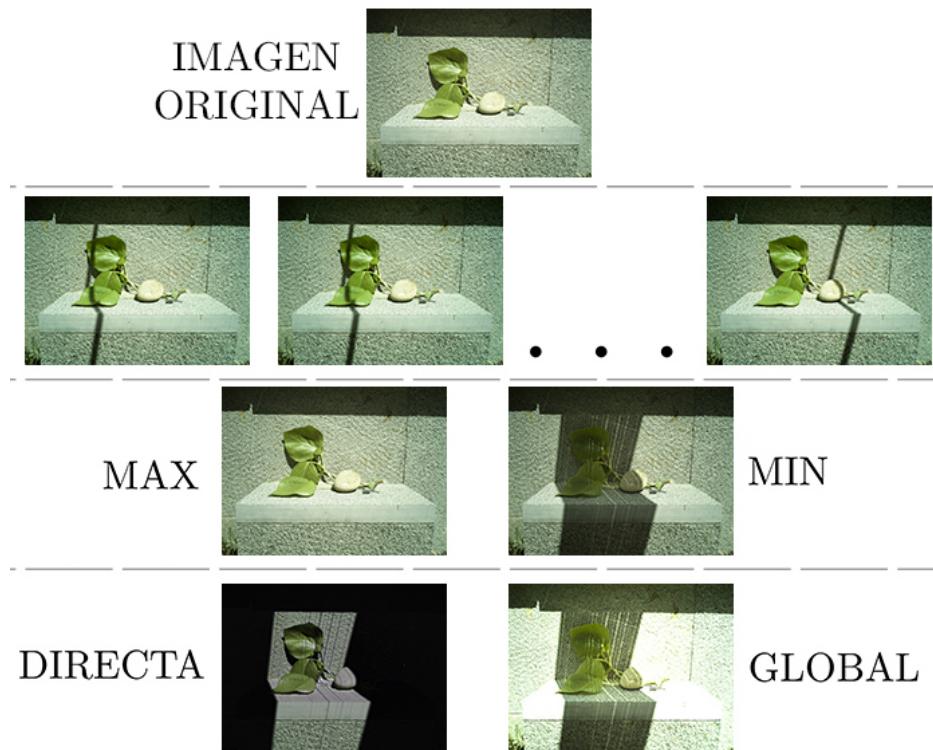


Figura 3.14: Pasos necesarios para realizar la descomposición con el tablet NVIDIA a plena luz del día. Los resultados correspondientes a este método se exponen en el capítulo 4.

Capítulo 4

Resultados

El primer aspecto importante a tener en cuenta es el formato digital con el que se almacenan las imágenes capturadas. A pesar de que el formato más popular es el JPG, éste utiliza un algoritmo de compresión para reducir el tamaño de las imágenes, lo que conlleva una pérdida en la calidad de las mismas. Para no perder información útil y obtener unos resultados de mayor calidad se almacenaron las imágenes mediante el formato RAW, que contiene la totalidad de los datos de la imagen tal y como ha sido captada por el sensor digital de la cámara fotográfica sin ningún tipo de compresión. En la figura 4.1 se observa la clara mejoría en los resultados obtenidos empleando el formato RAW en comparación con JPG. En el resultado de la descomposición obtenido a partir de imágenes JPG se observa pérdida de nitidez y degradación del color. Como consecuencia, algunos de los efectos interesantes quedan camuflados y resultan imperceptibles (p.ej: Apenas se observa que la servilleta colorea la parte superior de las pelotas de tenis como consecuencia de los rayos de luz en ella reflejados. Dicho fenómenos se observa claramente en la descomposición obtenida a partir de imágenes en formato RAW).

El algoritmo de descomposición fue probado en un amplio conjunto de escenas, cada una de ellas con diferentes tipos de materiales para comprobar la robustez del algoritmo. En las figuras 4.2 4.3 4.4 y 4.5 se muestran resultados obtenidos con el primero de los métodos explicados anteriormente, donde se obtenía la descomposición a partir de varias capturas de la escena (ver sección 3.2.1), en las figuras 4.7 y 4.7 resultados del segundo método, en el cual tan sólo se empleaba una única imagen para realizar la descomposición (ver sección 3.2.2) y finalmente en la figura 4.9 resultados obtenidos para la descomposición mediante el tablet NVIDIA Tegra (ver sección 3.3).

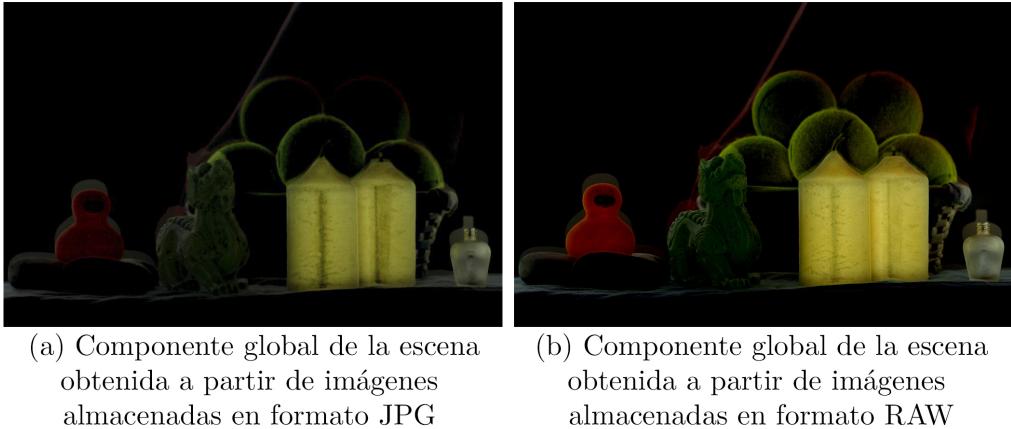


Figura 4.1: *Resultados obtenidos para la descomposición de una escena. En la figura se muestra la componente global empleando imágenes almacenadas con formato JPG (a) y con formato RAW (b).*

Analizando la figura 4.2 objeto por objeto, se observa cómo la apariencia del dragón y de las velas está fuertemente dominada por la gran cantidad de luz que se dispersa dentro de estos objetos y viaja por el interior de los mismos (*subsurface scattering*), tal y como se ve en la imagen de la componente global. Resulta especialmente llamativo cómo el color de las velas surge como consecuencia del efecto del *subsurface scattering*, mientras que únicamente aspectos relativos a cómo el objeto interacciona con la fuente de luz (reflejos y destellos de luz) quedan capturados en la componente directa. Se puede afirmar por lo tanto que los rayos de luz que inciden sobre las velas no son reflejados en la parte externa de la superficie, sino que penetran el objeto, viajan por el interior del mismo y emergen al exterior por otro punto distinto al de incidencia. El mismo fenómeno, aunque en menor medida, es observado en el dragón y en el patito naranja. Por su parte, los dos objetos translúcidos de la derecha (bote pequeño de perfume y vaso parcialmente lleno de agua) se muestran muy oscuros en la componente directa y muy brillantes en la componente global, síntoma claro de que dichos materiales dejan viajar la luz a través de ellos. No obstante, se observa una diferencia entre ambos; mientras que el bote de perfume es difuso, el vaso es bastante especular por lo que aparecen cáusticas a su alrededor. Por último, se puede observar también en la componente global cómo las pelotas de tenis quedan sutilmente coloreadas de rojo como consecuencia de la luz reflejada por la servilleta situada tras ellas. Este fenómeno se conoce como *Color Bleeding*. Estos fenómenos quedan descritos y marcados en la figura 4.3.

Resultados

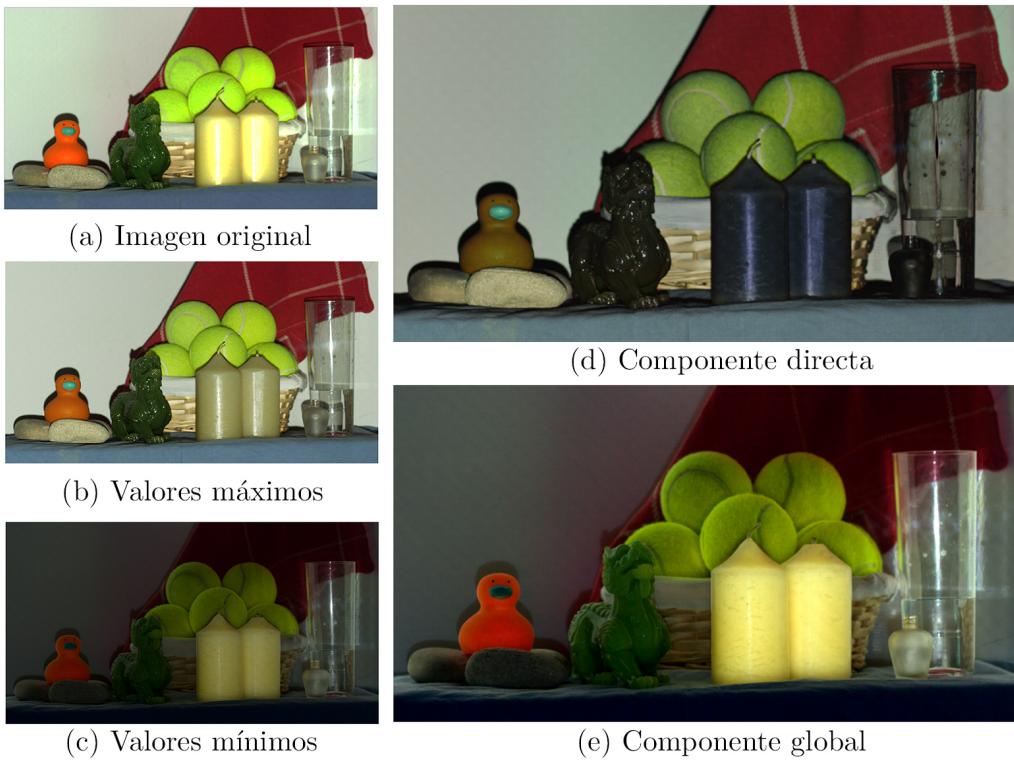


Figura 4.2: Escena con pelotas de tenis y objetos variados. Resultados obtenidos mediante el primero de los métodos (Descomposición a partir de varias capturas).

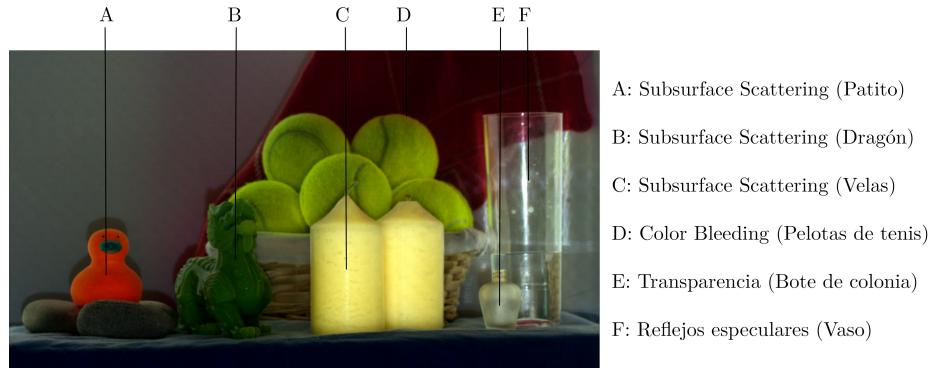


Figura 4.3: Efectos de la luz asociados a cada uno de los objetos pertenecientes a la escena analizada.

Las plantas, como la que aparece en la escena de la figura 4.4, están compuestas de múltiples capas superpuestas una a continuación de la otra formando un tejido tan complejo que cada capa puede poseer un índice de refracción diferente. La luz incidente penetra dichas capas y viaja a través de ellas. Este comportamiento es el responsable de que las hojas se muestren tan verdosas en la componente global. Si la planta transporta agua a través de sus filamentos, ésta actúa como medio participativo alterando la dirección y velocidad de la propagación de la luz. Dicho efecto queda capturado en la componente global haciendo que los filamentos luzcan especialmente brillantes.

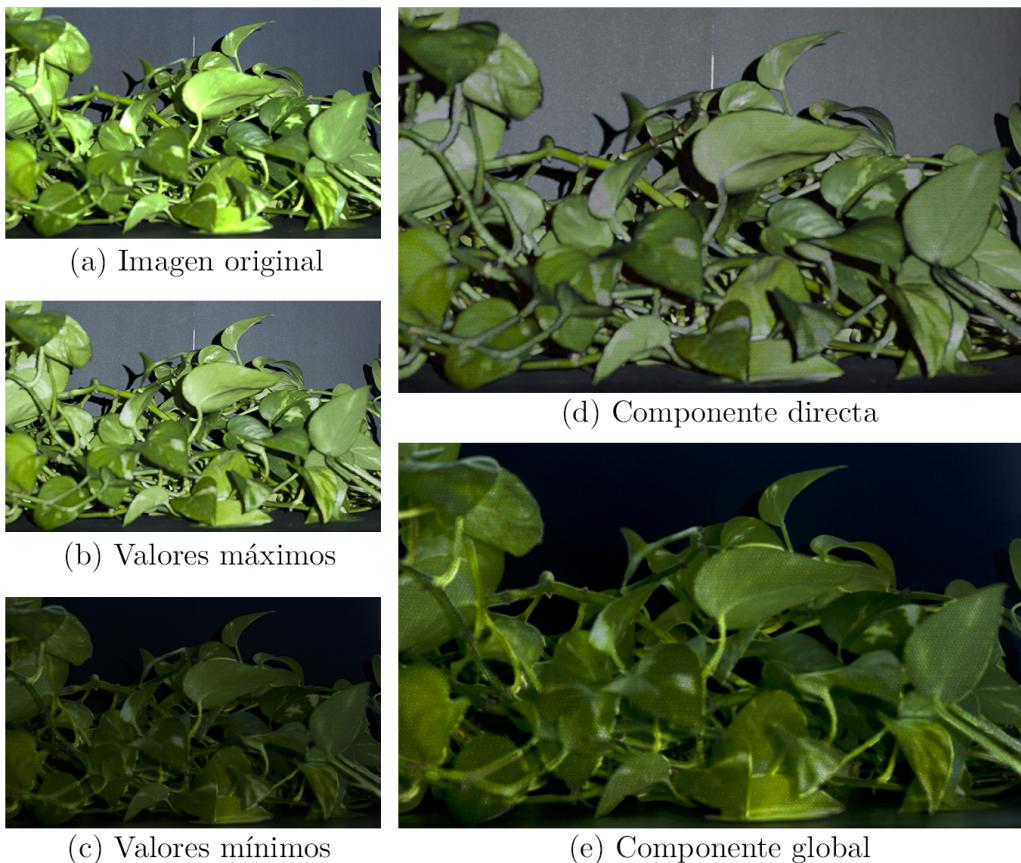


Figura 4.4: Escena con planta natural. Resultados obtenidos mediante el primero de los métodos (Descomposición a partir de varias capturas). En la componente global se observa cómo las hojas de la planta presentan un intenso color verdoso y cómo sus filamentos destacan por su brillo.

En la escena de la figura 4.5 aparecen varios envases translúcidos. Es decir, en mayor o menor medida, todos ellos permiten que los rayos de luz los atraviesen. Esta cualidad es la que hace que dichos objetos se muestren tan oscuros en la componente directa y tan brillantes en la componente global. En especial, el vaso de plástico situado en el centro de la escena sufre una fuerte variación entre la componente global y la componente directa mostrándose prácticamente oscuro en ésta última. Esto es un claro indicio de que su aspecto real, capturado por la cámara u observado por el ojo humano, se debe a la iluminación global, es decir, a los rayos de luz reflejados en los vasos situados próximos al mismo y a los rayos de luz dispersos en el agua que contiene en su interior. En óptica, una cáustica es la envolvente de los rayos de luz reflejados o refractados por una superficie curva. Si se observa detenidamente la componente global de la figura 4.5 se puede notar la aparición de este fenómeno, especialmente alrededor de los envases fuertemente especulares.

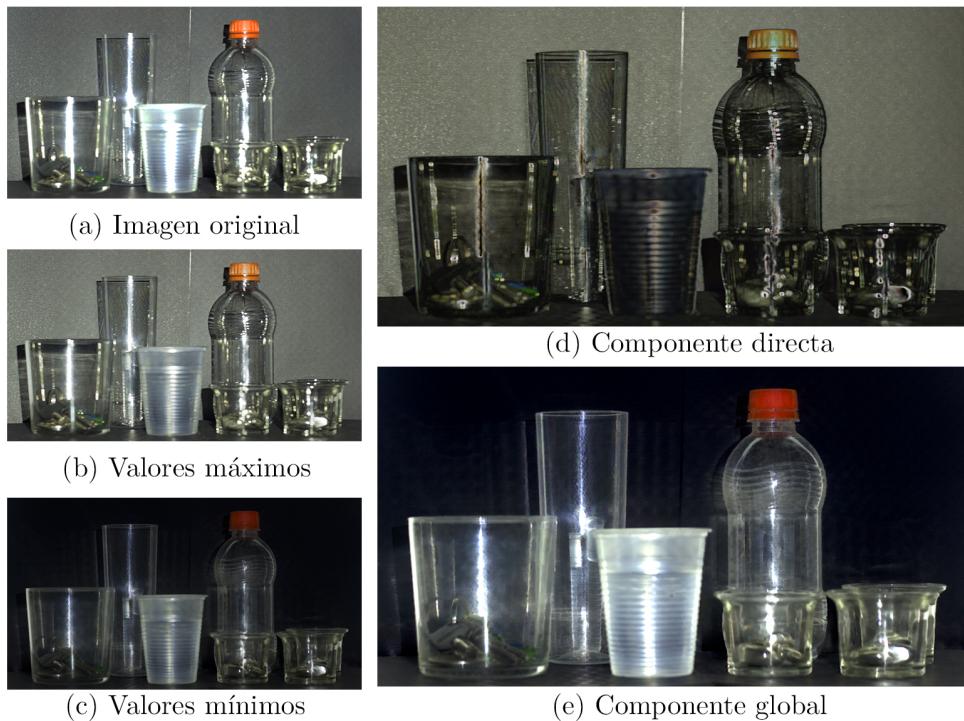


Figura 4.5: *Escena con envases translúcidos. Resultados obtenidos mediante el primero de los métodos (Descomposición a partir de varias capturas).* Los objetos que aparecen en esta escena se muestran oscuros en la componente directa y brillantes en la componente global indicando que dejan pasar la luz través de ellos. Además, alrededor de los objetos que son fuertemente especulares se observa la aparición de cáusticas.

A continuación se exponen los resultados obtenidos a partir de una única captura. En ellos, en lugar de analizar escenas con diferentes objetos y materiales, se han capturado imágenes de rostros humanos. La piel humana es una estructura compleja que se compone básicamente de tres capas distintas: epidermis, dermis e hipodermis, como muestra la figura 4.6. Es la dermis es donde las glándulas sudoríparas y las glándulas sebáceas producen la mayor parte de la grasa visible que presenta la piel. Esta capa intermedia está cubierta por la epidermis, cuyo grosor es variable en función de las zonas corporales, siendo más gruesa en los pies (de 1 a 5 mm) y más fina en la cara (0.02 mm). Esta capa es especialmente delicada en los párpados y la piel del contorno de los ojos (0.004 mm).

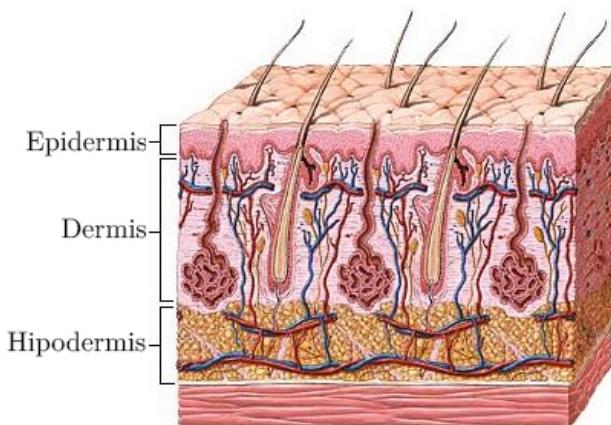


Figura 4.6: *Capas que componen el tejido de la piel humana. Imagen obtenida de la Enciclopedia Ilustrada de Salud (Health Illustrated Encyclopedia) de A.D.A.M*

Las figuras 4.7 y 4.8 muestran los resultados obtenidos para la descomposición en componentes directa y global de las caras de dos varones mediante el segundo de los métodos (una sola captura). En ellas se puede apreciar cómo en la componente directa quedan capturadas las zonas grasas de la piel, que son especialmente visibles en la frente y los párpados inferiores puesto que, como se ha mencionado anteriormente, la epidermis es más fina en esas zonas. Por su parte, en la imagen de la componente global, queda capturada la tonalidad de la piel y el color intrínseco de los labios. De este modo, una vez obtenida la descomposición podría oscurecerse (o aclararse) por separado la componente global y posteriormente fusionarse de nuevo con la componente directa para obtener una simulación acerca de cómo luciría la persona capturada con una tonalidad de piel distinta.



Figura 4.7: *Resultados obtenidos mediante el segundo de los métodos (Descomposición a partir de una captura) para la descomposición de una cara. La componente global revela las zonas grasas de la piel, especialmente localizadas en la frente y párpados inferiores. Por su parte, la tonalidad de la piel y de los labios queda capturada en la componente global.*



Figura 4.8: *Resultados obtenidos mediante el segundo de los métodos (Descomposición a partir de una captura) para la descomposición de otra cara. En este caso los resultados son de peor calidad dado que el sujeto capturado presenta una ligera inclinación respecto de la proyección del patrón provocando la aparición de zonas oscuras en su perfil izquierdo, tal y como se observa en la componente global.*

Finalmente se analizan los resultados logrados mediante el tablet NVIDIA Tegra. La figura 4.9 muestra dichos resultados obtenidos para la descomposición realizada a plena luz del día en un entorno móvil no controlado mediante el uso del tablet NVIDIA Tegra. El objeto empleado para proyectar su sombra a lo largo de la escena fue una fina varilla de madera. Las discontinuidades que se observan en los resultados se deben a que el desplazamiento de dicha varilla fue realizado manualmente. A pesar de estos pequeños saltos, y dado que casi la totalidad de la escena queda cubierta por las sucesivas sombras proyectadas, los resultados obtenidos presentan una calidad aceptable. Así pues, en estas hojas de árbol se pueden contemplar los mismos fenómenos previamente observados en la planta natural descompuesta mediante el primero de los métodos (patrón de ajedrez), es decir, las hojas se muestran muy verdosas y los filamentos se presentan especialmente brillantes en la componente global como consecuencia del líquido que transportan en su interior y que actúa como medio participativo.

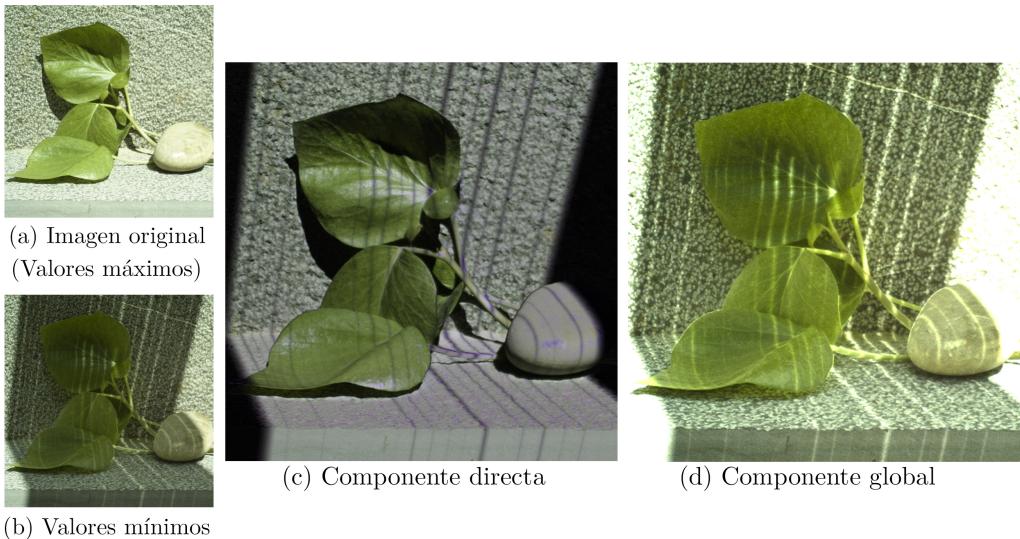


Figura 4.9: *Resultados obtenidos mediante el tercero de los métodos (tablet NVIDIA Tegra) para la descomposición de una rama de árbol.*

Capítulo 5

Conclusiones

A continuación se resume el trabajo realizado en este proyecto y se analizan las posibles líneas futuras de investigación. Además se muestra un resumen temporal del proyecto y las conclusiones y valoraciones personales del autor.

5.1. Trabajo realizado

Una vez finalizado el proyecto, se puede afirmar que se han alcanzado los objetivos establecidos al comienzo del proyecto (ver sección 1.2):

- Se ha implementado el algoritmo de descomposición de una escena en iluminación global e iluminación directa propuesto en Nayar et. al. [NR06] y dos de los métodos propuestos en dicho trabajo (patrón de ajedrez y una sola captura).
- Se ha adaptado dicho algoritmo a un entorno móvil no controlado para poder realizar la descomposición en el tablet NVIDIA Tegra.
- Por último cabe señalar que la exposición de este trabajo ha permitido conseguir una beca de colaboración con el prestigioso Max-Planck-Institut en Saarbrücken, Alemania. Además, destacar también que como consecuencia de la colaboración con la empresa NVIDIA Corp., el Graphics and Imaging Lab ha recibido otros diez tablets prototipos Tegra 3 para ser destinados a la investigación en fotografía computacional.

5.2. Resumen temporal del proyecto

En la figura 5.1 se muestra, mediante un diagrama de Gantt, la evolución temporal que ha sufrido el proyecto desde su comienzo en agosto de 2011 hasta su finalización en julio 2012. Nótese que este proyecto ha sido llevado a cabo conjuntamente con asignaturas pertenecientes al último año de carrera, por lo que se observan pausas correspondientes a cada uno de los períodos de exámenes.

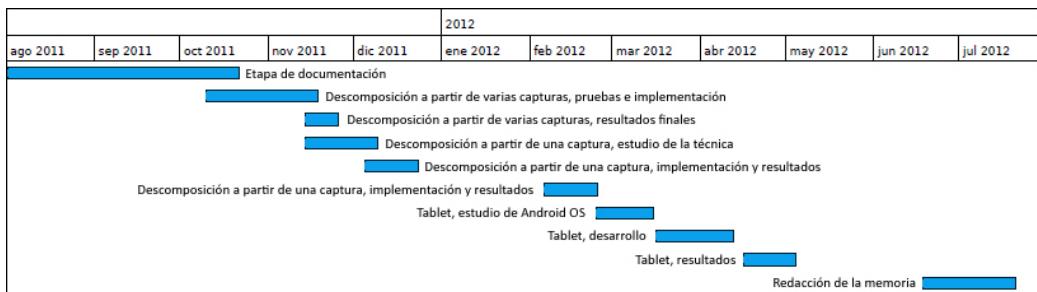


Figura 5.1: *Diagrama de Gantt*.

A continuación se definen las fases principales:

- **Etapa de documentación.** Hasta que fue fijado el propósito del proyecto, se revisaron multitud de trabajos existentes. Poco a poco el radio de acción se fue limitando hasta definir finalmente los objetivos para este trabajo. Una segunda fase de documentación comenzó entonces con ánimo de buscar soluciones relacionadas con ese problema en concreto.
- **Descomposición a partir de varias capturas, pruebas e implementación.** Se comenzó por implementar este método por ser aquél del que se disponía de mayor documentación. Los primeros pasos consistieron en implementar el algoritmo de descomposición en Matlab y obtener los primeros resultados.
- **Descomposición a partir de varias capturas, resultados finales.** Una vez refinado el algoritmo, distintas escenas fueron capturadas para su posterior procesado y así obtener resultados finales de la descomposición para este primer método.
- **Descomposición a partir de una captura, estudio de la técnica.** En cuanto a este segundo método, los detalles internos de su implementación estaban mucho más difusos. Era bien conocido que la descomposición con una sola captura se podía llevar a cabo, pero no estaba muy

claro el cómo. Es por esto que se invirtió cierta cantidad de tiempo en el estudio del mismo.

- **Descomposición a partir de una captura, implementación y resultados.** La implementación de este método, distinta a la propuesta original, causó ciertos quebraderos de cabeza hasta que finalmente se obtuvieron resultados aceptables con el algoritmo de interpolación propio.
- **Tablet, estudio de Android OS.** Llegado el momento de adaptar el algoritmo de descomposición para hacerlo funcionar en el tablet NVIDIA fue necesario un estudio detenido acerca del modelo de ejecución del sistema operativo Android y del entorno de desarrollo para programar aplicaciones válidas para dicho sistema.
- **Tablet, desarrollo.** Además, el desarrollo de la aplicación móvil estuvo fuertemente marcado por el lanzamiento de la segunda versión de las librerías externas de la Frankencamera (FCam) ya que con la primera versión alpha fue imposible compilar el ejecutable final.
- **Tablet, resultados.** Una vez ya se tenía una aplicación sólida y robusta se pasó a capturar varias escenas para obtener los resultados de su descomposición.
- **Redacción de la memoria.** Por último, una vez dado por finalizado el desarrollo del proyecto se procedió a la escritura de este documento.

5.3. Conclusiones y trabajo futuro

El trabajo desarrollado en este proyecto presenta resultados de buena calidad para realizar la descomposición de una escena en sus dos componentes de iluminación tanto en entornos controlados (entorno estático) como en entornos no controlados (entorno móvil).

Sin embargo, en el primero de estos escenarios, era requisito indispensable que la escena estuviese iluminada por una única fuente de luz. Recientemente, Gu et. al [GN11] han presentado un nuevo método que permite realizar la descomposición aquí tratada teniendo en cuenta múltiples fuentes de luz. Analizar dicho trabajo para extender este proyecto con nuevas funcionalidades, como por ejemplo obtener la geometría de la escena, podría resultar interesante.

En cuanto al segundo de los escenarios contemplados, un aspecto que se ha quedado pendiente, es el relativo a la sincronización del tablet NVIDIA con el barrido de la sombra proyectada sobre la escena. La aplicación desarrollada podría mostrar a través de su interfaz gráfica algún tipo de ayuda para que el usuario conozca en tiempo real si la sombra se está desplazando adecuadamente. También podría realizar un rápido pre-procesado de la captura para determinar si los datos son suficientemente buenos para continuar o si por el contrario, convendría repetir el proceso de captura.

5.4. Conclusiones personales

A nivel personal, la experiencia ha sido francamente agradable dado que me ha permitido introducirme en el mundo de la investigación dentro de la informática gráfica, siendo este mi propósito inicial. El hecho de poder abordar un prestigioso trabajo publicado por otros investigadores y ser capaz de entenderlo, reproducirlo e incluso construir sobre él es algo muy gratificante. He aprendido mucho gracias a este proyecto y gracias al Graphics and Imaging Lab, no sólo conocimientos propios del tema sino también relativos al trabajo de investigación en grupo y a la organización.

Bibliografía

- [AL10] JACOBS D. DOLSON J. TICO M. PULLI K. TALVALA E. AJ-DIN B. VAQUERO D. LENSCHE H. HOROWITZ M. PARK S. GELFAND N. BAEK J. MATUSIK W. ADAMS, A. and M. LEVOY. The frankencamera: An experimental platform for computational photography. *In Proc. of SIGGRAPH*, 2010.
- [AO05] FORSYTH D. ARIKAN, O. and J. O'BRIEN. Fast and detailed approximate global illumination by irradiance decomposition. *ACM Trans. on Graph*, pages 1108–1114, 2005.
- [AX08] D. ALIAGA and YI. XU. Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3d modeling. *In CVPR*, 2008.
- [BB89] M. BROOKS and HORN B. *Shape from shading*. 1989.
- [GN11] KOBAYASHI T. GUPTA M. GU, J. and S NAYAR. Multiplexed illumination for scene recovery in the presence of global illumination. *ICCV*, 2011.
- [JD06] GARDNER A. BOLAS M. MCDOWALL I. JONES, A. and P DEVÈBEC. Simulating spatially varying lighting on a live performance. *In Proc. of SIGGRAPH*, 2006.
- [KAJ86] J. T KAJIYA. The rendering equation. *In Proc. of SIGGRAPH*, pages 143–150, 1986.
- [LT07] CRISPELL D. LANMAN, D. and G TAUBIN. Surround structured lighting for full object scanning. *In 6th International Conference on 3D Digital Imaging and Modeling*, 2007.
- [MK05] SEITZ S. MATSUSHITA, Y. and K KUTULAKOS. A theory of inverse light transport. *In Proc. of ICCV, II*, pages 1440–1447, 2005.

BIBLIOGRAFÍA

- [NK91] IKEUCHI K. NAYAR, S. and T KANADE. Shape from interreflections. *IJCV*, pages 173–195, 1991.
- [NR06] KRISHNAN G. GROSSBERG D. M. NAYAR, S. and R RASKAR. Fast separation of direct and global components of a scene using high frequency illumination. *ACM SIGGRAPH*, 2006.
- [SF11] N. SILBERMAN and R FERGUS. Indoor scene segmentation using a structured light sensor. *In Proc. of ICCV.*, 2011.
- [SS02] KAUTZ J. SLOAN, P. and J SNYDER. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *In Proc. of SIGGRAPH*, pages 527–563, 2002.
- [SS03] D. SCHARSTEIN and R SZELISKI. High-accuracy stereo depth maps using structured light. *In Proc. of CVPR*, 2003.
- [WOO80] R WOODHAM. Photometric method for determining surface orientation from multiple images. *ptical Engineering 19, 1 (January)*, pages 139–144, 1980.