

TESIS

a ser presentada el día 3 de mayo de 2010 en la

UNIVERSIDAD DE LA REPÚBLICA

para obtener el título de

MAGISTER EN INGENIERÍA MATEMÁTICA

Resolución del problema de radiosidad usando matrices de rango bajo

Eduardo Fernández

Centro de Cálculo, Instituto de Computación

Facultad de Ingeniería, Universidad de la República

a realizarse el día 27 de mayo de 2010 por el comité de examinadores

| | |
|-----------------------|--------------------|
| Dr. Sergio Nesmachnow | Director de Tesis |
| Dr. Franco Robledo | Director Académico |
| Dr. José Vieitez | Presidente |
| Dr. Álvaro Pardo | |
| Dr. Pablo Musé | |
| Dr. Andrés Almansa | |

Resolución del problema de radiosidad utilizando matrices de rango bajo

Resumen

Esta tesis estudia la resolución del problema de radiosidad utilizando matrices de rango bajo. Los métodos de radiosidad forman parte de las técnicas de iluminación global, que tratan el problema de la generación de imágenes que contemplan la reflexión de la luz en los objetos de una escena gráfica tridimensional. En el problema de radiosidad, se considera que la reflexión de las superficies es exclusivamente lambertiana -superficies opacas que reflejan la luz de manera uniforme en todas direcciones-. Los métodos de radiosidad son utilizados en el diseño arquitectónico, en las animaciones digitales y en el diseño de escenarios tridimensionales para videojuegos, entre otras aplicaciones. Para resolver el problema de radiosidad existen múltiples técnicas, basadas principalmente en la resolución de un sistema lineal de grandes dimensiones. El componente de la matriz del sistema que expresa las propiedades de la escena es una matriz densa, por lo que su construcción, almacenamiento y utilización al resolver el problema de radiosidad plantea problemas informáticos de relevancia. En este trabajo se plantea la sustitución del componente mencionado por una aproximación de rango bajo, con la finalidad de simplificar la construcción del sistema lineal, reducir el volumen de información almacenada y reducir la complejidad de la resolución del sistema lineal. La propiedad de coherencia espacial permite deducir que la matriz sustituida tiene rango numérico bajo. La tesis presenta un relevamiento de técnicas para la construcción de aproximaciones de rango bajo y propone dos algoritmos eficientes con tal fin. Asimismo, se proponen metodologías iterativas y directas para la resolución del nuevo sistema lineal para escenas en blanco y negro, en color, estáticas -donde sólo varían las fuentes de luz- y dinámicas -que incluyen el movimiento de objetos-. El análisis experimental de los nuevos algoritmos muestra que los planteos realizados posibilitan el desarrollo de aplicaciones interactivas en tiempo real que utilicen las técnicas de radiosidad de rango bajo.

Palabras Clave: Radiosidad, tiempo real, coherencia espacial, matrices de rango bajo

Índice general

| | |
|------------------------------------------------------------------------------------|-----------|
| 1. Introducción | 3 |
| 2. El problema de radiosidad | 7 |
| 2.1. Introducción | 7 |
| 2.2. Radiosidad: aspectos físico-matemáticos | 8 |
| 2.2.1. Consideraciones básicas de la luz | 8 |
| 2.2.2. Unidades físicas | 9 |
| 2.2.3. La interacción de la luz con los objetos | 9 |
| 2.2.4. Las ecuaciones integrales de la luz | 12 |
| 2.3. Resolución de la ecuación de radiosidad | 13 |
| 2.3.1. Integración basada en reglas de cuadratura | 14 |
| 2.3.2. Integración basada en elementos finitos | 15 |
| 2.3.3. Integración basada en Monte Carlo | 19 |
| 2.4. Aplicación del método de Galerkin | 20 |
| 2.4.1. Cálculo de los factores de forma | 21 |
| 2.4.2. Propiedades de \mathbf{R} y \mathbf{F} | 24 |
| 2.5. Resolución de la ecuación matricial de radiosidad | 26 |
| 2.5.1. Métodos de relajación | 27 |
| 2.5.2. Métodos no estacionarios | 29 |
| 2.5.3. Métodos jerárquicos | 30 |
| 2.6. Trabajos relacionados | 34 |
| 2.6.1. Antecedentes de coherencia en computación gráfica | 34 |
| 2.6.2. Antecedentes en el uso de matrices de rango bajo en radiosidad . | 35 |
| 2.6.3. Antecedentes de algoritmos de iluminación global de tiempo real | 36 |
| 2.6.4. Conclusiones de los trabajos relacionados | 37 |
| 3. Radiosidad de Rango Bajo | 39 |
| 3.1. Introducción | 39 |
| 3.2. Aproximación de rango bajo a una matriz $n \times n$ | 39 |
| 3.2.1. Descomposición en valores singulares (SVD) | 40 |
| 3.2.2. Descomposición CUR | 41 |
| 3.2.3. DFT, DWT y otras factorizaciones con matrices ortonormales . . | 41 |
| 3.2.4. Notación para las matrices de rango bajo | 42 |
| 3.2.5. Conclusiones de las aproximaciones relevadas | 42 |
| 3.3. Nueva formulación del problema de radiosidad | 43 |
| 3.3.1. Coherencia espacial y su influencia en el rango numérico de \mathbf{RF} . | 43 |
| 3.3.2. Ecuación de radiosidad de rango bajo | 45 |

| | | |
|-----------|-----------------------------------------------------------------------------------------|------------|
| 3.4. | Resolución de la ecuación de radiosidad de rango bajo | 45 |
| 3.4.1. | Resolución iterativa de \tilde{B} | 46 |
| 3.4.2. | Resolución directa de \tilde{B} | 46 |
| 3.4.3. | Estimación del error relativo de \tilde{B} | 48 |
| 3.4.4. | Radiosidad en colores | 50 |
| 3.4.5. | Ecuación RRB en escenas con objetos en movimiento | 52 |
| 3.4.6. | Conclusiones sobre los métodos de radiosidad de rango bajo | 53 |
| 3.5. | Bases para el cálculo de nuevas aproximaciones de rango bajo de \mathbf{RF} | 54 |
| 3.5.1. | Influencia de la coherencia espacial en la emisión y la reflexión | 54 |
| 3.5.2. | Malla gruesa y malla fina de una escena | 58 |
| 3.5.3. | Factores de forma de parche a elemento y de elemento a parche | 58 |
| 3.6. | Algoritmos 2MF y 2MCE | 62 |
| 3.6.1. | Algoritmo 2MF: existencia de mallas con coherencia espacial | 62 |
| 3.6.2. | Evaluación de la existencia de mallas con coherencia espacial | 64 |
| 3.6.3. | Algoritmo 2MCE: generación de mallas con coherencia espacial | 67 |
| 3.7. | Conclusiones del capítulo | 75 |
| 4. | Análisis experimental | 77 |
| 4.1. | Introducción | 77 |
| 4.2. | Relación entre la emisión y la reflexión | 77 |
| 4.2.1. | Evaluación de la sensibilidad entre la emisión y la reflexión | 77 |
| 4.2.2. | RRB basada en mallas con coherencia espacial | 81 |
| 4.3. | Comparación de las factorizaciones | 83 |
| 4.3.1. | Generación de las factorizaciones surgidas de 2MF y 2MCE | 86 |
| 4.3.2. | Generación de las factorizaciones SVD | 86 |
| 4.3.3. | Generación de las factorizaciones CUR | 86 |
| 4.3.4. | Generación de las emisiones de prueba | 87 |
| 4.3.5. | Cálculo del error relativo de radiosidad | 88 |
| 4.3.6. | Resultados obtenidos | 88 |
| 4.4. | Evaluación de eficiencia computacional en CPU y GPU | 90 |
| 4.4.1. | Hardware y Software | 90 |
| 4.4.2. | Casos de prueba | 92 |
| 4.4.3. | Resultados de eficiencia computacional | 92 |
| 4.5. | Escenas con objetos en movimiento | 94 |
| 4.5.1. | Definición de $\Delta(\mathbf{RF})$ y $\Delta(\widetilde{\mathbf{RF}})$ | 96 |
| 4.5.2. | Reducción del rango de $\Delta(\mathbf{RF})$ | 97 |
| 4.5.3. | Cálculo de \hat{B}_F luego de conocer $\Delta(\widetilde{\mathbf{RF}})$ | 98 |
| 4.5.4. | Algunos resultados del cálculo de \hat{B}_F | 98 |
| 5. | Conclusiones y trabajo futuro | 101 |
| 5.1. | Conclusiones | 101 |
| 5.2. | Trabajo futuro | 104 |

Índice de principales conceptos

| | |
|----------------------------------------------|----|
| algoritmo del hemicubo | 22 |
| coherencia | 34 |
| coherencia espacial | 43 |
| descomposición CUR | 41 |
| descomposición SVD | 40 |
| DFT | 41 |
| DWT | 41 |
| ecuación matricial de radiosidad | 19 |
| ecuación de radiosidad | 12 |
| ecuación de radiosidad de rango bajo | 45 |
| ecuación de rendering | 12 |
| ecuación integral de Fredholm | 12 |
| ecuación RRB | 45 |
| ecuación RRB con geometría fija | 48 |
| ecuaciones RRB en colores con geometría fija | 52 |
| parches y elementos | 58 |
| factor de forma | 18 |
| GPU | 36 |
| mallado de grano fino | 58 |
| mallado de grano grueso | 58 |
| mallados compatibles | 58 |
| mallados con coherencia espacial | 61 |
| matriz de radiosidad | 25 |
| métodos basados en elementos finitos | 15 |
| método de Galerkin | 17 |
| método de Gauss-Seidel | 27 |
| método de Jacobi | 27 |
| método de puntos de colocación | 16 |
| método de radiosidad | 3 |
| métodos iterativos | 27 |
| métodos de Monte Carlo | 19 |
| métodos de relajación estocástica | 28 |
| parche con coherencia espacial | 61 |
| problema de radiosidad | 3 |
| radio espectral | 25 |
| radiosidad | 9 |
| radiosidad jerárquica | 30 |
| rango numérico | 39 |
| reflexión difusa | 10 |
| reflexión especular | 10 |
| reflexión lambertiana | 10 |
| wavelets | 33 |

Índice de figuras

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1. Modelos de interacción entre la luz y las superficies. La piel y el mármol están mejor representados en (a) y los metales y materiales dieléctricos en (b). | 10 |
| 2.2. Haz de luz con área de sección dA y área de intersección $dA/\cos\theta$ | 10 |
| 2.3. Distintos tipos de reflexión. Los dos tipos superiores corresponden a superficies rugosas o con difusión debajo de la superficie. Los dos tipos inferiores corresponden a superficies metálicas o dieléctricas, como el agua y el vidrio. | 11 |
| 2.4. Elementos considerados para el cálculo de la función $G(x, x')$ | 13 |
| 2.5. Ejemplo de fórmula analítica exacta para la transferencia de radiosidad entre dos objetos (figura extraída de Lienhard y Lienhard [52]). | 13 |
| 2.6. Cálculo geométrico de factores de forma (extraídas de Cohen et al. [15]). | 21 |
| 2.7. El hemicubo (figura extraída de Cohen et al. [15]). | 22 |
| 2.8. Cálculo de factores de forma con Monte Carlo. | 24 |
| 2.9. Representación gráfica del método de Malley. | 29 |
| 2.10. División del parche A para reflejar mejor el intercambio de energía con cada parche de la escena. | 31 |
| 2.11. División de dos parches cercanos para reflejar correctamente el intercambio de radiosidad existente entre ellos. | 32 |
| 2.12. Quadtree de uno de los parches de la Figura 2.11. | 33 |
| 2.13. Primeros miembros de la base de Haar y aproximación de una función por esos primeros miembros. | 34 |
| 3.1. Dos parches resaltados en (a) y la escena vista desde ellos ((b) y (c)). . . | 44 |
| 3.2. Descomposiciones SVD de \mathbf{RF} y de $\mathbf{I} - \mathbf{RF}$ | 44 |
| 3.3. Una misma escena con dos implementaciones de radiosidad en colores. . . | 50 |
| 3.4. Pequeñas variaciones en C implican grandes variaciones en E | 55 |
| 3.5. Dos E sensiblemente diferentes generan dos C prácticamente iguales. . . | 56 |
| 3.6. Dos reflexiones muy parecidas generan emisiones extremadamente distintas. 57 | |
| 3.7. Producto matricial de dos columnas de \mathbf{U}_k por dos filas de una matriz \mathbf{V}_k^T dispersa. Cada columna de \mathbf{U}_k se replica cuatro veces en la matriz producto. | 59 |
| 3.8. Elementos que determinan el factor de forma entre dos áreas. | 60 |
| 3.9. Entre la pared y el piso se pierde coherencia espacial, forzando a generar mallas especiales. | 65 |
| 3.10. En (a) los triángulos de la escena son muy grandes. En (b) y (c) se representa la misma escena con mallas gruesas de distintas densidades. . . . | 68 |

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.11. Cinco niveles de subdivisión de un parche. | 69 |
| 3.12. Dos representaciones de la estructura jerárquica. En (a) cada parche nivel cero genera un árbol. En (b) sólo se aprecia la jerarquía. | 70 |
| 3.13. Estructura de datos para implementar el método jerárquico. | 70 |
| 3.14. Acciones que se realizan en cada nodo de la estructura jerárquica. | 73 |
| 3.15. Los parches se distribuyen entre el nivel cero y <i>maxnivel</i> . Estos nodos así como sus ancestros (que están en la zona de color gris claro) pertenecen a la categoría <i>tiene parches</i> | 73 |
| 3.16. Parches generados al aplicar el algoritmo 2MCE | 74 |
| 4.1. Esquema de una escena con 896 parches. | 78 |
| 4.2. Diferentes imágenes que muestran el comportamiento de la reflexión de los parches ante emisiones diferentes. En (a) y (b) se muestran las emisiones, en (c) y (d) las reflexiones correspondientes, en (e) se muestra la diferencia entre ambas reflexiones amplificadas por un factor de 20, y en (f) cómo se ven las reflexiones si son interpoladas con Gouraud (por ser muy parecidas sólo se muestra una de las interpolaciones). | 79 |
| 4.3. Mallas de grano fino y grano grueso utilizados en el cálculo de la aproxi- mación $\mathbf{U}_k \mathbf{V}_k^T$ de la matriz \mathbf{RF} | 81 |
| 4.4. Etapas del método directo de radiosidad basado en la ecuación RRB con geometría fija, con la aproximación a \mathbf{RF} basada en mallas con coherencia espacial. | 82 |
| 4.5. Etapas del método iterativo de radiosidad. En (a) se muestra la emisión inicial y en (b) se realiza el pasaje de la emisión de malla fina a malla gruesa. En (c) se culmina la primera iteración. De (d) a (g) se realizan los pasos dos y tres de la iteración. En (h) e (i) se muestra el resultado final en malla gruesa y malla fina respectivamente. | 83 |
| 4.6. Escenas generadas para evaluar las factorizaciones de la matriz \mathbf{F} | 84 |
| 4.7. En (a) se muestra una vista de la escena. El resto de las figuras muestran distintas vistas de la escena, durante un recorrido que atraviesa la escena. Cada parche tiene un color único que permite generar un vector E por imagen. | 87 |
| 4.8. Una escena donde un cubo se mueve de una posición inicial a otra final. | 94 |
| 4.9. La matriz $\Delta\mathbf{F}$ y los parches de la escena con el cubo en su posición final. | 95 |
| 4.10. Matriz $\Delta(\widehat{\mathbf{RF}})$ que surge de la resta de dos aproximaciones de rango $k = 860 \mathbf{R}\mathbf{U}_F \mathbf{V}_F^T - \mathbf{R}\mathbf{U}_I \mathbf{V}_I^T$ | 96 |
| 4.11. Generación de la matriz aproximada a $\Delta\mathbf{RF}$ a partir del producto de matrices de rango bajo. | 98 |
| 4.12. Cálculo de B_f , \tilde{B}_f y \hat{B}_f para una iluminación dada. | 99 |

Índice de cuadros

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.1. Factorizaciones realizadas para evaluar la conveniencia de 2MF y 2MCE con respecto a SVD y CUR. | 85 |
| 4.2. Promedio de los errores de radiosidad para cada factorización en norma de Frobenius. | 88 |
| 4.3. Distancia de las factorizaciones a las matrices \mathbf{F} evaluadas. | 89 |
| 4.4. Comparación de factorizaciones con aproximadamente la misma cantidad de información para una misma escena. Se comparan los errores relativos del cálculo de B en las factorizaciones CUR, 2MF y 2MCE con respecto al error en la factorización SVD con rango $\approx \frac{k}{2}$ | 89 |
| 4.5. Combinaciones posibles de n y k consideradas en los experimentos. . . . | 92 |
| 4.6. Números de punto flotante y memoria utilizada en cada uno de los casos considerados en los experimentos. | 93 |
| 4.7. Tiempos de ejecución para el cálculo de \tilde{B} | 93 |

Agradecimientos

Quiero expresar mi más profundo agradecimiento a mi director de tesis Sergio Nesmachnow y a mi director académico Franco Robledo, quienes me convencieron de que las investigaciones que venía realizando podían enmarcarse en las actividades de una tesis de maestría, tuvieron la valentía de guiarme en el proceso y me dieron más tiempo del que merecía para bajar a tierra mis ideas, soportando mis ausencias e impertinencias. Del mismo orden son mis agradecimientos a los compañeros del Centro de Cálculo que ayudaron en la tesis: a Pablo, por su insistencia en que publique y en ser un oído atento cuando lo necesitaba; a Tomás, Aldo y Mauricio, por encargarse del Curso de Computación Gráfica Avanzada en el semestre que pedí libre; a Martín, por ayudarme con algunas formalidades de la tesis; a los demás compañeros, por hacer agradable lo cotidiano y organizar asados en lo de Gerardo.

En el ámbito de las ideas, un agradecimiento especial le corresponde a Ian Ashdown, quién a través de su trabajo académico y su correspondencia me señaló que matrices de bajo rango y radiosidad se llevan bien, y me convenció de que en radiosidad no estaba todo hecho.

En el terreno familiar, mis agradecimientos superan el infinito y el más allá. En primer lugar quiero agradecer a mis amadas hijas Juliana y Natalia, a quienes les robé demasiado tiempo, demasiado... (qué sería de mí sin sus alegrías y sus penas), y a mi compañera de vida Mariela, que ha sido la base de la familia mientras yo andaba distraído (más distraído), haciendo garabatos en hojas sueltas, de mal humor y llevando la compu a las vacaciones, entre otros desaires ingratos. Como hijo de una familia con padre, madre, cuatro hermanos, diez sobrinos, cuñados, suegra, etc., tener un concepto con cada uno es aburrido y un poco injusto. Por lo tanto, voy a agradecer especialmente a mi madre y a mi hermano Gerardo, por todo el tiempo en que me recibieron y cuidaron cuando necesitaba tranquilidad para trabajar. No me olvido de mi padre, quien está presente en todo momento a pesar de ya no estar.

Por último, en el ámbito laboral, organizacional y financiero, le agradezco al Instituto de Computación de la Facultad de Ingeniería por otorgarme un semestre libre de actividades de enseñanza para avanzar en esta tesis. También agradezco a la Comisión Sectorial de Investigación Científica (CSIC), Universidad de la República, Uruguay y al proyecto “Computación de alto desempeño aplicada a discretizaciones con volúmenes finitos del Río de la Plata” (financiado por CSIC), por apoyar financieramente la publicación de los artículos que surgieron durante el proceso de investigación.

Capítulo 1

Introducción

En los últimos treinta años, la computación gráfica ha tenido un importante desarrollo potenciado por tres fuerzas principales: la facilidad intrínseca de las personas de comprender y manipular elementos visuales; el incremento a tasas exponenciales de la potencia gráfica de las computadoras; y el desarrollo de metodologías y algoritmos gráficos eficientes. Entre las metodologías y algoritmos gráficos desarrollados, poseen relevancia aquellos que buscan resolver o simular la iluminación global de una escena, es decir, generar imágenes que contemplen la reflexión de la luz en los objetos de una escena, el transporte de la luz en medios participativos, etc. Los métodos de iluminación global a su vez pueden dividirse entre aquellos físicamente realistas, que están asentados sobre principios físico-matemáticos extraídos de la realidad, y aquellos que logran resultados efectistas que a “simple vista” parecen realistas, pero que adolecen de defectos importantes.

El **método de radiosidad** es una de las técnicas pioneras de iluminación global físicamente realistas. El método surgió como una variante de las técnicas empleadas para el estudio de la transferencia de calor a través de la radiación térmica. En el método de radiosidad se propone una formulación general de la iluminación global, donde las superficies poseen reflexión difusa lambertiana y se excluyen aquellas superficies que tienen componentes de reflexión especular. Esta restricción no ha invalidado su uso en diversas áreas como el diseño arquitectónico, el desarrollo de películas animadas y el diseño de escenarios tridimensionales para videojuegos.

Formulado como un problema matemático, el **problema de radiosidad** para una escena implica hallar la función solución de una ecuación integral de Fredholm de segunda especie. Una ecuación de Fredholm puede ser transformada en una ecuación con operadores lineales, que en su versión discreta se transforma en un sistema de ecuaciones lineales. Con estas transformaciones, el cálculo de la radiosidad de una escena pasa a ser el cálculo de la incógnita de un sistema lineal.

Transformar el problema de radiosidad continuo en un sistema lineal tiene características particulares, surgiendo inconvenientes propios de la técnica. La transformación de la escena continua en otra compuesta por polígonos (parches) con características especiales debe contemplar la minimización del error inherente al proceso de discretización, obligando a utilizar entre miles y millones de parches en escenas sencillas. Por otra parte, las matrices que se construyen a partir de ecuaciones integrales suelen contener del orden de n^2 números de punto flotante no nulos, siendo n el número de parches utilizados para discretizar la escena, lo que obliga a un manejo cuidadoso de la memoria principal y

otros recursos computacionales. Otro inconveniente importante resulta de la resolución de sistemas lineales de grandes dimensiones. El empleo de métodos iterativos para la resolución de estos sistemas es un reto computacional en sí mismo. Todas estas dificultades se agudizan desde un punto de vista computacional cuando se pretende resolver el problema de radiosidad en menos de $\frac{1}{20}$ de segundo para incluir la radiosidad en aplicaciones gráficas interactivas que funcionan en tiempo real. Por ejemplo, en el ámbito de la arquitectura, los iluminadores, escenógrafos, decoradores y arquitectos se beneficiarían con aplicaciones gráficas interactivas que les permitan experimentar con cientos de configuraciones de luz para una arquitectura dada, sin tener que esperar algunos segundos para el cálculo de cada radiosidad. Por último, en la búsqueda del realismo visual la satisfacción del usuario nunca se alcanza por completo. La imagen realista de hoy será mañana una caricatura de la realidad, por lo tanto, siempre se requerirán más parches y menor tiempo de cálculo.

En este contexto, las técnicas de radiosidad de rango bajo proponen un enfoque novedoso para el tratamiento de los inconvenientes planteados. Partiendo de una geometría cualquiera, generan un sistema lineal que ocupa relativamente poca memoria y resuelven el problema de radiosidad en tiempo real en escenarios con cientos de miles de parches. Algunos trabajos que han incluido consideraciones sobre el rango bajo de las matrices involucradas son los de Ashdown [2] y más recientemente los trabajos de Hašan et al. [40] y Hašan [39].

Los trabajos previos no han aplicado métodos sistemáticos para reducir el rango de la matriz de radiosidad, como medio para reducir la memoria necesaria para resolver el problema de radiosidad, mejorar la eficiencia de los métodos existentes y para desarrollar aplicaciones de tiempo real. Por lo tanto, está abierta la posibilidad de contribuir en esta línea de investigación proponiendo métodos que trabajen con información reducida y permiten calcular iluminaciones realistas aún para escenas complejas y de grandes dimensiones.

Esta tesis propone la sustitución del sistema lineal del problema de radiosidad por otro sistema lineal donde la matriz densa es reemplazada por una aproximación de rango bajo. La resolución del sistema lineal resultante posee una complejidad lineal en escenarios con geometría fija, donde sólo varían la configuración de los parches emisores.

Las principales contribuciones del trabajo son:

1. El estudio de las técnicas de radiosidad de rango bajo y un análisis de los trabajos relacionados.
2. El estudio de las técnicas de factorización existentes para el cálculo de aproximaciones de rango bajo a matrices.
3. La propuesta de dos algoritmos para realizar aproximaciones de rango bajo específicas para las matrices surgidas del problema de radiosidad.
4. La propuesta de unas metodologías directas e iterativas para la resolución eficiente del problema de radiosidad, para escenarios estáticos y dinámicos.
5. El análisis experimental de las técnicas propuestas, que demuestra la capacidad de los nuevos métodos para resolver el problema de radiosidad en pocos milisegundos con cientos de miles de parches, lográndose de esta manera la resolución del problema de radiosidad en tiempo real.

El contenido de la tesis se estructura de la siguiente manera. El capítulo 2 presenta una introducción al problema de radiosidad. Comienza con una descripción de los aspectos físico matemáticos de la luz más relacionados con la tesis, y continúa con la exposición de las metodologías más generales para la resolución del problema de radiosidad. A partir de la utilización de elementos finitos se establecen estrategias para la construcción de los elementos del sistema lineal de radiosidad (llamado *ecuación matricial de radiosidad*). Luego de construido el sistema lineal, se desarrollan algunos de los métodos más relevantes para la resolución de la ecuación matricial de radiosidad. El capítulo culmina con la exposición de un relevamiento de aquellos desarrollos más relacionados con esta tesis. El capítulo 3 desarrolla los conceptos principales alcanzados en el curso de este trabajo de investigación. Comienza exponiendo el concepto de rango bajo numérico y desarrollando algunas de las formas más utilizadas para aproximar una matriz de dimensión $n \times n$ por otra de rango bajo. Continúa explicando una propuesta de formulación matricial alternativa para el problema de radiosidad, basada en el uso de matrices de rango bajo, llamada . Esta propuesta surge de la propiedad de coherencia espacial de las escenas, que se manifiesta en el rango numérico bajo de la matriz de radiosidad. Luego se exponen nuevas metodologías para la resolución eficiente del problema de radiosidad, incluyendo métodos iterativos y directos, que son luego ampliados para escenas con superficies coloreadas y escenas con geometría variable. Las últimas secciones del capítulo 3 están dedicadas a explicar los conceptos sobre los que se puede desarrollar una aproximación alternativa de rango bajo a la matriz de radiosidad. Por último se desarrollan dos algoritmos para generar aproximaciones de rango bajo a la matriz de radiosidad. El capítulo 4 está dedicado al análisis experimental, realizándose evaluaciones de los conceptos desarrollados en el capítulo 3. Se estudia la relación entre la emisión y reflexión para una escena genérica, y se realiza un análisis comparativo de las aproximaciones calculadas por los algoritmos y las basadas en descomposiciones tradicionales. Luego se evalúa la eficiencia computacional de uno de los métodos implementados para el cálculo de la ecuación RRB para escenas con geometría fija y donde la luz puede variar su posición. Por último, se realiza un análisis de la variación de la matriz de radiosidad cuando la geometría de la escena es modificada, y se realiza el cálculo de la radiosidad de la escena. En el capítulo 5 se exponen las conclusiones de la investigación así como las principales líneas de trabajo futuro.

Capítulo 2

El problema de radiosidad

2.1. Introducción

La generación de imágenes realistas es el proceso informático de creación de imágenes que son indistinguibles de aquellas surgidas de la realidad [44]. Las primeras técnicas desarrolladas para la generación de imágenes realistas estaban limitadas por la tecnología de su tiempo y no eran más que un conjunto de “trucos” que resolvían algunos temas puntuales. La determinación de las superficies visibles desde el punto de vista del observador, el cálculo de la incidencia directa de la luz en los objetos y los primeros cálculos de sombras fueron grandes avances, que fueron realizados en la década de 1970 y principios de la década de 1980 [24]. Luego, con el advenimiento de las técnicas de traza de rayos (*ray tracing*) y radiosidad (*radiosity*), la síntesis digital de imágenes comenzó a generar simulaciones basadas en la física con un grado de realismo apreciables. Las técnicas de traza de rayos y de radiosidad existieron con anterioridad a su utilización en computación gráfica: la traza de rayos era utilizada en óptica y la radiosidad en los problemas de transferencia de calor [15].

Las técnicas de **iluminación global** son aquellas que calculan la difusión de la luz contemplando las particularidades de las escenas sobre las que actúa. Con estas técnicas se pretende lograr una precisa predicción de la luz en cada punto de una escena. Diversos algoritmos de iluminación global se han desarrollado, la mayoría de ellos basados en dos técnicas principales ya mencionadas: las técnicas de muestreo de puntos a partir de la traza de rayos y las técnicas de elementos finitos para el problema de radiosidad. También existen técnicas híbridas que combinan ambos métodos.

Las técnicas de traza de rayos emiten rayos infinitesimales a través de la escena. Las primeras técnicas exitosas, desarrolladas por Whitted en 1980 [73], emiten rayos en sentido inverso desde el observador a las fuentes de luz. Con estas técnicas, sólo se pueden considerar reflexión y refracción en superficies especulares. Para calcular la iluminación indirecta proveniente de superficies de reflexión difusa, cáusticas y reflexiones brillosas (*glossy*) se amplió la traza de rayos de manera de incluir a los métodos de Monte Carlo, en los que los rayos se distribuyen de manera estocástica para considerar todos los caminos posibles de la luz [45, 49].

Esta tesis está basada en las técnicas de radiosidad, por lo tanto, en este capítulo se desarrollan sus principales conceptos básicos y se profundiza en algunos aspectos específicos. El capítulo comienza con una breve introducción de los aspectos físicos de la luz y de las ecuaciones integrales, que son la base del método de radiosidad. Luego

se describen algunas soluciones a la ecuación de radiosidad, en especial explicándose los detalles de la construcción de un sistema lineal que aproxime a su solución. Por último, se establecen algunas metodologías para la reducción del rango de la matriz de los factores de forma, empleadas en algunos métodos de iluminación global.

2.2. Radiosidad: aspectos físico-matemáticos

Cohen et al. [15] y Glassner [28] han escrito excelentes referencias en los aspectos físico-matemáticos de la radiosidad y de la síntesis digital de imágenes. Esta sección no pretende abarcar la generalidad de esta temática, sino aquellos aspectos necesarios para comprender los conceptos presentados en el resto de la tesis.

2.2.1. Consideraciones básicas de la luz

La luz es un tipo de energía electromagnética. Como tal, sus propiedades pueden ser interpretadas de dos maneras: como ondas y como partículas. Muchas de las propiedades de la luz se pueden comprender al utilizar la metáfora de la luz como una onda que viaja a 3×10^8 m/s. Algunas otras propiedades pueden ser comprendidas si se considera a la luz como un conjunto de partículas (fotones) que se desplazan por el espacio.

La luz se emite desde fuentes luminosas como el Sol, algunas reacciones químicas, objetos calientes, etc. Los objetos, dependiendo de las propiedades del material, pueden absorber, reflejar o refractar la luz que incide en ellos. Cuando la luz es absorbida por un objeto se transforma generalmente en calor.

Cada fotón contiene una cantidad de energía que está relacionada directamente con su frecuencia. A partir de los trabajos de Planck y Einstein [52] se puede deducir que una lámpara de 100 Watts emite como máximo unos $2,5 \times 10^{20}$ fotones por segundo. Este número es muy alto como para realizar una simulación computacional realista de las interacciones entre los fotones y las superficies, y es uno de los motivos del surgimiento de la computación gráfica como el conjunto de técnicas que generan imágenes realistas a partir de recursos computacionales limitados.

Luego de emitido un fotón, puede interactuar con uno o más objetos y posteriormente, puede arribar al ojo del observador. En una escena con un observador, solo una parte muy pequeña de todos los fotones emitidos arriba a sus ojos. De esos fotones a su vez sólo una pequeña porción corresponde al rango de frecuencias comprendido entre 380 y 770 nanómetros, que es el rango en que el ojo es sensible a la luz. A partir de la reacción de los ojos ante la presencia de la luz, comienza un proceso muy complejo en el que interviene el sistema nervioso, que interpreta las señales recibidas y las transforma en colores, formas, texturas, sensación de profundidad, etc.

No toda la energía electromagnética es percibida por la vista, y señales muy diferentes pueden ser percibidas como iguales por el ojo humano. A continuación se brinda el conjunto de unidades físicas más representativas de la luz. Estas unidades contemplan la luz como fenómeno físico y la luz como fenómeno sensorial humano. Las unidades a exponer son radiométricas cuando están basadas en la radiometría (la ciencia que estudia la medición física de la energía electromagnética) y fotométricas cuando están basadas en la fotometría (la ciencia que estudia la medición psicofísica de la sensación visual humana producida por el espectro electromagnético) [15].

2.2.2. Unidades físicas

Existe un conjunto de unidades radiométricas y fotométricas que ayudan a medir distintos aspectos de la luz [15, 44].

Las unidades radiométricas de energía radiante y potencia radiante (o flujo radiante) son *joules* [$J = \text{kgm}^2/\text{s}^2$] y *watts* [$W = J/\text{s}$] respectivamente, en cambio la unidad fotométrica de potencia luminosa es el *lumen* y la unidad fotométrica de energía luminosa es el *talbot*. Una de las diferencias más notorias entre las unidades radiométricas y las fotométricas está en que el ojo es más sensible a la luz de 550nm (color verde), que a otras frecuencias, siendo casi insensible cerca de los extremos del espectro visible. Para la misma cantidad de watts hay más lumens a 550nm que en otras frecuencias.

La *radiancia* [$W/\text{m}^2\text{sr}$] mide la cantidad de luz que pasa a través de o es emitida por un área particular y se emite en direcciones contenidas dentro en un ángulo sólido dado. Se mide en watts por estereorradianes por metro cuadrado. La radiancia se utiliza para indicar cuánta de la potencia emitida por una superficie va a ser recibida por un sistema óptico que mira desde determinada posición. La radiancia y su equivalente fotométrico la luminancia son buenos indicadores de cuán brillante el objeto parecerá al observador.

La *irradiancia* [W/m^2] es la potencia radiante por unidad de área que incide sobre una superficie. El equivalente fotométrico es la iluminancia que se mide en *lux* [$\text{lumens}/\text{m}^2\text{sr}$].

La **radiosidad** [W/m^2] es muy similar a la irradiancia. La diferencia es que radiosidad es la potencia por unidad de área que deja una superficie. La radiosidad también es llamada *exitancia radiante*. El equivalente fotométrico es la *luminosidad*, que también se mide en *lux*.

La *intensidad radiante* [W/sr] es una medida de intensidad en la radiación electromagnética. Se define como potencia por unidad de ángulo sólido. El equivalente fotométrico es *intensidad luminosa* y se mide en *candelas=lumens/sr*.

2.2.3. La interacción de la luz con los objetos

Cuando la luz interacciona con un objeto, es absorbida por éste o se difunde [44]. Cuando se difunde dentro del objeto, recorre parte de su interior antes de salir nuevamente a la superficie y continuar su recorrido (véase la Figura 2.1 (a)). Esto ocurre principalmente en objetos translúcidos como el mármol o la piel. Se puede realizar una simulación de este fenómeno a través de la función *Bidirectional Scattering Surface Reflectance Distribution Function* o (BSSRDF) cuya expresión se presenta en la Ecuación (2.1), que relaciona el diferencial de radiancia reflejada en el punto x en dirección \vec{w} ($dL_r(\vec{x}, \vec{w})$), con el diferencial de flujo incidente en el punto \vec{x}' y desde la dirección \vec{w}' , ($d\Phi_i(\vec{x}', \vec{w}')$). Las dos direcciones y las dos posiciones son datos de entrada, y la salida es el cociente entre los diferenciales de radiancia reflejada y de flujo incidente.

$$BSSRDF(x, \vec{w}, \vec{x}', \vec{w}') = \frac{dL_r(\vec{x}, \vec{w})}{d\Phi_i(\vec{x}', \vec{w}')} \quad (2.1)$$

La función BSSRDF es costosa de evaluar por lo que suele ser sustituida por una función más simple, comúnmente utilizada en computación gráfica, llamada *Bidirectional Reflectance Distribution Function* o (BRDF) cuya expresión se presenta en la Ecuación (2.2).

$$BRDF(x, \vec{w}, \vec{w}') = \frac{dL_r(\vec{x}, \vec{w})}{dE_i(\vec{x}, \vec{w}')} \quad (2.2)$$

La función BRDF asume que la luz incidente y la reflejada parten del mismo punto, y establece cuál es la relación existente entre la radiancia reflejada ($dL_r(x, \vec{w})$) y la potencia recibida por unidad de área o irradiancia ($dE_i(x, \vec{w}')$) (véase la Figura 2.1 (b)).

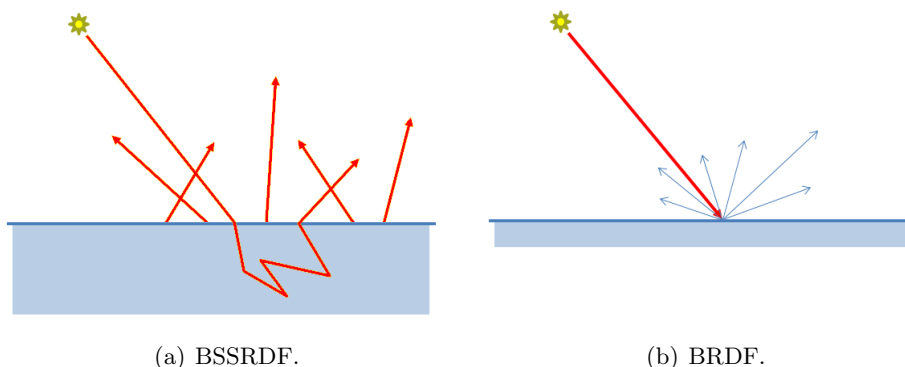


Figura 2.1: Modelos de interacción entre la luz y las superficies. La piel y el mármol están mejor representados en (a) y los metales y materiales dieléctricos en (b).

Las funciones BSSRDF y BRDF brindan grandes libertades en el modelado de la interacción de la luz con los objetos. De todas las interacciones posibles entre la luz y los objetos, la **reflexión difusa** y la **reflexión especular** agrupan a la mayor parte de las interacciones existentes. La reflexión difusa ocurre cuando la luz se refleja en todas direcciones una vez que un rayo incide en una superficie, y se da principalmente en superficies rugosas o en superficies donde la luz se difunde por debajo de la superficie (véase la Figura 2.3 (a)). Un caso especial de reflexión difusa es la **reflexión lambertiana**, donde la reflexión es igual en todas direcciones (véase la Figura 2.3 (b)). En la reflexión lambertiana la radiancia es constante en todas direcciones y su valor de la radiancia es proporcional a la irradiancia. El ángulo de incidencia de la irradiancia influye debido a que cada rayo de luz incide en un área que varía de tamaño según el ángulo de incidencia. Un haz de luz con sección igual a un diferencial de área dA interseca una superficie con área $dA/\cos\theta$ (véase la Figura 2.2). Para un haz luminoso incidente, la potencia luminosa que incide por unidad de área es proporcional a $\cos\theta$.

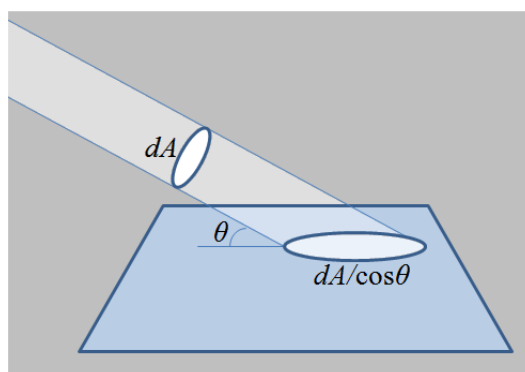


Figura 2.2: Haz de luz con área de sección dA y área de intersección $dA/\cos\theta$.

La reflexión especular ocurre cuando la luz incide sobre una superficie suave metálica o dieléctrica (como el agua o el vidrio). Debido a las imperfecciones de la superficie, un material especular refleja la luz en un pequeño cono alrededor de la dirección especular ideal, brindando una reflexión brillante o *glossy* (véase la Figura 2.3 (c)). Si la superficie es perfectamente lisa, entonces la reflexión es perfectamente especular: el rayo incidente se refleja en la dirección simétrica respecto a la normal de la superficie, como un espejo (véase la Figura 2.3 (d)).

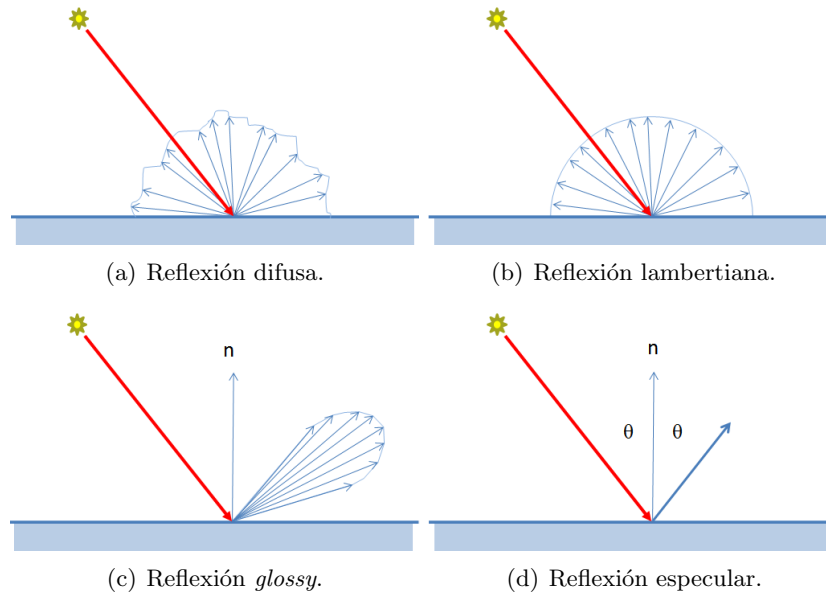


Figura 2.3: Distintos tipos de reflexión. Los dos tipos superiores corresponden a superficies rugosas o con difusión debajo de la superficie. Los dos tipos inferiores corresponden a superficies metálicas o dieléctricas, como el agua y el vidrio.

En el problema de radiosidad se asume que todas las superficies de la escena tienen exclusivamente reflexión lambertiana. Los otros tipos de reflexión son ignorados por las técnicas de radiosidad, eliminando de esta manera la necesidad de incluir el punto de vista del observador en la resolución de la iluminación de la escena. Por lo tanto, el uso de la reflexión lambertiana simplifica la formulación del problema de radiosidad, como observa en la siguiente sección. Si hay superficies con otros tipos de reflexión en una escena, éstas son tratadas por otras técnicas y luego combinadas para generar la imagen final.

La cantidad de luz reflejada por una superficie con reflexión lambertiana depende exclusivamente del ángulo de incidencia de la luz y del coeficiente de reflexión difusa de la superficie (en el caso de superficies coloreadas el coeficiente varía según la frecuencia de la luz). El coeficiente de reflexión es un valor entre cero y uno, donde cero significa que el objeto es un cuerpo negro que no refleja luz y uno significa que es un objeto blanco que refleja absolutamente toda la luz que recibe.

2.2.4. Las ecuaciones integrales de la luz

Una **ecuación integral** es aquella en la que existe una función incógnita dentro de una integral. En la práctica, unas pocas estructuras generales de ecuaciones integrales capturan la mayoría de los modelos matemáticos de la naturaleza, incluyendo la ecuación de radiosidad.

Según Glassner [28], la Ecuación (2.3) tiene la estructura general más útil para la síntesis de imágenes.

$$y(t) = g(t) + \lambda \int_a^b k(t, u)y(u)du \quad (2.3)$$

En la Ecuación (2.3), la incógnita es $y(t)$, una función real de la variable independiente y real t . La función $g(t)$ es llamada *término libre*. El valor λ es en general un número complejo. La función $k(t, u)$ es una función real de dos números reales y es llamada el *núcleo de la integración*.

Existe una taxonomía de ecuaciones integrales. Por ejemplo, si $g(t) = 0$ en el dominio, entonces la ecuación es *homogénea*. Si la integral es lineal en $y(t)$, entonces la ecuación es *lineal*. La ecuación integral es de **Fredholm** si la cota superior del dominio es algún número real b , y es de *Volterra* si la cota superior del dominio es el punto de evaluación t . La ecuación integral es de *primera especie* si la función incógnita aparece sólo dentro de la integral y es de *segunda especie* si la función incógnita aparece dentro y fuera de la integral. En iluminación global, las ecuaciones más utilizadas son las ecuaciones integrales de Fredholm de segunda especie, lineales y no homogéneas.

La **ecuación de rendering** es considerada la ecuación de iluminación global más general. Es una ecuación de Fredholm de segunda especie, publicada por primera vez por Kajiyá en 1986 [46] y que tiene la expresión presentada en la Ecuación (2.4).

$$I(x, x') = g(x, x') \left[\epsilon(x, x') + \int_S \rho(x, x', x'')I(x', x'')dx'' \right] \quad (2.4)$$

En la Ecuación (2.4), el término $I(x, x')$ está relacionado con la intensidad de la luz que va de manera directa del punto x' al punto x ; $g(x, x')$ es un término geométrico (si x y x' se ven vale $1/r^2$ donde r es la distancia entre x y x' , y vale 0 en caso contrario); $\epsilon(x, x')$ está relacionado con la intensidad de la luz emitida por x' en dirección hacia x ; $\rho(x, x', x'')$ es una expresión de la función BRDF, donde se relaciona la intensidad de luz proveniente de x'' que incide en x' y se refleja en dirección a x . Las reflexiones en las superficies consideradas en la Ecuación (2.4) no están limitadas, dado que no hay restricciones respecto al tipo de función que debe ser $\rho(x, x', x'')$.

La ecuación de radiosidad es un caso particular de la ecuación de rendering, desarrollado para el caso en que las superficies de la escena tienen exclusivamente reflexión lambertiana [15]. Al ser la reflexión lambertiana de igual intensidad en todas direcciones, es posible simplificar la ecuación de rendering. No tiene sentido el cálculo específico de la intensidad de luz que va desde el punto x' en dirección hacia cada x de la escena (como en la ecuación de rendering) debido a que su intensidad es igual en todas direcciones. Alcanza con saber cuál es la radiosidad emitida/reflejada por cada punto.

La **ecuación de radiosidad** tiene la expresión general presentada en la Ecuación (2.5), que es una ecuación de Fredholm de segunda especie, donde $B(x)$ es la radiosidad del punto x , $E(x)$ es la emisión del punto x y $\rho(x)$ es la reflectividad difusa en x .

$$B(x) = E(x) + \rho(x) \int_S B(x')G(x, x')dA' \quad (2.5)$$

El término $G(x, x')$ expresa la visibilidad y la distancia entre x y x' , de acuerdo a la expresión de la Ecuación (2.6) donde la función $V(x, x')$ vale 1 si los puntos x y x' son mutuamente visibles y vale 0 si no lo son, y el resto de la expresión establece la relación existente de la intensidad radiada entre los puntos, que está en función de los ángulos y de la distancia existente entre ellos (véase la Figura 2.4).

$$G(x, x') = G(x', x) = V(x, x') \frac{\cos \theta' \cos \theta}{\|x - x'\|^2} \quad (2.6)$$

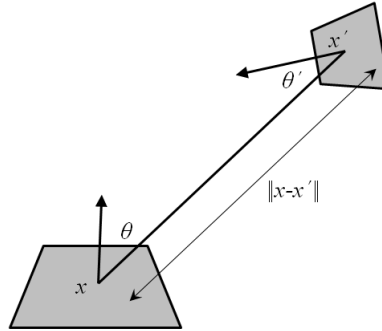
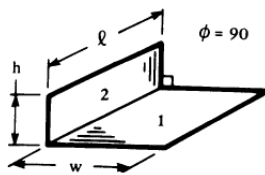


Figura 2.4: Elementos considerados para el cálculo de la función $G(x, x')$.

Las ecuaciones de rendering y de radiosidad expresan la conservación de la energía luminosa en todos los puntos de una escena: la energía reflejada, absorbida y transmitida por cada superficie y volumen de la escena debe ser igual a la energía recibida.

2.3. Resolución de la ecuación de radiosidad

La ecuación de radiosidad es difícil de calcular analíticamente para la mayoría de las escenas. En las publicaciones sobre transferencia de calor existen tablas con fórmulas complejas que expresan la transferencia de energía entre cientos de configuraciones posibles de pares de elementos. Como referencias cabe mencionar a los trabajos de Siegel et al. [64], Howell [42], y Lienhard y Lienhard [52]. En la Figura 2.5 se presenta el diagrama y la fórmula analítica de la interacción 3D entre objetos perpendiculares.



$H = h/\ell$ y $W = w/\ell$. entonces

$$F_{1-2} = \frac{1}{\pi W} \left\{ W \tan^{-1} \frac{1}{W} - \sqrt{H^2 + W^2} \tan^{-1} (H^2 + W^2)^{-1/2} \right. \\ \left. + H \tan^{-1} \frac{1}{H} + \frac{1}{4} \ln \left[\frac{(1 + W^2)(1 + H^2)}{1 + W^2 + H^2} \right] \right. \\ \left. \times \left[\frac{W^2(1 + W^2 + H^2)}{(1 + W^2)(W^2 + H^2)} \right]^{W^2} \left[\frac{H^2(1 + H^2 + W^2)}{(1 + H^2)(H^2 + W^2)} \right]^{H^2} \right\}$$

Figura 2.5: Ejemplo de fórmula analítica exacta para la transferencia de radiosidad entre dos objetos (figura extraída de Lienhard y Lienhard [52]).

La utilización de las metodologías analíticas en escenas con muchos elementos resulta poco práctica. En la actualidad, los métodos analíticos han sido superados por métodos aproximados, que resuelven problemas de integración numérica utilizando la capacidad de cálculo de las computadoras.

Para la resolución de las ecuaciones de Fredholm de segunda especie por métodos numéricos, se han desarrollado tres líneas metodológicas principales: la integración basada en reglas de cuadratura, la integración basada en elementos finitos y la integración basada en los métodos de Monte Carlo. En los métodos de integración basados en reglas de cuadratura y en elementos finitos, se lleva la ecuación integral a un sistema lineal que luego hay que resolver. En la integración basada en los métodos de Monte Carlo existen heurísticas que eliminan la necesidad de conocer los términos del sistema lineal para hallar la radiosidad de la escena.

A continuación se desarrollan los principios básicos de los métodos de integración basados en reglas de cuadratura, en elementos finitos y en Monte Carlo. Luego se presentan las principales técnicas para la resolución del sistema lineal resultante.

2.3.1. Integración basada en reglas de cuadratura

La integración de la Ecuación (2.3) utilizando reglas de cuadratura o reglas de integración numérica [60] transforma la ecuación integral en un sistema lineal. Al resolver el sistema lineal se halla una aproximación a la función incógnita (en la ecuación de radiosidad, la función $B(x)$). Resulta entonces la Ecuación (2.7), donde los términos w_j son los pesos de la regla de cuadratura utilizada para el cálculo de la integral.

$$y(t) = g(t) + \lambda \sum_{j=0}^{N-1} w_j k(t, u_j) y(u_j) \quad (2.7)$$

Luego de realizada esta simplificación, se calcula la ecuación para N puntos en t (Ecuación (2.8)), obteniéndose un sistema de N ecuaciones presentado en la Ecuación (2.9), donde Y contiene las incógnitas $y(t_i)$, \mathbf{I} es la matriz identidad, \mathbf{K} contiene los términos $w_j k(t_i, u_j)$ y G se compone de los términos $g(t_i)$.

$$y(t_i) = g(t_i) + \lambda \sum_{j=0}^{N-1} w_j k(t_i, u_j) y(u_j), i \in 0 \dots N - 1 \quad (2.8)$$

$$(\mathbf{I} - \lambda \mathbf{K})Y = G \quad (2.9)$$

Este sistema de ecuaciones está generalmente bien condicionado, salvo cuando λ es muy cercano a un valor propio de \mathbf{K} . En el caso de la ecuación de radiosidad, λ es 1 y el máximo valor propio de \mathbf{K} es menor que 1 debido a la absorción de energía de las superficies (cuanto más reflectivas sean las superficies, será más cercano a 1) [15].

Björck [9] sugiere utilizar como regla de cuadratura la regla de Gauss-Legendre, debido a que posibilita una mayor precisión, empleando menos puntos que otras cuadraturas. Como la resolución del sistema lineal involucra $O(n^3)$ operaciones, donde n es la cantidad de puntos, cualquier reducción de puntos afecta notoriamente los tiempos de resolución del problema.

Si el núcleo \mathbf{K} no posee singularidades y varía suavemente en su recorrido, se facilita la obtención de resultados de buena precisión con un número reducido de puntos utilizando reglas de cuadratura. Si, por ejemplo, \mathbf{K} poseyera algunas discontinuidades, éstas deberían reflejarse en los puntos elegidos para su discretización, dado que deberían concentrarse los puntos cercanos a la discontinuidad para realizar un buen relevamiento de sus propiedades. En la ecuación de radiosidad, la función \mathbf{K} puede variar bruscamente y poseer singularidades, debido a la existencia de esquinas, aristas y uniones en la mayoría de las superficies de una escena. Por lo tanto, las reglas de cuadratura no son una buena metodología para aplicar en el problema de radiosidad.

Una vez calculados los valores $y(t_i)$ en los puntos t_i , puede ser de interés calcular y para cualquier t . En lugar de utilizar cualquier función interpolante tradicional (como Lagrange, Hermite, splines, etc. [18]), se sugiere utilizar el método de Nyström [60]. El método de Nyström consiste en utilizar la Ecuación (2.7) como fórmula interpolatoria. A partir de la resolución del sistema lineal, se conocen los valores de $y(u_j)$ utilizados en la Ecuación (2.7) y por lo tanto se conoce por completo la expresión a la derecha de la igualdad en (2.7), con la que se puede calcular y para todo t .

En la ecuación de radiosidad no resulta tan sencillo utilizar Gauss-Legendre ni el método de Nyström debido a que en lugar de una variable t unidimensional se trabaja con una variable x de dos dimensiones, si se consideran sólo las superficies de la escena y de tres dimensiones si hay medios participativos -como el humo y la niebla, que ocupan una porción del espacio y sus puntos interiores reflejan y transmiten luz-. También hay problemas respecto a la suavidad del recorrido de \mathbf{K} , en especial en las aristas y uniones de las superficies. Por lo tanto, para resolver la ecuación de radiosidad se sugiere aplicar otros métodos, por ejemplo aquellos basados en elementos finitos.

2.3.2. Integración basada en elementos finitos

Los **métodos basados en elementos finitos** (MEF) son útiles en general para el tratamiento de problemas físicos basados en ecuaciones diferenciales. Han mostrado ser útiles cuando el dominio del problema tiene una estructura compleja o cuando las funciones involucradas se comportan de maneras distintas en las diferentes partes del dominio. En MEF, el dominio se representa como la unión de un número finito de subdominios de estructura simple (por ejemplo, triángulos). En computación gráfica a cada subdominio se los denomina **parche**. En cada subdominio las funciones involucradas se aproximan por funciones continuas, solamente definidas en ciertos puntos a partir del valor de la función y eventualmente de alguna de sus derivadas. Estos puntos se llaman nodos y pertenecen al interior o a la frontera del subdominio [74].

Al aplicar el método de elementos finitos al problema de radiosidad, una escena se transforma en otra aproximada compuesta por parches, y la función B de la ecuación de radiosidad se aproxima por otra función $\tilde{B}(x)$ continua dentro de cada parche pero discontinua entre parches [15]. Por lo tanto, se puede formular $\tilde{B}(x)$ mediante la expresión de la Ecuación (2.10), donde n indica el total de parches e i es el índice de cada parche, donde B_i es el pesos por el que está multiplicado N_i , y donde

$$N_i(x) = \begin{cases} p(x), & \text{polinomio de bajo orden, cuando } x \text{ esta dentro del parche } i \\ 0 & \text{cuando } x \text{ está fuera del parche } i \end{cases}$$

$$B(x) \approx \tilde{B}(x) = \sum_i^n B_i N_i(x) \quad (2.10)$$

Al aplicar estas transformaciones a una escena y a la ecuación de radiosidad, resulta difícil que se cumpla la igualdad $B(x) = \tilde{B}(x)$ en todos los puntos de la escena. En la Ecuación (2.10) $\tilde{B}(x)$ es una combinación de los polinomios $N_i(x)$, aspecto que en general no se cumple para una función $B(x)$ que cumple con la ecuación de radiosidad. Para poder salvar este inconveniente, se define la **ecuación de residuo** (2.11) y, en lugar de exigir $r(x) = 0 \forall x$, se busca una función $\tilde{B}(x)$ que minimice su valor según cierta métrica.

$$r(x) = \tilde{B}(x) - E(x) - \rho(x) \int_S \tilde{B}(x') G(x, x') dA' \quad (2.11)$$

Por ejemplo, una posibilidad es tratar de encontrar un $\tilde{B}(x)$ tal que cumpla $r(x_i) = 0 \forall i$, siendo x_i los puntos nodos de la escena. Otra posibilidad es hallar un $\tilde{B}(x)$ tal que el $r(x)$ resultante sea ortogonal a las funciones $N_i(x)$. La primera de las posibilidades se llama *método de puntos de colocación* (*point collocation method*) y la segunda *método de Galerkin*. Ambas técnicas se presentan a continuación.

Integración basada en el método de puntos de colocación

En el método de puntos de colocación, se cumple que

$$r(x_i) = \tilde{B}(x_i) - E(x_i) - \rho(x_i) \int_S \tilde{B}(x') G(x_i, x') dA' = 0, \forall i \quad (2.12)$$

Sustituyendo \tilde{B} por su formulación en la Ecuación (2.10), resulta

$$r(x_i) = \sum_j^n B_j N_j(x_i) - E(x_i) - \rho(x_i) \int_S \sum_j^n B_j N_j(x') G(x_i, x') dA' = 0, \forall i \quad (2.13)$$

Luego, agrupando por B_i , que es una constante (independiente de x y x'), se obtiene

$$r(x_i) = \left[\sum_j^n B_j \left[N_j(x_i) - \rho(x_i) \int_S N_j(x') G(x_i, x') dA' \right] \right] - E(x_i) = 0, \forall i \quad (2.14)$$

Esta ecuación puede ser replanteada simplemente como un conjunto de n ecuaciones

$$r(x_i) = \sum_j^n B_j \mathbf{K}_{ij} - E(x_i) = 0, \forall i \quad (2.15)$$

donde

$$\mathbf{K}_{ij} = N_j(x_i) - \rho(x_i) \int_S N_j(x') G(x_i, x') dA', \forall i, j \quad (2.16)$$

o matricialmente como

$$\mathbf{K}B = E, \text{ con } \mathbf{K} = \{\mathbf{K}_{ij}\}, B = \{B_j\}, E = \{E(x_i)\} \quad (2.17)$$

Integración basada en el método de Galerkin

Como se expresó anteriormente, en el método de Galerkin los residuos deben ser ortogonales a las funciones $N_i(x)$. Galerkin es un caso particular del método de los *weighted residuals*, donde el residuo debe ser ortogonal a ciertas funciones W_i predefinidas. La ortogonalidad entre las n funciones N_i y r se define como

$$\langle N_i(x), r(x) \rangle = \int_S N_i(x)r(x)dA = 0, \forall i \quad (2.18)$$

Utilizando la expresión (2.13) para expandir $r(x)$ resulta

$$\int_S N_i(x) \left[\sum_{j=1}^n B_j N_j(x) - E(x) - \rho(x) \int_S \sum_{j=1}^n B_j N_j(x') G(x, x') dA' \right] dA = 0, \forall i \quad (2.19)$$

Finalmente, agrupando por B_j se obtiene

$$\left[\sum_{j=1}^n B_j \left[\int_S N_i(x) N_j(x) dA - \int_S N_i(x) \rho(x) \int_S N_j(x') G(x, x') dA' dA \right] \right] - \int_S N_i(x) E(x) dA = 0, \forall i \quad (2.20)$$

La Ecuación (2.20), al igual que la surgida del método de los puntos de colocación, también puede aproximarse a un sistema lineal

$$\mathbf{K}B = E \quad (2.21)$$

donde

$$\mathbf{K}_{ij} = \int_S N_i(x) N_j(x) dA - \int_S N_i(x) \rho(x) \int_S N_j(x') G(x, x') dA' dA, \forall i, \forall j \quad (2.22)$$

y E_i es el promedio de emisión por unidad de área en el parche i .

$$E_i = \int_S N_i(x) E(x) dA \quad (2.23)$$

Integración basada en Galerkin con N_i constante

Si las funciones $N_i(x)$ expresadas en (2.3.2), se simplifican de manera tal que

$$N_i(x) = \begin{cases} 1 & \text{cuando } x \text{ está dentro del parche } i \\ 0 & \text{cuando } x \text{ está fuera del parche } i \end{cases} \quad (2.24)$$

la Ecuación (2.20) se simplifica considerablemente [15]. En primer lugar, se cumple la Ecuación (2.25) siendo A_i el área del parche i .

$$\int_S N_i(x) N_j(x) dA = A_i \delta_{ij}, \forall i, j \quad (2.25)$$

En segundo lugar, la Ecuación (2.23) se transforma en

$$\int_S N_i(x)E(x)dA = E_i A_i \quad (2.26)$$

donde E_i es el promedio de emisión por unidad de área en el parche i . En tercer lugar, si se asume que la reflectividad $\rho(x)$ es igual a la constante ρ_i dentro de cada área i , se tiene que

$$\int_S N_i(x)\rho(x) \int_S N_j(x')G(x, x')dA'dA = \rho_i \int_{A_i} \int_{A_j} G(x, x')dA'dA, \forall i, j \quad (2.27)$$

Con estas simplificaciones, la Ecuación (2.20) se transforma en

$$\left[\sum_{j=0}^n B_j \left[\delta_{ij} A_i - \rho_i \int_{A_i} \int_{A_j} G(x, x')dA'dA \right] \right] - E_i A_i = 0, \forall i \quad (2.28)$$

que a su vez se puede expresar como

$$\left[\sum_{j=0}^n B_j [\delta_{ij} - \rho_i \mathbf{F}_{ij}] \right] = E_i, \forall i \quad (2.29)$$

donde \mathbf{F}_{ij} es llamado **factor de forma**, que se calcula mediante la fórmula en la Ecuación (2.30).

$$\mathbf{F}_{ij} = \frac{1}{A_i} \int_{A_i} \int_{A_j} G(x, x')dA'dA \quad (2.30)$$

Un factor de forma \mathbf{F}_{ij} indica qué fracción del total de la energía luminosa que abandona el parche i llega al parche j . Los factores de forma tienen siempre valores no negativos (tienen un valor positivo o valen cero si los parches no se ven entre sí). Los factores de forma fueron calculados analíticamente previamente al advenimiento de las computadoras, para posibilitar los cálculos de radiación de energía calórica en problemas de termodinámica. Como ejemplo, las fórmulas presentadas en la Figura 2.5, refieren a la expresión analítica de los factores de forma entre dos placas perpendiculares entre sí, en 3D.

La Ecuación (2.29) se expresa también como

$$B_i = E_i + \rho_i \sum_{j=0}^n \mathbf{F}_{ij} B_j, \forall i \quad (2.31)$$

o como un sistema lineal de ecuaciones, que fuera presentado por primera vez por Goral et al. [30]

$$\begin{pmatrix} 1 - \rho_1 \mathbf{F}_{11} & -\rho_1 \mathbf{F}_{12} & \cdots & -\rho_1 \mathbf{F}_{1n} \\ -\rho_2 \mathbf{F}_{21} & 1 - \rho_2 \mathbf{F}_{22} & \cdots & -\rho_2 \mathbf{F}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -\rho_n \mathbf{F}_{n1} & -\rho_n \mathbf{F}_{n2} & \cdots & 1 - \rho_n \mathbf{F}_{nn} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{pmatrix} = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{pmatrix} \quad (2.32)$$

La Ecuación (2.32) también puede ser expresada a través de la Ecuación (2.33), donde \mathbf{I} es la matriz identidad de dimensión $n \times n$, \mathbf{R} es una matriz diagonal con los índices de reflectividad ρ_i , \mathbf{F} es una matriz con los factores de forma \mathbf{F}_{ij} , B es un vector con los valores de radiosidad de cada parche B_i y E es un vector donde E_i es la emisión promedio por unidad de área del parche i . En este trabajo, la Ecuación (2.33) es llamada **ecuación matricial de radiosidad** con el fin de distinguirla de la ecuación de radiosidad (2.5).

$$(\mathbf{I} - \mathbf{R}\mathbf{F})B = E \quad (2.33)$$

2.3.3. Integración basada en Monte Carlo

Los **métodos de Monte Carlo** [55, 47] comprenden una clase de algoritmos computacionales que son utilizados para la resolución de problemas genéricos, interpretándolos como problemas probabilísticos, considerando como incógnita el valor esperado de un conjunto de variables aleatorias.

El tratamiento de un problema como probabilístico puede realizarse independientemente de si el problema original es probabilístico. Por ejemplo, la integración basada en los métodos de Monte Carlo evalúa la función a integrar en una muestra aleatoria de puntos y estima la integral en base a esas evaluaciones [60]. La integración basada en los métodos de Monte Carlo es útil cuando las funciones a integrar tienen muchas dimensiones, cuando no hay picos en regiones pequeñas, cuando la estructura de la función es compleja o cuando no se requiere precisión extrema en los resultados. Los métodos de Monte Carlo no son muy eficientes en relación a su precisión, debido a que la precisión en el valor calculado de la incógnita es inversamente proporcional a la raíz cuadrada del número de puntos de la muestra.

La utilidad de los métodos de Monte Carlo en las ecuaciones de Fredholm y de radiosidad surge de que éstas se pueden representar como integrales múltiples de infinitas dimensiones. En efecto, en el caso de la ecuación de radiosidad se tiene la expresión en la Ecuación (2.5) donde, expandiendo de manera iterativa el término B se tiene que

$$\begin{aligned} B(x) &= E(x) + \rho(x) \int_S \left[E(x') + \rho(x') \int_S B(x'') G(x', x'') dA'' \right] G(x, x') dA' \\ &\vdots \\ B(x) &= E(x) + \rho(x) \int_S \left[E(x') + \rho(x') \int_S [E(x'') + \dots] G(x', x'') dA'' \right] G(x, x') dA' \end{aligned} \quad (2.34)$$

Cada paso de la iteración representa una emisión o una reflexión de los rayos de luz. La Ecuación (2.34) expresa la radiosidad de x , generada por la luz emitida por x o por la luz reflejada por x proveniente de otras partes.

Una propiedad interesante de la luz, es que aquella luz que ha realizado muchas reflexiones suele influir poco en la radiosidad final de la escena. Si las superficies no son extremadamente reflectoras, en unas pocas reflexiones la luz pierde significativamente su intensidad, influyendo poco en la radiosidad final. Esta propiedad posibilita limitar el número de integrales de la Ecuación (2.34). Por ejemplo, si se establece como máximo dos reflexiones, entonces los términos en la expansión se sustituyen por cero en la Ecuación (2.34). La luz emitida en algún punto x'' sólo se reflejará en x si previamente se reflejó en algún x' visible desde x .

La Ecuación (2.34) es una integral múltiple, y por lo tanto es una de las ecuaciones para las que los métodos de Monte Carlo son relativamente eficientes. Entre los métodos de Monte Carlo más exitosos se encuentran los de *caminata aleatoria* (*random walk*) [63], que explicitan el camino recorrido por los rayos de luz entre las fuentes emisoras y las superficies de la escena. En los métodos de caminata aleatoria, se simulan los caminos que recorre una partícula de luz de acuerdo a las leyes físicas de emisión y difusión. Se lanzan rayos desde las superficies emisoras en todas direcciones y cada rayo cambia de dirección cada vez que interseca una superficie o mientras se desplaza por un medio participativo, siguiendo un comportamiento aleatorio basado en la física. La radiosidad de cada parche está relacionada fuertemente con la suma de la potencia del total de los rayos con los que interaccionó.

Como se mencionara en la sección 2.2.1, una lámpara de 100 Watts de potencia emite del orden de 10^{20} fotones por segundo. Para evitar lanzar 10^{20} rayos para obtener imágenes realistas, se deben tener en cuenta los siguientes aspectos [68]:

- *Muestras densas*: Tomar muestras que cubran densamente todo el espacio de caminos de luz, para no ignorar posibilidades importantes de transferencia de luz.
- *Muestreo de importancia* (*importance sampling*): Concentrar el cómputo en aquellos caminos de luz que transporten potencia luminosa significativa.
- *Reutilización*: Aprovechar la información obtenida al recorrer otros caminos de luz, en lugar de comenzar cada vez la búsqueda de los caminos de luz importantes.

Sobre las estrategias de recorrido de los caminos en los métodos de iluminación global, las dos principales son las estrategias denominadas *depth-first search* y *breadth-first search* [68]. La estrategia *depth-first-search* sigue un camino de luz por vez hasta que el rayo encuentre el ojo o cámara que capta la imagen, la potencia del rayo sea muy debil o hasta que el número de reflexiones llegue a un valor preestablecido. Por otra parte, en la estrategia *breadth-first search* se lanzan todos los rayos a la vez y se siguen todos los caminos simultaneamente, agregándose un paso por vez en cada camino.

Para aplicar cualquiera de estas técnicas al problema de radiosidad donde las superficies son de reflexión difusa, se divide la escena en parches (siguiendo los conceptos de elementos finitos) y se almacena en cada paso de cada camino en qué parche incide el rayo y cuánta potencia contiene el rayo reflejado por el parche. La imagen final se genera a partir de los valores que cada parche tiene como potencia radiada, que se corresponde a la potencia emitida más la potencia reflejada.

2.4. Aplicación del método de Galerkin

En la sección anterior se introdujeron algunos de los métodos existentes para resolver el problema de radiosidad. En la resolución basada en elementos finitos utilizando el método de Galerkin se llegó a la formulación de un sistema lineal presentado en la Ecuación (2.32), cuya matriz tiene entre sus componentes a los factores de forma \mathbf{F}_{ij} que se calculan a través de una integral doble presentada en la Ecuación (2.30).

Los aspectos de mayor dificultad al aplicar el método de Galerkin son el cálculo de los factores de forma y la resolución del sistema lineal. En esta sección se describen algunos de los principales enfoques para resolver la construcción de los factores de forma. Se detallan las principales propiedades de la matriz \mathbf{F} y de la matriz (o producto matricial) \mathbf{RF} . La resolución del sistema lineal generado es tratada en la sección 2.5.

2.4.1. Cálculo de los factores de forma

El cálculo de los factores de forma es relativamente complejo debido a la necesidad de resolver una integral doble (véase la Ecuación (2.30)) y a que es necesario calcular la visibilidad entre los diferenciales de área de los parches, para determinar la obstrucción visual entre pares de puntos debida a los parches de la escena. A esta complejidad se añade que hay que calcular n^2 factores de forma, siendo n la cantidad de parches, por lo que resulta imprescindible encontrar un método de cálculo eficiente.

Los factores de forma generan la matriz \mathbf{F} , que suele ser densa, sobre todo en espacios geométricos amplios. En los espacios geométricos amplios, cada punto de la escena puede observar una gran parte del resto de los puntos de la escena. En esta situación, existe una probabilidad alta de que dos parches i y j cualesquiera se vean y que los factores de forma asociados (es decir \mathbf{F}_{ij} y \mathbf{F}_{ji}) no valgan cero. Si este razonamiento se extiende a todos los pares de parches posibles, resulta probable que la matriz generada sea densa. Es común encontrar artículos donde se ha trabajado con escenas que contienen del orden de millones de polígonos [7, 40]. Por lo tanto, no resulta extraño hablar de órdenes de 10^{12} factores de forma para una escena. Estos órdenes de magnitud son un llamado de atención sobre la imposibilidad de generar todos los factores de forma e incentivan la búsqueda de metodologías que se basen en la utilización de sólo algunos factores de forma o, mejor aún, de ninguno.

A continuación se mencionarán algunos conceptos básicos relacionados con los factores de forma, así como algunas de las técnicas más exitosas para su cálculo.

Analogía de Nusselt

La analogía de Nusselt es un argumento geométrico para el cálculo de los factores de forma tal como está expresado en la Ecuación (2.30), entre un diferencial de área y un área (sin incluir visibilidad). Su representación gráfica se presenta en la Figura 2.6.

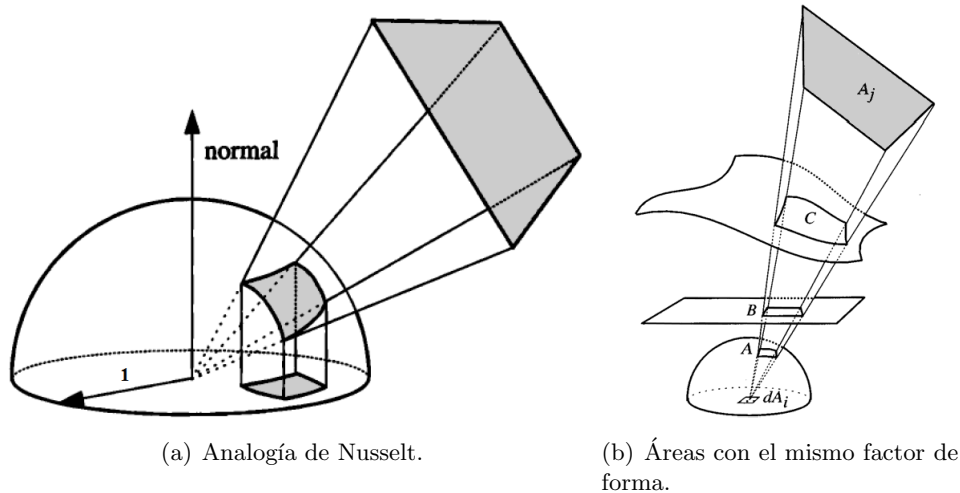


Figura 2.6: Cálculo geométrico de factores de forma (extraídas de Cohen et al. [15]).

La Figura 2.6 (a) muestra una esfera de radio 1, centrada en un diferencial de superficie, que contiene un parche proyectado sobre un hemisferio de radio 1. Esa proyección está proyectada ortogonalmente a la base del hemisferio. La fracción del área de la base que está cubierta por la proyección del parche es igual al valor del factor de forma entre el diferencial de área y el parche, es decir, es igual a la porción de potencia emitida del diferencial de área hacia el parche.

La Figura 2.6 (b) muestra otra propiedad interesante: cualquier objeto que cubra la misma área en el hemisferio tendrá el mismo factor de forma, por ocupar el mismo ángulo sólido.

Los factores de forma son no negativos. Si sobre un parche i cualquiera se genera un hemisferio de radio 1 y se proyectan todos los parches visibles desde i , la suma de todos los factores de forma obtenidos no puede ser mayor a 1 (Ecuación (2.35)). Esto se debe a que todos los parches visibles se proyectan en áreas disjuntas en la base del hemisferio, y por lo tanto, la suma de todas las áreas no puede ser superior al área de la base.

$$\sum_{j_1}^n \mathbf{F}_{ij} \leq 1, \forall i \quad (2.35)$$

Algoritmo del hemicubo

Para incluir la visibilidad entre parches en el cálculo de los factores de forma, se puede utilizar la técnica desarrollada por Catmull [12, 24], denominada *algoritmo de Z-buffer*. Este algoritmo agrega información en cada píxel, de la distancia (en el eje Z) que hay entre el centro de proyección (el ojo del observador) y el polígono proyectado en él. Dado un polígono proyectado sobre una ventana de un dispositivo gráfico (un plano de vista), cada píxel contiene información sobre el color del polígono en la zona abarcada por el píxel, y el Z-buffer almacena la distancia promedio (en Z) entre esa zona del polígono y el centro de proyección. Si hay varios polígonos dibujados y éstos se superponen, en cada píxel se ve el polígono más cercano, así como la distancia más cercana. El algoritmo de Z-buffer está implementado en la mayoría de las tarjetas gráficas actuales.

Utilizando el algoritmo de Z-buffer, Cohen et al. desarrollaron el algoritmo del hemicubo [14] (véase la Figura 2.7).

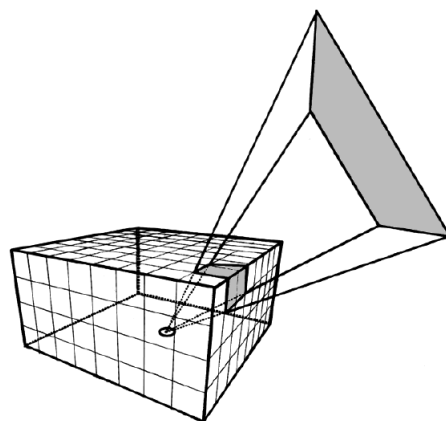


Figura 2.7: El hemicubo (figura extraída de Cohen et al. [15]).

El algoritmo del hemicubo, centrado en un parche i cualquiera, calcula los factores de forma \mathbf{F}_{ij} con todos los parches j de la escena. El algoritmo de Z-buffer se basa en realizar cinco proyecciones de los parches de la escena en las cinco superficies de un medio cubo, con proyecciones que utilizan el algoritmo Z-buffer. Al final del proceso, cada píxel del hemicubo contiene la información de los parches que son visibles desde el diferencial de área. Para cada píxel del hemicubo se calcula el factor de forma correspondiente, y se lo almacena en una tabla para su posterior utilización. Entonces, el factor de forma entre el parche i y un parche j cualquiera es aproximadamente igual a la suma de los factores de forma de los píxeles en los que está proyectado.

El algoritmo del hemicubo ha resultado muy exitoso y es ampliamente utilizado en las técnicas de radiosidad que implican el cálculo de los factores de forma.

Un problema de esta técnica son los *artefactos de discretización* o *aliasing* que se puedan generar al trabajar con píxeles. Por ejemplo, pueden haber objetos pequeños de tamaño inferior al píxel que no queden registrados en el hemicubo, pero que emitan cantidades de luz significativas como para influir en la escena. Otro problema importante del algoritmo del hemicubo está relacionado con el costo de realizar cinco proyecciones por cada parche (una por cada cara del hemicubo). Este problema fue estudiado entre otros por Coombe et al. [16] quienes propusieron cambiar la proyección en un hemicubo por una proyección estereográfica, que permite proyectar todo el hemisferio visible en una superficie acotada, al estilo de la analogía de Nusselt. Otra proyección reportada es la proyección parabólica, donde la escena se proyecta sobre un paraboloides. Esta proyección fue desarrollada por Heidrich y Seidel [41] y originalmente fue pensada para su aplicación en la construcción de los *environment maps* [10].

Cálculo de los factores de forma con los métodos de Monte Carlo

Un método usualmente utilizado para el cálculo de los factores de forma está basado en la utilización de la integración de Monte Carlo. En este método, dados dos parches i y j entre los que se quiere calcular el factor de forma \mathbf{F}_{ij} (véase la Figura 2.8), se toman al azar n pares de puntos $[x_i(k), x_j(k)]$, donde $x_i(k)$ pertenece al parche i y $x_j(k)$ pertenece al parche j . Se evalúa si hay visibilidad entre ellos, y si hay visibilidad se calcula el factor de forma mediante la Ecuación (2.36), siendo $\theta_{x_i(k)}$ y $\theta_{x_j(k)}$ los ángulos entre las normales de los parches i y j en $x_i(k)$ y $x_j(k)$ respectivamente, y el vector $\overrightarrow{(x_i(k), x_j(k))}$, r es la distancia entre $x_i(k)$ y $x_j(k)$, A_j es el área del parche j y n es el total de pares de puntos. Este algoritmo puede ser mejorado tomando muestras de pares de puntos que estén distribuidos de manera no uniforme y adaptativa (para más detalles véase el trabajo de Cohen et al. [15]).

$$\mathbf{F}_{ij} = A_j \sum_{\substack{[x_i(k), x_j(k)] \text{ visibles} \\ \cos \theta_{x_i(k)} \cos \theta_{x_j(k)} > 0}} \frac{\cos \theta_{x_i(k)} \cos \theta_{x_j(k)}}{\pi r^2 + \frac{A_j}{n}} \quad (2.36)$$

Además de los métodos basados en el uso de los hemicubos y en los métodos Monte Carlo, otros métodos analíticos y numéricos han sido desarrollados para hallar los factores de forma. Existen métodos que hallan los factores de forma entre áreas y otros que los hallan entre diferencial de área y área; hay métodos basados en Monte Carlo, métodos jerárquicos, métodos basados en integrales de contorno, etc. Cohen et al. [15] han propuesto una amplia taxonomía para los métodos existentes.

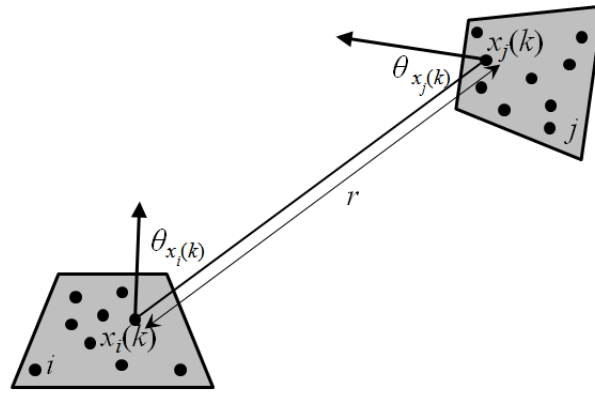


Figura 2.8: Cálculo de factores de forma con Monte Carlo.

2.4.2. Propiedades de \mathbf{R} y \mathbf{F}

La ecuación matricial de radiosidad, obtenida a partir de la ecuación de radiosidad, tiene como componentes importantes a las matrices \mathbf{R} y \mathbf{F} . La matriz \mathbf{R} se calcula a partir de la información disponible de las superficies de la escena, y la matriz \mathbf{F} se compone de factores de forma, que se calculan según los conceptos vistos en la sección anterior.

A continuación se presentan las propiedades de \mathbf{R} y \mathbf{F} que serán útiles en el desarrollo de los conceptos de radiosidad de rango bajo, que se presentan en el Capítulo 3.

1. \mathbf{R} y \mathbf{F} son matrices $n \times n$ donde n es el número de parches de la escena.
2. \mathbf{R} es una matriz diagonal positiva.
3. $0 \leq \mathbf{R}_{ii} = \rho_i \leq 1$, donde ρ_i indica el índice de reflexión del parche i .
4. \mathbf{F} es una matriz densa (con relativamente pocos valores nulos) y positiva. El espacio necesario para almacenar \mathbf{F} es $O(n^2)$ números de punto flotante, especialmente en espacios geométricos abiertos (véanse las propiedades de \mathbf{F}_{ij}). La matriz \mathbf{F} puede no ser densa en casos especiales, como por ejemplo en un laberinto de túneles donde la iluminación de cada parche sólo afecte a unos pocos parches cercanos.
Si la escena se compone de varias habitaciones incomunicadas entre sí, entonces la matriz es diagonal por bloques, con tantos bloques como habitaciones incomunicadas haya, y el sistema lineal se puede separar en varios sistemas lineales que pueden resolverse en forma independiente.
5. \mathbf{F}_{ij} es un factor de forma e indica la porción de radiosidad que incide en el parche j del total de la radiosidad emitida y reflejada por el parche i .
6. Si $\mathbf{F}_{ij} \neq 0$ entonces $\mathbf{F}_{ji} \neq 0$. Si el parche i “ve” al parche j entonces también j “ve” a i , y por lo tanto siempre hay intercambios de energía luminosa en ambos sentidos. En la práctica, al generar los factores de forma puede suceder que esta propiedad no se cumpla, debido a errores inherentes al método de generación.

7. $0 \leq \mathbf{F}_{ij} \leq 1, \forall i, j$ dado que, según la analogía de Nusselt, se puede calcular como un cociente entre una porción del área de la base del hemisferio y el área total de esa base.
8. $\mathbf{F}_{ii} = 0$ en la mayoría de las escenas, debido a que se utilizan parches planos. Dos diferenciales de área pertenecientes a un mismo parche plano no se ven entre sí ni intercambian radiosidad.
9. $0 \leq \sum_{j=1}^n \mathbf{F}_{ij} \leq 1, \forall i$. La sumatoria permite calcular cuál es la proporción de toda la energía emitida por el parche i que incide en el resto de los parches. La sumatoria es menor estricto que 1 para algún parche i cuando parte de la energía emitida por este no incide en ningún parche, “perdiéndose” en el espacio.
10. \mathbf{AF} es una matriz simétrica, donde \mathbf{A} es una matriz diagonal y \mathbf{A}_{ii} es el área del parche i . En la práctica, la simetría sólo se cumple de manera aproximada, debido a que \mathbf{F} se calcula a través de métodos que pueden tener importantes valores de error. La distancia $\|\mathbf{AF} - (\mathbf{AF})^T\|$ podría ser utilizada como un indicador de la calidad de \mathbf{F} . A partir de una matriz \mathbf{F} se puede hallar otra \mathbf{F}_* tal que \mathbf{AF}_* sea simétrica, si se aplica la Ecuación (2.37). En ese caso hay que cuidar que se sigan cumpliendo las otras propiedades de la lista, en especial $0 \leq \mathbf{F}_{ij} \leq 1$ o $0 \leq \sum_{j=1}^n \mathbf{F}_{ij} \leq 1$, que al no cumplirse comprometen la resolución del sistema lineal (véase el siguiente ítem).

$$\mathbf{F}_* = \frac{1}{2} \mathbf{A}^{-1} (\mathbf{AF} + (\mathbf{AF})^T) \quad (2.37)$$

11. La **matriz de radiosidad**, definida en la Ecuación (2.38), es diagonal dominante, es decir, se cumple la relación presentada en la Ecuación (2.39), que implica que la matriz es invertible [29] y por lo tanto en la ecuación matricial de radiosidad existe una única solución B por cada E , como la física sugiere.

$$\mathbf{K} = (\mathbf{I} - \mathbf{RF}) \quad (2.38)$$

$$|\mathbf{K}_{ii}| > \sum_{j=1, j \neq i}^n |\mathbf{K}_{ij}|, \forall i \quad (2.39)$$

Desde el punto de vista teórico podría ocurrir la igualdad $|\mathbf{K}_{ii}| = \sum_{j=1, j \neq i}^n |\mathbf{K}_{ij}|$ para algún i , como por ejemplo cuando $\rho_i = 1$, $\mathbf{F}_{ii} = 0$ y $\sum_{j=1}^n \mathbf{F}_{ij} = 1$. Cuando esta igualdad se cumple, es posible que \mathbf{K} tenga uno o más valores propios iguales a 0 (véase el teorema de Gershgorin en Saad et al. [62]). Bajo estas hipótesis la matriz \mathbf{K} no sería invertible y por lo tanto la ecuación matricial de radiosidad (2.33) no tiene solución. El que la matriz \mathbf{K} sea diagonal dominante posibilita la resolución del sistema lineal por los métodos iterativos de Jacobi y Gauss-Seidel (véase la Sección 2.5).

12. $\|\mathbf{RF}\| < 1$ para alguna norma. Esto se debe a que el radio espectral de \mathbf{RF} es menor que 1. El **radio espectral** de una matriz es igual al máximo valor propio en valor absoluto. Si una matriz tiene un radio espectral menor que 1, se puede demostrar que siempre existe una norma menor que 1 para esa matriz [62]. Esta propiedad es fundamental para la propiedad del cálculo de la inversa de la matriz

de radiosidad utilizando la serie de Neumann. Físicamente hablando, cuanto más cercano a uno sea el radio espectral de \mathbf{RF} , la luz reflejará más veces antes de ser absorbida por las superficies [15].

13. $(\mathbf{I} - \mathbf{RF})^{-1} = \sum_{i=0}^{\infty} (\mathbf{RF})^i$ (serie de Neumann). Esta igualdad se cumple debido a que

$$(\mathbf{I} - \mathbf{RF}) \sum_{i=0}^{\infty} (\mathbf{RF})^i = [\mathbf{I} + \mathbf{RF} + (\mathbf{RF})^2 + \dots] - [\mathbf{RF} + (\mathbf{RF})^2 + \dots] = \mathbf{I} \quad (2.40)$$

por lo tanto, existe un método iterativo de calcular la inversa de la matriz de radiosidad (2.38). La serie $\sum_{i=0}^{\infty} (\mathbf{RF})^i$ converge siempre, porque el radio espectral de \mathbf{RF} es menor que uno. La norma de la serie está acotada por

$$\left\| \sum_{i=1}^{\infty} (\mathbf{RF})^i \right\| \leq \sum_{i=1}^{\infty} \|\mathbf{RF}\|^i = \frac{1}{1 - \|\mathbf{RF}\|} \quad (2.41)$$

que es un valor positivo para aquellas normas en que $\|\mathbf{RF}\| < 1$.

14. Según Cohen et al. [15], la matriz $\mathbf{I} - \mathbf{RF}$ está generalmente bien condicionada, o sea que el cociente entre el mayor y el menor valor propio (en valor absoluto) no es en general un número grande. Al resolver un sistema lineal por un método numérico con notación de punto flotante, el que la matriz esté bien condicionada implica que la solución calculada no es muy sensible a los errores que puedan surgir de la perturbación de los datos.

Dadas las características de la matrices \mathbf{R} y \mathbf{F} , muchas propiedades de \mathbf{F} son cumplidas por \mathbf{RF} . Por ejemplo, en la lista anterior las propiedades 1, 4, 6, 7, 8 y 9 son cumplidas por \mathbf{F} y por \mathbf{RF} .

2.5. Resolución de la ecuación matricial de radiosidad

Luego de hallados los factores de forma se tienen todos los elementos necesarios para resolver el sistema lineal expresado en las ecuaciones de radiosidad (Ecuaciones (2.31), (2.32) y (2.33)). Esta sección presenta una breve exposición de algunas ideas básicas y métodos clásicos para la resolución de sistemas lineales.

Los principales inconvenientes que tiene el sistema lineal generado para el problema de radiosidad, están relacionados con el número de parches (n) que involucra el modelado de escenas de la vida real. Como se mencionara en la sección 2.4.1, los parches pueden llegar a ser millones y los factores de forma billones. Almacenar un billon de números de punto flotante en memoria principal y resolver sistemas lineales densos por métodos directos (al estilo de la escalerización gaussiana) es sólo posible en clusters de computadoras. La escalerización gaussiana, al tener una complejidad de $O(n^3)$ [29], implica un costo del orden del exaflop (10^{18} flop) cuando se trabaja con millones de parches. Para evitar estos inconvenientes, se han desarrollado metodologías que, a un costo computacional relativamente bajo, hallan soluciones aproximadas al problema de radiosidad. Las soluciones aproximadas son habitualmente suficientes para simular el comportamiento de la luz en los escenarios utilizados.

2.5.1. Métodos de relajación

En esta sección se desarrollan algunos conceptos de los principales métodos iterativos para la resolución iterativa de la ecuación matricial de radiosidad. Estos métodos son los métodos iterativos (también llamados métodos de relajación) y los métodos de relajación estocástica, que eliminan la necesidad de calcular los factores de forma.

Métodos iterativos

Los principales métodos para la resolución de la ecuación matricial de radiosidad son los métodos iterativos. Los métodos iterativos más populares son el método de Jacobi, el método de Gauss-Seidel y la sobrerelajación sucesiva (SOR)[29, 8].

Los *métodos iterativos* más utilizados proponen transformar el sistema $\mathbf{A}x = b$ en otro sistema $x = \mathbf{C}x + d$, para resolverlo mediante las iteraciones $x^{k+1} = \mathbf{C}x^k + d$ a partir de cierto x^k preestablecido. Para que x^k converja a la solución, el radio espectral de \mathbf{C} tiene que ser menor que uno.

La matriz \mathbf{C} y el vector \mathbf{d} se pueden construir de diferentes maneras. Construyendo un par de matrices \mathbf{M} y \mathbf{N} tal que $\mathbf{M} - \mathbf{N} = \mathbf{A}$, para que el sistema lineal se pueda expresar como $\mathbf{M}x = \mathbf{N}x + b$ y por lo tanto $x = \mathbf{M}^{-1}\mathbf{N}x + \mathbf{M}^{-1}b$. Los métodos de Jacobi, Gauss-Seidel y SOR son sólo algunos métodos populares entre todos los posibles. Su éxito radica en que las matrices \mathbf{M} involucradas son fácilmente invertibles.

Dada la matriz \mathbf{A} se define la igualdad $\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U}$, donde \mathbf{D} se compone de la diagonal de \mathbf{A} , y $-\mathbf{L}$ y $-\mathbf{U}$ contienen respectivamente las partes estrictamente triangular inferior y superior de \mathbf{A} .

Jacobi cumple que $\mathbf{M} = \mathbf{D}$ y que $\mathbf{N} = \mathbf{L} + \mathbf{U}$. Cada paso de la iteración es

$$x_i^{k+1} = \frac{1}{a_{ii}} \left(b_i - \sum_{j \neq i} a_{ij} x_j^k \right), i = 1, \dots, n \quad (2.42)$$

Gauss-Seidel cumple con $\mathbf{M} = \mathbf{D} - \mathbf{L}$ y con $\mathbf{N} = \mathbf{U}$. Cada paso de la iteración es

$$x_i^{k+1} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{k+1} - \sum_{j=i+1}^n a_{ij} x_j^k \right), i = 1, \dots, n \quad (2.43)$$

Como se puede apreciar, cada valor x_i^{k+1} calculado sustituye inmediatamente al valor previo x_i^k .

SOR asume que $x_i^{k+1} = w x_i^{k+1} + (1-w)x_i^k$ cumpliéndose que

$$\mathbf{C} = (\mathbf{D} - w\mathbf{L})^{-1} [(1-w)\mathbf{D} + w\mathbf{U}] , d = w(\mathbf{D} - w\mathbf{L})^{-1} b$$

donde w es un parámetro a ajustar que vale entre 0 y 2 (cuando $w = 1$, SOR es igual al método de Gauss-Seidel). Cada paso de la iteración consiste en

$$x_i^{k+1} = (1-w)x_i^k + \frac{w}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{k+1} - \sum_{j=i+1}^n a_{ij} x_j^k \right), i = 1, \dots, n \quad (2.44)$$

En los métodos presentados, las iteraciones continúan hasta que $\|x^{k+1} - x^k\|$ o la norma del residuo sean menores a ciertos valores prefijados. Al **residuo** r^k se lo define como $b - \mathbf{A}x^k$. Una variante del método de Gauss-Seidel, llamado *método de Southwell*, en lugar de calcular todos los x_i^k en secuencia para $i = 1, \dots, n$, establece en qué valor de i se halla el mayor residuo r_i^k , y calcula el correspondiente x_i^k . El cálculo del valor del residuo se realiza de forma incremental a partir del valor anterior. Así, el cálculo de cada nuevo residuo tiene una complejidad $O(n)$, como ocurre en cada paso de Jacobi y Gauss-Seidel. Al igual que en Gauss-Seidel, el nuevo valor hallado de x_i^k sustituye inmediatamente al valor anterior.

A su vez, el método de Southwell posee una variante denominada *refinamiento progresivo*, que es aplicada a la resolución de la ecuación de radiosidad con el fin de permitir la visualización de los cambios luego de cada paso del proceso iterativo. El método de refinamiento progresivo brinda retroalimentación inmediata en su proceso iterativo. En lugar de modificar el vector r por cada B_i hallado, se modifica el vector B . Según Cohen et al. [15], el vector B obtenido en cada paso de refinamiento progresivo es igual al vector $B + r$ en el método de Southwell. Más información sobre el método de refinamiento progresivo se puede consultar en los trabajos de Cohen et al. [13, 15].

Relajación estocástica

Los métodos de **relajación estocástica** son una combinación entre los métodos de relajación tradicionales (Jacobi, Gauss-Seidel, etc.) y los métodos de Monte Carlo. Como se observara en las Ecuaciones (2.42), (2.43) y en (2.44), en los métodos de relajación se realizan sumas del tipo $\sum_j a_{ij}x_j$. En la ecuación matricial de radiosidad, los coeficientes a_{ij} son de la forma $(\delta_{ij} - \rho_i \mathbf{F}_{ij})$ (Ecuación (2.29)) y las sumas tienen la forma

$$B_i - \rho_i \sum_j \mathbf{F}_{ij} B_j \quad (2.45)$$

donde x es sustituido por B . La idea básica de la relajación estocástica es la de estimar esas sumas utilizando Monte Carlo, sin tener que calcular cada coeficiente a_{ij} .

Para el caso de la ecuación matricial de radiosidad, la relajación estocástica hace posible la resolución del sistema lineal sin conocer los factores de forma. Para lograr este objetivo, se utiliza la Ecuación (2.46).

$$\lim_{N \rightarrow \infty} \left(\frac{\sum_{k=1}^N B_{P_k}}{N} \right) = \sum_{j=1}^n \mathbf{F}_{ij} B_j \quad (2.46)$$

La igualdad de la Ecuación (2.46) se cumple porque los P_k son puntos distribuidos uniformemente en la base del hemisferio de Nusselt y B_{P_k} es la radiosidad proyectada según la analogía de Nusselt en cada punto P_k (como en la Figura 2.9). Por lo tanto, en el límite se obtiene la radiosidad promedio de la base del hemisferio. Del lado derecho de la igualdad se tiene la sumatoria de la radiosidad B_i de cada parche multiplicada por la porción del área que el parche ocupa en la base del hemisferio (que es \mathbf{F}_{ij} por la analogía de Nusselt). El resultado de esta última sumatoria es también la radiosidad promedio de la base del hemisferio.

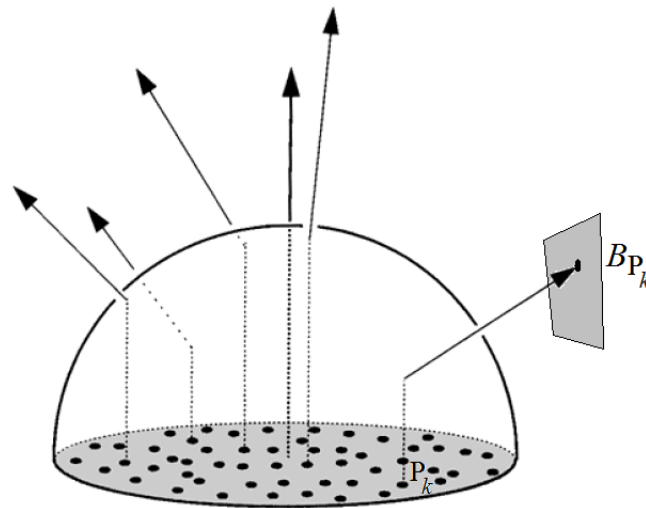


Figura 2.9: Representación gráfica del método de Malley.

Para calcular $(\sum_{k=1}^N B_{P_k})/N$ de forma eficiente y con la suficiente cantidad de puntos como para tener una varianza pequeña, se utiliza el método de Malley [54]. Este método parte de los puntos P_k de la base del hemisferio y los proyecta hacia el espacio, haciendo el recorrido inverso al propuesto por Nusselt. Los rayos se dirigen hacia la escena y por cada rayo se recoge el valor de radiosidad B_{P_k} correspondiente al polígono más cercano que interseca el rayo.

2.5.2. Métodos no estacionarios

Los métodos de relajación presentados en la subsección anterior también son conocidos como métodos estacionarios. Este nombre proviene de que en estos métodos, salvo el vector x (B en el caso de radiosidad), los términos utilizados permanecen constantes a lo largo del proceso iterativo. Por el contrario, los métodos no estacionarios involucran información que cambia en cada iteración [6].

Hay una variedad mayor de métodos estacionarios que de métodos no estacionarios. En esta última categoría se encuentran los métodos de Gradiente Conjugado (CG), los de residuos mínimos ($MINRES$), los de residuos mínimos generalizados ($GMRES$) y la iteración de Chebyshev.

Para la resolución de la ecuación matricial de radiosidad se han presentado desarrollos basados en Chebyshev [3], hay trabajos que utilizan $GMRES$ [56], y otros que utilizan a $MINRES$ [2], entre otros. La poca utilización de los métodos no estacionarios para la resolución de la ecuación matricial de radiosidad quizá se deba al éxito conseguido con los métodos estacionarios.

2.5.3. Métodos jerárquicos

El principal reto en los métodos que utilizan factores de forma para resolver la ecuación de radiosidad, radica en la relación existente entre la precisión de la solución y la cantidad de factores de forma utilizados. Para generar soluciones con buena precisión es necesario tener un número elevado de parches. La cantidad de factores de forma al utilizar la matriz \mathbf{F} es cuadrática en relación al número de parches. Esto implica que cuanto mayor sea la precisión buscada, existirá un mayor consumo de memoria y mayores costos asociados al cálculo de los factores de forma y a la resolución del sistema lineal.

Para afrontar estos problemas se han desarrollado métodos jerárquicos. Estos métodos surgieron como inspiración de los algoritmos desarrollados por Esselink [21], Barnes y Hut [4], y Greengard y Rokhlin [36], quienes presentaron algoritmos con $O(n \log n)$ para el problema de los n cuerpos (*n-body problem*). El **problema de los n cuerpos** trata sobre la descripción de la dinámica de n objetos en movimiento, atraídos entre sí por fuerzas gravitatorias. Este problema tiene una complejidad $O(n^2)$ si se consideran las fuerzas gravitatorias entre todos los pares de cuerpos, pero la complejidad puede reducirse hasta $O(n \log n)$ si los objetos cercanos entre sí se consideran como un único cuerpo al interactuar con objetos lejanos. Hanrahan et al. [38] observaron que puede desarrollarse una variante del problema de los n cuerpos que en lugar de operar con la fuerza gravitatoria de cada cuerpo considere la radiosidad de cada cuerpo.

A continuación se realiza una breve introducción al método de radiosidad jerárquica (*hierarchical radiosity*) y se mencionan los puntos principales de la radiosidad basada en wavelets (*wavelets radiosity*).

Radiosidad jerárquica

Para conocer la interacción de energía luminosa existente entre un parche A y el resto de la escena (véase la Figura 2.10 (a)) se comienza calculando los factores de formas entre pares de parches. Luego, con algún criterio se decide dividir o no a los parches de la escena. Se comienza con una escena con parches grandes (por ejemplo, con los parches A , B , C y D) y estos se van dividiendo según las necesidades que surjan en la escena.

Las divisiones de los parches se pueden hacer a la medida de cada interacción. Por ejemplo, en las Figuras 2.10 (b) y (c) se observa que una división eficiente de A realizada para analizar su interacción con B , es diferente a la división que se debe hacer de A para analizar su interacción con C . En la Figura 2.10 (d) se observa que no hace falta dividir los parches A y D , por estar muy alejados entre sí.

Para evaluar si un parche A se debe dividir o no, Hanrahan et al. [38] propusieron un método que primero estima los valores de los factores de forma a través de una fórmula sencilla. La estimación no considera la oclusión por otros parches, entre otras simplificaciones que aceleran considerablemente la velocidad de cálculo. A pesar de estas simplificaciones, el método permite obtener buenos resultados.

Cuando se calcula la relación del parche A con el parche B , en primer término se estima el factor de forma \mathbf{F}_{BA} , que indica la porción de energía que incide en A del total emitido por B . Si este valor es mayor a un umbral predeterminado, entonces se divide el parche A . Luego, se estima el factor de forma \mathbf{F}_{AB} para saber si es necesario dividir al parche B .

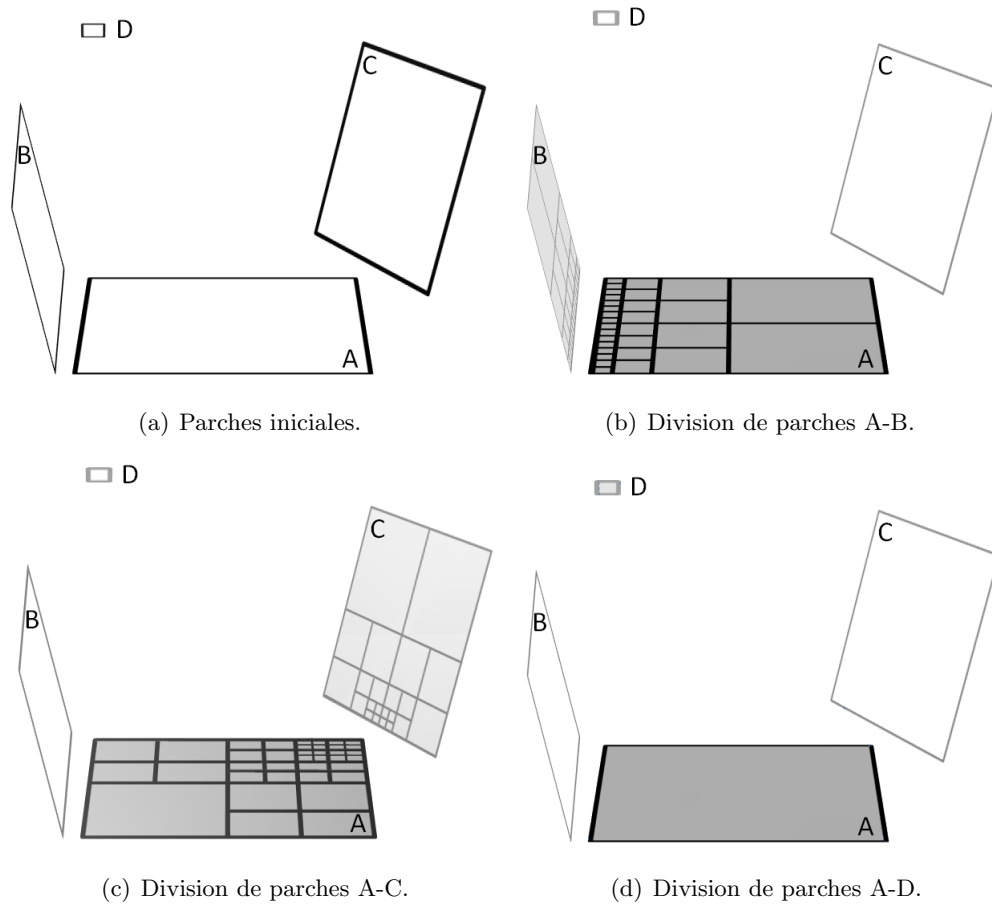


Figura 2.10: División del parche A para reflejar mejor el intercambio de energía con cada parche de la escena.

La conveniencia de dividir parches está dada porque, si el factor de forma \mathbf{F}_{BA} es grande, la radiosidad que emite B puede no incidir de manera uniforme en A . Entre las zonas más cercanas de A y B hay mayor intercambio de energía que entre las zonas más alejadas, por lo tanto, resulta conveniente dividir ambos parches para disminuir el error de la distribución de radiosidad.

Los parches se pueden dividir siguiendo una estructura jerárquica de *quadtrees*. Geométricamente, un *quadtrees* consiste en la división de un parche en cuatro partes iguales. Cada parte se la considera como un nuevo parche y por lo tanto es posible de ser también dividida. Como estructura de datos, un *quadtrees* es un árbol donde el nodo raíz es el parche original y cada nodo tiene cuatro hijos. Las sucesivas divisiones de parches generan un conjunto de *quadtrees*, un *quadtrees* por cada parche original.

Para optimizar el número de factores de forma calculados, conviene realizar en cada parche una división diferente según contra qué parche se está evaluando su interacción. En la Figura 2.10 se muestra que no es necesario calcular los factores de forma entre las hojas de A y las hojas de B , C o D , sino que casi siempre alcanza con calcular factores de forma para los nodos superiores. En la figura se aprecia que, dada la distancia entre A y D , alcanza con calcular los factores de forma entre los nodos raíces de A y D

para obtener un buen resultado. En este caso son innecesarios los nodos de los quatrees asociados a A y D . En cambio, los factores de forma entre los nodos raíz de A y B tienen valores muy altos y por lo tanto conviene calcular los factores de forma correspondientes a otros nodos en cada árbol.

Debido a estas particularidades, para realizar un cálculo eficiente de los factores de forma resulta conveniente almacenar en cada nodo de cada quadtree una lista de los nodos de los otros quadtrees con los que interactúa. En la Figura 2.10 (d) es el nodo raíz de A el que interactúa con el nodo raíz de D .

En la Figura 2.11 se ejemplifica la división de dos parches cercanos.

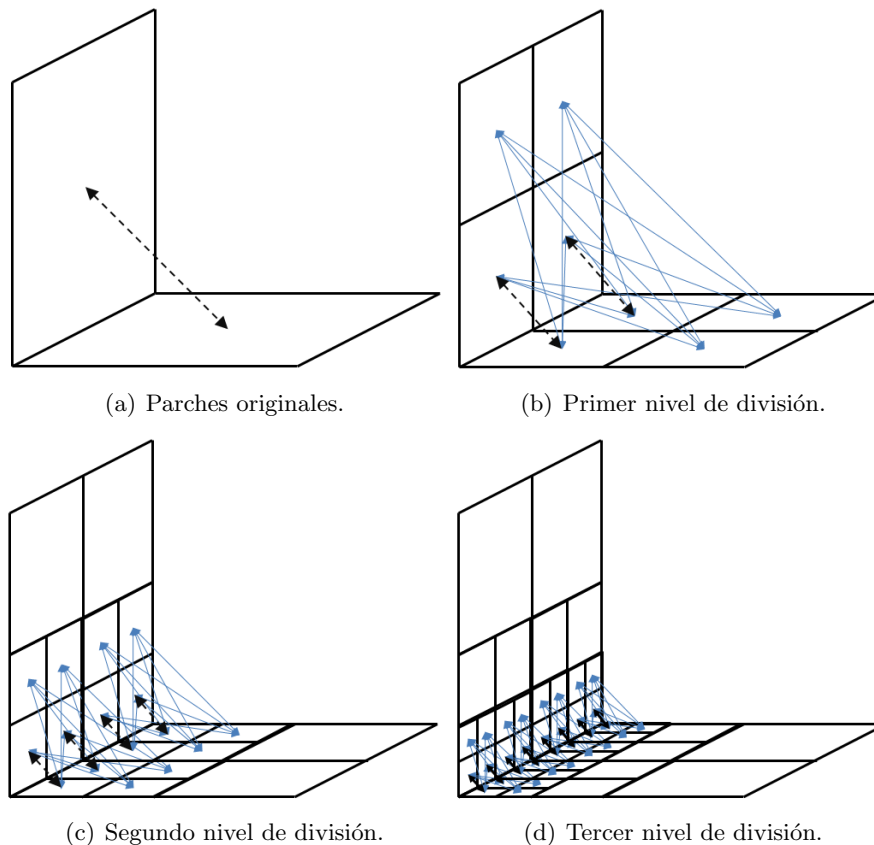


Figura 2.11: División de dos parches cercanos para reflejar correctamente el intercambio de radiosidad existente entre ellos.

Como se mencionara, los factores de forma entre dos parches no deben superar un umbral. Si el umbral de un factor de forma es superado, entonces el parche destino es dividido en cuatro. En la Figura 2.11, las líneas punteadas negras unen parches cuyos factores de forma superaron el umbral (son parches que hay que dividir) y las líneas azules continuas unen parches cuyos factores de forma están por debajo del umbral (donde la división no es necesaria). Por ejemplo, en la Figura 2.11 (a) hay dos parches ortogonales cuyos factores de forma son mayores al umbral, por lo tanto deben ser divididos. En la Figura 2.11 (b) se muestra cómo la mayoría de los factores de forma están por debajo del umbral establecido, salvo en las dos interacciones entre los pares de parches más cercanos. Por lo tanto, se dividen esos cuatro parches (y los parches resultantes

de la división sólo se utilizan para el cálculo de los factores de forma faltantes). En la Figura 2.11 (c) se tienen los nuevos factores de forma hallados. En estos factores de forma ocurre lo mismo que en 2.11 (b) a otra escala, lo que obliga a una nueva división y nuevos cálculos de factores de forma (véase la Figura 2.11 (d)). El proceso continúa hasta que los nuevos parches sean menores a un área prefijada o cuando los factores de forma sean menores a un valor preestablecido. En la Figura 2.12 se visualiza el quadtree asociado a uno de los parches de la Figura 2.11.

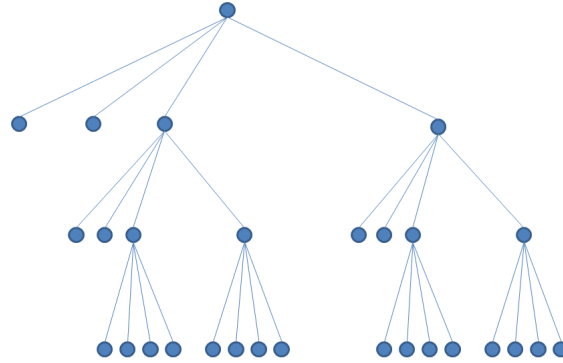


Figura 2.12: Quadtree de uno de los parches de la Figura 2.11.

La lista de nodos de los otros quatees que se incluye en cada nodo está representado en las líneas continuas de la Figura 2.11. Para que el cálculo de factores de forma sea eficiente y se eviten los cálculos innecesarios, sólo hay que calcular los factores de forma asociados a esas líneas continuas. Según Hanrahan et al. [38] la cantidad de factores de forma que se calculan de esta manera es proporcional al número de nodos de los árboles y no al cuadrado del número de nodos. Esta es una ventaja fundamental del método de radiosidad jerárquica respecto a otros métodos basados en el cálculo de la matriz \mathbf{F} : la reducción en un orden de la información que es necesario almacenar para el cálculo del sistema lineal.

Una vez calculados todos los factores de forma se está en condiciones de evaluar la radiosidad de la escena. El cálculo se realiza de forma iterativa, donde en cada paso de la iteración se realizan dos procesos para cada nodo de cada quadtree. Un primer proceso acumula en cada nodo toda la energía proveniente de cada link ya definido. Luego, un segundo proceso transmite la radiosidad recibida en cada nodo hacia sus hijos, y la transmite también hacia su padre (en este último caso se la pondera por el cociente entre el área del parche y el área del parche padre).

Radiosidad basada en wavelets

Gortler et al. [34] establecieron que el proceso de construcción de una estructura jerárquica de parches como la recién mostrada, es un caso particular de la aproximación de los factores de forma a wavelets.

Las **wavelets** son bases funcionales, como las transformaciones de Fourier, pero que a diferencia de éstas, que sólo varían en frecuencia, varían en frecuencia y tiempo. Existen múltiples bases funcionales del tipo wavelet, las más comúnmente utilizadas son la wavelet de Haar y la wavelet del tipo sombrero mejicano.

En la Figura 2.13 (a) se observan los primeros miembros de la base ortonormal de Haar (que tiene infinitos elementos). Como se observa, cada miembro de la base tiene una escala (relacionada con la frecuencia) y actúa sobre un área o período de tiempo específico. En la Figura 2.13 (b) se observa el resultado de aproximar una función a través de una combinación lineal de los primeros miembros de la base de Haar.

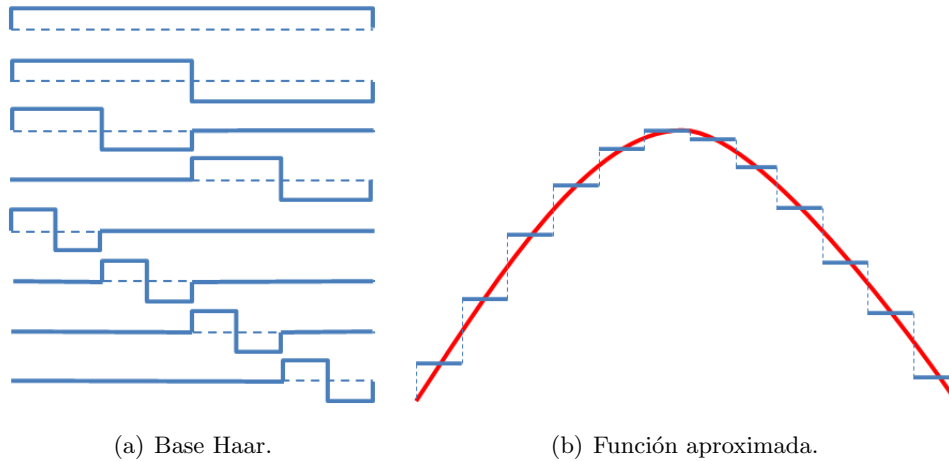


Figura 2.13: Primeros miembros de la base de Haar y aproximación de una función por esos primeros miembros.

Cuando se aproxima una función a través de una base de wavelets cualquiera, a cada miembro de la base le corresponde un peso diferente. La esencia del uso de wavelets consiste en seleccionar (utilizando algún método sencillo) la combinación de las funciones base que mejor modelan la variación de los factores de forma.

Si se modela la función de los factores de forma entre dos puntos cualquiera de la escena, la mayoría de los elementos de una base wavelet serán muy cercanos a cero [34]. Por lo tanto una matriz \mathbf{F} de factores de forma se puede aproximar esencialmente por una matriz dispersa, lográndose el objetivo de calcular unos pocos factores de forma, almacenarlos y utilizarlos para la resolución de la ecuación de radiosidad de rango bajo, cuya solución es cercana a la solución de la ecuación matricial de radiosidad.

2.6. Trabajos relacionados

En esta sección se reseñan brevemente algunos trabajos relacionados con las propuestas de esta tesis. Se analizan los trabajos que fueron fuente de inspiración para las ideas propuestas y aquellos que buscan objetivos similares (eficiencia, reducción de información, tiempo real) pero utilizando otros medios.

2.6.1. Antecedentes de coherencia en computación gráfica

El término coherencia ha sido muy utilizado en la comunidad gráfica, pese a lo cual resulta difícil encontrar una definición precisa del concepto. Uno de los primeros trabajos que utilizaron el concepto de coherencia espacial fue el de Sutherland et al. [67], quienes realizaron un análisis de diez algoritmos para la determinación de superficies visibles en

una escena. En dicho trabajo el término **coherencia** refiere al *grado en que el medio ambiente o su imagen es localmente constante*. Luego se distinguen distintos tipos de coherencia que cumplen los objetos, de forma tal que posibilite el ahorro de trabajo a los algoritmos para la determinación de superficies visibles.

Entre los algoritmos para la determinación de superficies visibles se destaca el algoritmo de Z-buffer [12, 24] -que es utilizado en el algoritmo del hemicubo- que aprovecha la coherencia de cada polígono para acelerar el cálculo de profundidad en cada píxel.

En radiosidad, el término coherencia ha sido utilizado para el desarrollo de algunos conceptos básicos. Cohen y Wallace [15] mencionan la coherencia en la geometría de las superficies visibles para el cálculo eficiente de los factores de forma, la coherencia entre celdas cercanas de la matriz y en las propiedades de las coordenadas de vista de los hemicubos.

Los algoritmos basados en los métodos de Monte Carlo aprovechan las propiedades de coherencia espacial para que la reducción del tamaño de las muestras relevadas no perjudique la calidad de las imágenes generadas. Como se expusiera en este capítulo, los métodos de Monte Carlo son utilizados en radiosidad para el cálculo de los factores de forma y la resolución de la ecuación matricial de radiosidad.

En las técnicas de iluminación global basadas en la traza de rayos también se utilizan conceptos relativos a la coherencia espacial. Como ejemplos se pueden mencionar las técnicas de *photon-mapping* [44] e *irradiance caching* [70]. En photon mapping, la cantidad de rayos lanzados -y de fotones almacenados- depende fuertemente de las propiedades de coherencia espacial de las superficies. En irradiance caching se aprovecha la coherencia espacial entre rayos para reducir el costo del muestreo de radiancia. Esto es verdadero en la iluminación indirecta, que varía suavemente. Por otra parte, en la técnica de *radiance caching* [50] se almacenan además muestras de la radiancia direccional utilizando armónicos esféricos.

Para el estudio de definiciones y usos de la coherencia en computación gráfica, es relevante el trabajo de Gröller [37], quien realizó un estudio exhaustivo de algoritmos y de estructuras de datos que explotan la coherencia espacial en el área.

2.6.2. Antecedentes en el uso de matrices de rango bajo en radiosidad

Ashdown [2] utilizó la propiedad de que el producto matricial $\mathbf{A}\mathbf{F} = \mathbf{G}$ genera una matriz simétrica, donde \mathbf{A} es una matriz diagonal con el área de los parches y \mathbf{F} es la matriz de los factores de forma, para transformar la ecuación matricial de radiosidad en otra que incluya una matriz simétrica. Ashdown se basó en trabajos anteriores que realizaron un análisis experimental de casos donde la matriz \mathbf{F} es simétrica, observando que unos pocos de sus valores propios tienen valores significativamente altos. El análisis de Ashdown mostró que en muchas escenas la matriz \mathbf{G} posee una cantidad pequeña de valores propios significativamente grandes. En base a estos conceptos, propuso un método iterativo que calcula soluciones aproximadas a B a partir del uso de los principales valores propios de \mathbf{G} .

En Hašan et al. [40] y Hašan [39] se define la matriz *unidades de superficie* \times *luces*, donde cada fila contiene información sobre la iluminación que una unidad de superficie recibe de todas las luces y cada columna contiene información sobre iluminación que una luz envía a todas las unidades de superficie. Esta matriz, que está muy emparentada con la matriz de factores de forma, posee un rango numérico bajo, por lo que puede ser convenientemente aproximada por matrices de rango bajo.

2.6.3. Antecedentes de algoritmos de iluminación global de tiempo real

La búsqueda de algoritmos de iluminación global que generen imágenes a velocidades compatibles con el tiempo real (al menos 20 imágenes por segundo) ha sido una constante en el área de computación gráfica. En los últimos años, el advenimiento de unidades de proceso gráfico con cientos de procesadores impulsó nuevas líneas de investigación en la temática. Entre las líneas de investigación desarrolladas se destacan las siguientes:

- El trabajo de Keller [48] sobre *instant radiosity*, basados en la idea de convertir la iluminación de manera aproximada en un número enorme de puntos de luz generados automáticamente. Las implementaciones realizadas en tiempo real tienen el inconveniente de ignorar la interreflexión de la luz entre los objetos.
- El trabajo de Sloan et al. [65] propone el precómputo de la transferencia de la radiancia (*precomputed radiance transfer*). La técnica se basa en la idea de realizar el precómputo y la compresión de los efectos de iluminación y sombreado que luces distantes provocan en la escena, utilizando los resultados para el cálculo eficiente de la iluminación global de la escena ante cambios de luz. La principal limitación de este enfoque es que solamente trabaja con luces distantes.
- El trabajo de Dachsbacher y Stamminger [17] sobre mapas de sombra reflectivos (*reflective shadow maps*). Los mapas de sombra son considerados fuentes de luz indirecta, con el objetivo de lograr la generación de iluminación indirecta, pero se niega la existencia de múltiples interreflexiones de luz entre las superficies.
- El trabajo de Nijasure et al. [57] basados en el uso de procesadores gráficos que realizan un muestreo de la radiancia incidente en grillas uniformes tridimensionales. En estos trabajos no se generan imágenes de buena calidad debido a la necesidad de utilizar mallas de baja resolución para conseguir velocidades compatibles con el tiempo real.
- El trabajo de Gautron et al. [27] sobre *radiance cache splatting*, basados en la técnica de *radiance caching* y en el uso de imágenes para recoger la radiancia final de la escena. Si bien los métodos son rápidos, ignoran oclusiones en la etapa final del algoritmo, las cuales son importantes para la producción de sombras indirectas.
- El trabajo de Ritschel et al. [61] sobre mapas de sombra imperfectos (*imperfect shadow maps*), que evalúan mapas aproximados de sombra de muchos puntos de luz en un solo paso de la *unidad de procesamiento gráfico* (GPU), mejorando notablemente la velocidad de cómputo. Al igual que en las técnicas de radiosity, los métodos de Ritschel et al. sólo son adecuados para superficies difusas y tienen la desventaja de ser métodos iterativos, por lo que la implementación de múltiples reflexiones tienen un costo computacional importante.
- El trabajo realizado por Wang et al. [69] sobre la implementación de la técnica de mapas de fotones (*photon mapping*) en GPU, generando imágenes a velocidades cercanas al tiempo real (entre 1 y 5 imágenes por segundo).

2.6.4. Conclusiones de los trabajos relacionados

Del análisis del relevamiento realizado surge en primera instancia que, la coherencia es un concepto comúnmente utilizado como herramienta para formulación de algoritmos y estructuras de datos gráficas. Pero, cuando se trabaja con la matriz \mathbf{F} de radiosidad, no se han encontrado trabajos relevantes que relacionen la coherencia espacial a la existencia de pocos valores singulares significativos. Los resultados encontrados son basados principalmente en evidencia experimental y no están acompañados por una teoría que los explique. Además, son escasos los trabajos que utilizan algún tipo de descomposición o aproximación de matrices en los algoritmos de iluminación global. De los trabajos relevados, ninguno aprovecha la reducción de información y la simplificación de las matrices para deducir un cálculo directo de la iluminación global en escenarios con geometría fija.

Las líneas de trabajo actuales proponen el desarrollo de algoritmos de iluminación global en tiempo real que en su mayoría siguen siendo iterativos. Aquellos que no son en realidad utilizan la ejecución de procesos iterativos de un solo paso, reduciendo el realismo de la imagen calculada. Por otra parte, a pesar del componente algebraico implícito en la resolución del problema de radiosidad, pocas propuestas se basan en el uso de bibliotecas de álgebra lineal. Existen desarrollos como CUBLAS [58], donde se aprovecha la capacidad de los procesadores gráficos de realizar cálculos en paralelo, que podrían ser utilizadas por la comunidad gráfica para la resolución de sistemas lineales.

En este trabajo se propone un retorno a las líneas de investigación pioneras en radiosidad, donde se definió al problema de radiosidad como un problema de álgebra lineal y donde se combinaron métodos gráficos (como el algoritmo del hemisferio) con métodos surgidos del álgebra y de los métodos numéricos (como Gauss-Seidel) para su resolución. En este sentido, el aporte de esta propuesta se centra en la sustitución de la matriz \mathbf{RF} -una matriz densa, de dimensiones potencialmente enormes, de costosa creación y manejo- por una matriz aproximada de rango bajo. Para ello, se comienza realizando un análisis teórico para establecer si existen aproximaciones cercanas a \mathbf{RF} para rangos relativamente bajos. Luego, a partir de esta sustitución, se proponen metodologías que reducen la complejidad de los cálculos y los órdenes de memoria utilizados para la resolución del problema de radiosidad. La propuesta continúa introduciendo dos algoritmos para la generación de las matrices de rango bajo que sean cercanas a \mathbf{RF} , debido a que el análisis de los métodos existentes -cuyos resultados se presentan en el Capítulo 3- no brindó resultados satisfactorios. Se realiza un análisis experimental de las propuestas realizadas donde se confirman las hipótesis planteadas. Y se observa que las GPU posibilitan el cálculo de radiosidad en tiempo real, en escenas con geometría estática y cientos de miles de elementos.

Capítulo 3

Radiosidad de Rango Bajo

3.1. Introducción

Este capítulo introduce las técnicas de radiosidad de rango bajo (RRB), que se centran en la sustitución de la matriz \mathbf{RF} por una aproximación de rango bajo. El capítulo comienza explicando las bases teóricas de la propuesta y continúa realizando una nueva formulación del problema de radiosidad, donde la sustitución mencionada es aplicada. Luego se estudian fórmulas directas e iterativas para la resolución de la nueva formulación. En el resto del capítulo se desarrollan los conceptos de coherencia espacial y otros que son utilizados para el cálculo de nuevas aproximaciones de rango bajo a la matriz \mathbf{RF} . El capítulo culmina con el desarrollo de dos algoritmos que construyen aproximaciones de rango bajo a \mathbf{RF} .

3.2. Aproximación de rango bajo a una matriz $n \times n$

El *rango de una matriz* es la dimensión del espacio definido por sus columnas [29]. Si una matriz tiene rango k , todas sus columnas se encuentran en un espacio de dimensión k y se pueden expresar como combinación lineal de k de ellas.

Desde el punto de vista numérico, el rango de una matriz es difícil de calcular. Cuando en una matriz $n \times n$ las columnas pertenecen a un subespacio de dimensión k más una pequeña perturbación de orden ϵ , el rango de la matriz puede ser n , no reflejando que la matriz se encuentra a un ϵ o menos de ser de rango k . Para lidiar con este problema se ha desarrollado el concepto de ϵ -rango o **rango numérico** [29, 9], definido por la expresión en la Ecuación (3.1).

$$\text{rango}(\mathbf{A}, \epsilon) = \min_{\|\mathbf{A}-\mathbf{B}\|_2 \leq \epsilon} \text{rango}(\mathbf{B}) \quad (3.1)$$

Con la definición precedente, si los datos de \mathbf{A} tienen un error menor a 0,001, entonces se debe buscar $\text{rango}(\mathbf{A}, 0,001n)$ [9].

El cálculo de una aproximación de rango bajo a una matriz ha sido utilizado frecuentemente como técnica de compresión de información con pérdida. Si una matriz \mathbf{A} de dimensión $n \times n$ puede ser aproximada por el producto de dos matrices \mathbf{U}_k y \mathbf{V}_k de dimensión $n \times k$, con $n \gg k$, entonces pueden sustituirse los n^2 números de punto flotante de una matriz \mathbf{A} por los $2nk$ números de punto flotante de la aproximación, con el consiguiente ahorro de memoria.

En esta sección se desarrollarán los conceptos principales de la descomposición en valores singulares, la descomposición CUR, la descomposición basada en transformada discreta de Fourier, la basada en la transformada discreta de wavelets y las basadas en matrices ortonormales.

3.2.1. Descomposición en valores singulares (SVD)

Una de las factorizaciones más utilizadas para hallar una aproximación a una matriz \mathbf{A} expresable como el producto de dos matrices, es la **descomposición en valores singulares** (descomposición SVD). La descomposición SVD factoriza a una matriz \mathbf{A} de dimensión $m \times n$ en el producto de tres matrices $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, donde \mathbf{U} es una matriz con columnas ortonormales y de dimensión $m \times n$, \mathbf{D} es una matriz diagonal de dimensión $n \times n$ (los elementos de la diagonal son del tipo $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$), y \mathbf{V} es una matriz con columnas ortonormales y de dimensión $n \times n$ [29]. Para llevar de tres matrices a dos, alcanza con multiplicar la matriz \mathbf{D} por la matriz \mathbf{U} o la matriz \mathbf{V}^T .

Otra forma de expresar el producto matricial es

$$\mathbf{A} = \sum_{i=1}^n \sigma_i U_i V_i^T$$

donde U_i y V_i son las i -ésimas columnas de las matrices \mathbf{U} y \mathbf{V} respectivamente. Dada esta expresión y sabiendo que los σ_i están ordenados de mayor a menor y son no negativos, una aproximación de rango k de la matriz \mathbf{A} es

$$\mathbf{A} \approx \mathbf{A}_k = \sum_{i=1}^k \sigma_i U_i V_i^T \quad (3.2)$$

La proximidad entre \mathbf{A} y \mathbf{A}_k depende de qué tan grandes sean los σ_i de los términos no incluidos en la sumatoria (aquellos donde $i > k$). Si esos σ_i son nulos entonces ambas matrices son iguales entre sí. Si los σ_i no son nulos, sus valores permiten medir la norma de la diferencia $\|\mathbf{A} - \mathbf{A}_k\|$ para algunas normas. Para norma-2 y norma Frobenius las normas de la diferencia tienen los valores expresados en las Ecuaciones (3.3) y (3.4), respectivamente [66]. Con ambas normas basta conocer el último σ_k utilizado para construir \mathbf{A}_k , para acotar superiormente la norma del error.

$$\|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1} \leq \sigma_k \quad (3.3)$$

$$\|\mathbf{A} - \mathbf{A}_k\|_{Frobenius} = \sqrt{\sum_{i=k+1}^n \sigma_i^2} \leq \sigma_k \sqrt{n - k - 1} \quad (3.4)$$

Otro resultado importante sobre la matriz \mathbf{A}_k es que es la matriz de rango k que mejor aproxima a la matriz \mathbf{A} . Este teorema es llamado de *Eckart y Young* y alternativamente de *Schmidt y Mirsky* [66] y se lo suele expresar mediante la Ecuación (3.5).

$$\min_{\text{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\| = \|\mathbf{A} - \mathbf{A}_k\| \quad (3.5)$$

El resultado previo se cumple para todas las normas que son invariantes a transformaciones unitarias, en particular la norma-2 y la norma Frobenius [66]. La importancia de este resultado ha motivado la búsqueda de algoritmos eficientes para el cálculo de la descomposición SVD. La complejidad de ejecución del algoritmo básico de la descomposición SVD es $O(\min\{mn^2, nm^2\})$ [26]. Este orden elevado ha orientado la búsqueda hacia métodos que calculen soluciones aproximadas con menor complejidad.

3.2.2. Descomposición CUR

La descomposición **CUR** [20, 53] o **CGR** [33], se basa en la selección de k filas y k columnas de la matriz **A** para encontrar una factorización de rango k que se aproxime a la matriz **A**. Esta factorización es utilizada cuando es importante conocer el significado de sus componentes. En diversos campos de investigación los resultados de la descomposición SVD no resultan satisfactorios para los investigadores, quienes prefieren utilizar factorizaciones con información más comprensible, como las filas y columnas más significativas (según algún criterio) de una matriz.

La descomposición CUR se expresa como $\mathbf{A} = \mathbf{CUR}$, donde la matriz **C** posee dimensión $n \times k$ y está compuesta por k columnas de la matriz **A**, y la matriz **R** posee dimensión $k \times n$ y está compuesta por k filas de la matriz **A**. La matriz **U** posee dimensión $k \times k$ y se puede construir de diversas formas, entre las que se destacan dos que se definen a continuación: 1) **U** es la pseudoinversa de la matriz **W** (donde **W** es la intersección de las columnas **C** con las filas **R**); 2) **U** es la pseudoinversa de $\mathbf{C}^T \mathbf{A} \mathbf{R}^T$). Para seleccionar las filas y columnas de la matriz **A** que forman las matrices **C** y **R** se suele utilizar el método denominado de *volumen máximo*, desarrollado por Goreinov et al. [32, 31], que busca la combinación de filas y columnas que maximicen $|\det(\mathbf{W})|$. Existen heurísticas con complejidad $O(n)$, que hallan una buena combinación de columnas y filas sin necesidad de disponer de todas las columnas y filas de una matriz [11, 59].

3.2.3. DFT, DWT y otras factorizaciones con matrices ortonormales

La descomposición basada en la transformada de Fourier (o **DFT** por sus siglas en inglés), puede ser expresada de forma matricial si se construye una matriz **W** (véase la Ecuación (3.6)) de dimensión $n \times n$, donde la celda (k, l) contiene el complejo $e^{\frac{2\pi i}{n} kl}$.

$$\mathbf{W} = \begin{pmatrix} e^{\frac{2\pi i}{n}} & \cdots & e^{\frac{2\pi i l}{n}} & \cdots & e^{\frac{2\pi i n}{n}} \\ \vdots & & \vdots & & \vdots \\ e^{\frac{2\pi i k}{n}} & \cdots & e^{\frac{2\pi i k l}{n}} & \cdots & e^{\frac{2\pi i k n}{n}} \\ \vdots & & \vdots & & \vdots \\ e^{\frac{2\pi i n}{n}} & \cdots & e^{\frac{2\pi i n l}{n}} & \cdots & e^{\frac{2\pi i n^2}{n}} \end{pmatrix} \quad (3.6)$$

Aplicando la matriz **W** a una matriz cualquiera **A**, el equivalente a la *transformada de Fourier en dos dimensiones* se resuelve realizando el producto \mathbf{WAW}^T . El resultado del producto es una matriz de dimensión $n \times n$, que si tiene un conjunto pequeño de elementos con valores relativamente grandes, puede ser sustituida por una matriz dispersa que sólo contenga dichos elementos.

Cuando se tiene una *matriz ortonormal* **N** (que posee columnas ortogonales cuya norma-2 vale uno), se pueden aplicar los mismos razonamientos utilizados en la descomposición SVD para hallar una aproximación de rango bajo a una matriz **A**. Como se viera en la descomposición SVD, una matriz **A** se factoriza como el producto de tres matrices \mathbf{UDV}^T , donde **U** y **V** son matrices ortonormales. En particular si se realiza el producto \mathbf{AV} se obtiene por resultado a **UD** debido a la ortonormalidad de **V**. Luego es posible seleccionar las k columnas con mayor norma \mathbf{AV} y construir la misma matriz \mathbf{A}_k expresada en la Ecuación (3.2). Si en lugar de utilizar la matriz **V** se utiliza una matriz ortonormal **N** y se hallan las k mayores columnas de \mathbf{AN} , también se puede hallar $\mathbf{A}_k^{(\mathbf{N})}$ que es la mejor aproximación de rango k de la matriz **A** utilizando **N**, como

se observa en la Ecuación (3.7), donde $(\mathbf{AN})_j$ y \mathbf{N}_j representan a las j -ésimas columnas de las matrices \mathbf{AN} y \mathbf{N} respectivamente.

$$\mathbf{A} = \sum_{j=1}^n (\mathbf{AN})_j \mathbf{N}_j^T \approx \sum_{\substack{k \ll n \\ j=1 \\ \|(\mathbf{AN})_j\| \geq \|(\mathbf{AN})_{j+1}\|, \forall j \in \{1 \dots n-1\}}} (\mathbf{AN})_j \mathbf{N}_j^T = \mathbf{A}_k^{(\mathbf{N})} \quad (3.7)$$

Un caso particular de matriz ortonormal es la matriz $\mathbf{N} = \mathbf{W}/\sqrt{n}$, donde \mathbf{W} es la matriz definida en la Ecuación (3.6). La ventaja de utilizar la transformada DFT radica en que las operaciones \mathbf{AW}/\sqrt{n} y \mathbf{WAW}^T pueden realizarse con un complejidad y $O(n^2 \log n)$. Este resultado surge de los trabajos pioneros de Danielson y Lanczos [60], quienes descubrieron la técnica denominada *Fast Fourier Transformation*.

Los razonamientos anteriores basados en DFT, también se pueden realizar con la transformada discreta de wavelets (o **DWT** por sus siglas en inglés), con la que se pueden construir una variedad infinita de matrices \mathbf{W} y donde los productos matriciales también pueden ser llevados a $O(n^2 \log n)$. Para la construcción de \mathbf{W} , en lugar de utilizar $e^{\frac{2\pi i}{n} kl}$ se utiliza una familia de funciones llamadas *wavelets*, que poseen una variedad mucho más amplia de propiedades y características que las surgidas de Fourier. Con la transformada DWT también pueden realizarse cálculos con $O(n^2 \log n)$. Las wavelets están presentadas en los trabajos de Daubechies [19] y de Press et al. [60].

Como resumen, se puede establecer que una factorización basada en DFT o DWT tiene una complejidad menor a aquella basada en la descomposición SVD o en un matriz ortonormal cualquiera. Por otra parte, un inconveniente de la utilización de una matriz ortonormal genérica como las que surgen de DFT o DWT radica en que dichas matrices no fueron generadas para factorizar específicamente a la matriz A (a diferencia de la descomposición SVD donde \mathbf{V} varía según la matriz). Por lo tanto, nada asegura que estas nuevas factorizaciones puedan encontrar buenas aproximaciones con rango bajo ($k \ll n$).

3.2.4. Notación para las matrices de rango bajo

Si bien existen aproximaciones de rango bajo a una matriz que se expresan como el producto de tres matrices -donde la matriz interior del producto suele ser más pequeña que las otras- a lo largo de la tesis se utiliza la fórmula

$$\mathbf{A} \approx \mathbf{U}_k \mathbf{V}_k^T \quad (3.8)$$

donde \mathbf{A} es una matriz cuadrada de dimensión $n \times n$ (como la matriz de radiosidad) y \mathbf{U}_k y \mathbf{V}_k son matrices de dimensión $n \times k$, con $k \ll n$. Las matrices surgidas de las descomposiciones SVD, CUR, DFT, DWT y las matrices ortonormales, pueden llevarse a un producto de dos matrices como las mencionadas.

3.2.5. Conclusiones de las aproximaciones relevadas

Salvo la descomposición CUR, todas las factorizaciones mostradas en esta sección (SVD, DFT, DWT y cualquiera basada en el uso de matrices ortonormales genéricas) parten de la hipótesis de que se dispone por completo de la matriz a factorizar. Esto resulta en un inconveniente cuando la matriz contiene una gran cantidad de números de punto flotante. La descomposición CUR surge como la opción más prometedora entre las evaluadas.

En la sección 3.6 se presenta el desarrollo de una metodología novedosa que permiten el cálculo de aproximaciones de las matrices \mathbf{F} y \mathbf{RF} que evitan su almacenamiento previo en memoria. Esta nueva aproximación combina propiedades de la descomposición CUR y de las descomposiciones basadas en matrices ortonormales (SVD, DFT, DWT, etc.). La matriz \mathbf{U}_k contiene columnas muy cercanas a las columnas de la matriz original (en CUR las columnas de la matriz \mathbf{U}_k son iguales a las de la matriz a aproximar) y la matriz \mathbf{V}_k es ortonormal (como en SVD, DFT, DWT) y dispersa, posibilitando un mejor uso de memoria que el de las otras factorizaciones.

Por último, resulta importante destacar que al sustituir el producto matricial \mathbf{RF} por alguna factorización $\mathbf{U}_k\mathbf{V}_k^T$ en la ecuación matricial de radiosidad, el objetivo final es la reducción de la memoria necesaria para operar con la matriz del sistema lineal. Este ahorro de memoria puede ser superfluo si el cálculo del sistema lineal sigue requiriendo una memoria temporal de n^2 números de punto flotante (como ocurriría al utilizar la eliminación gaussiana), pero es relevante cuando se utilizan métodos de cálculo más eficientes. También resulta importante conocer la relación entre el error relativo de \tilde{B} (valor aproximado de B) y la distancia $\|\mathbf{RF} - \mathbf{U}_k\mathbf{V}_k^T\|$, y si es posible acelerar el tiempo de cálculo \tilde{B} con respecto al tiempo de cálculo de B . Estos puntos se verán en las secciones siguientes.

3.3. Nueva formulación del problema de radiosidad

En esta sección se establecen las bases de la utilización de matrices de rango bajo para la resolución de la ecuación de radiosidad. Comienza introduciendo el concepto de coherencia espacial en el problema de radiosidad, llegando a la conclusión que la matriz \mathbf{RF} suele tener un rango numérico bajo y culmina mostrando una nueva formulación del problema de radiosidad, en donde la matriz \mathbf{RF} es sustituida por otra de rango bajo.

3.3.1. Coherencia espacial y su influencia en el rango numérico de \mathbf{RF}

La **coherencia espacial** es el grado en que las partes de una escena exhiben similitudes locales. Esta propiedad es utilizada por Sutherland et al. [67] para la determinación de superficies ocultas, y también puede ser utilizada en la radiosidad de una escena con superficies lambertianas. La existencia de coherencia espacial en la radiosidad de una escena presupone que para buena parte de las superficies los valores de radiosidad entre puntos cercanos son también cercanos. Si se cumple la coherencia espacial en radiosidad, no existen marcadas diferencias (o discontinuidades) de los valores de radiosidad entre puntos cercanos de una misma superficie. Esta es una propiedad clave para el desarrollo de las técnicas de RRB, ya que, complementada con técnicas de interpolación, posibilita el desarrollo de algoritmos que realicen ahorros significativos en el almacenamiento información y en la complejidad de los cálculos de radiosidad.

Si en lugar de considerar puntos cercanos se consideran parches cercanos, también se aplica el concepto de coherencia espacial: en dos parches cercanos i_1 e i_2 la irradiancia recibida del resto de la escena es similar y la radiosidad reflejada hacia la escena se distribuye de forma similar. Esta propiedad garantiza que en la matriz de factores de forma \mathbf{F} los elementos de las filas asociadas a los parches i_1 e i_2 cumplen que, $\mathbf{F}_{i_1j} \approx \mathbf{F}_{i_2j} \forall j$. Las filas se consideran *parecidas* si se cumple esta aproximación, verificándose que $\|\mathbf{F}_{i_1} - \mathbf{F}_{i_2}\| \approx 0$.

En la Figura 3.1 se observa cómo dos parches cercanos i_1 e i_2 “ven” prácticamente la misma escena (las vistas se representan a través de dos hemicubos). Los parches ven aproximadamente los mismos otros parches y cualquier parche j ocupa aproximadamente la misma área en ambos hemicubos. Con estas áreas se determinan los factores de forma \mathbf{F}_{i_1j} y \mathbf{F}_{i_2j} .

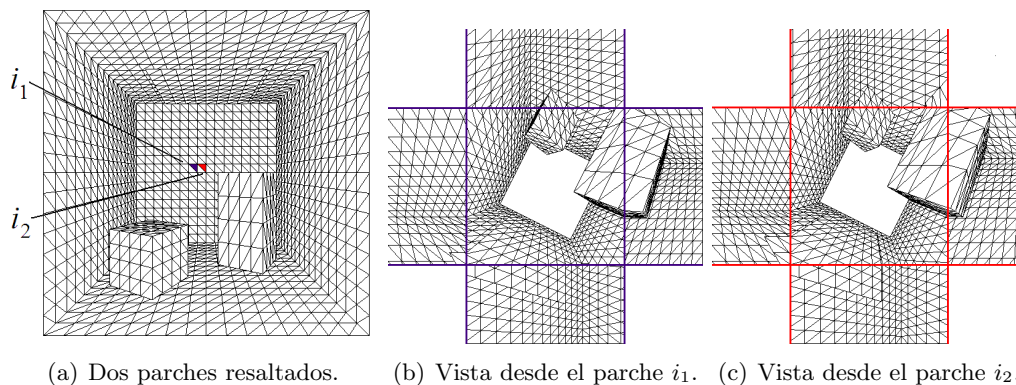


Figura 3.1: Dos parches resaltados en (a) y la escena vista desde ellos ((b) y (c)).

El fenómeno presentado previamente puede duplicarse en los miles de pares de parches cercanos que hay en una escena. Tener miles de filas parecidas trae como consecuencia la reducción del rango numérico de \mathbf{F} (y de \mathbf{RF}).

Experimentalmente se observa que en general las matrices \mathbf{RF} asociadas a las escenas tienen un rango numérico bajo: la mayoría de sus valores singulares son pequeños y sólo algunos son relativamente grandes (véase la Figura 3.2 (a)). Hašan et al. [40] obtuvieron resultados similares para las matrices de *unidades de superficies* \times *luces*.

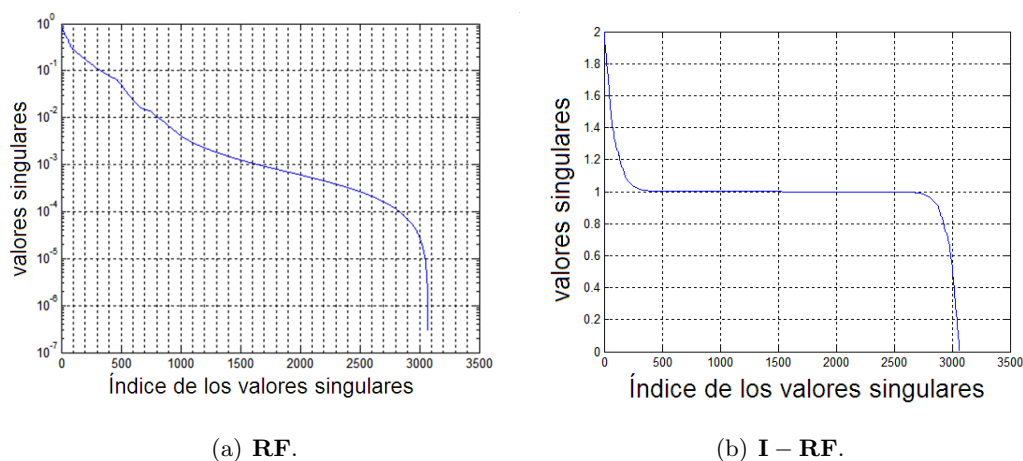


Figura 3.2: Descomposiciones SVD de \mathbf{RF} y de $\mathbf{I} - \mathbf{RF}$.

El rango numérico bajo de \mathbf{RF} no afecta a la resolución del sistema lineal. El rango numérico de una matriz está fuertemente asociado a su número de condición, y por tanto, a las dificultades para resolver un sistema lineal que la contenga. En la ecuación matricial de radiosidad, la matriz del sistema es $(\mathbf{I} - \mathbf{RF})$ y esta matriz no posee un rango numérico bajo debido a que la mayoría de sus valores singulares están cercanos a uno (véase la Figura 3.2 (b)).

Al existir muchas filas y columnas parecidas en \mathbf{RF} , su rango numérico es bajo y por lo tanto puede aproximarse a una matriz de rango bajo, que puede construirse como el producto $\mathbf{U}_k \mathbf{V}_k^T$. Las técnicas RRB tienen utilidad para aquellas escenas en que se logran buenas aproximaciones a \mathbf{RF} con valores de $2k$ notoriamente inferiores a n .

3.3.2. Ecuación de radiosidad de rango bajo

Como se expresó en la sección anterior, es probable que la matriz \mathbf{RF} tenga un rango numérico bajo, por lo que es posible hallar una buena aproximación de rango bajo $\mathbf{U}_k \mathbf{V}_k^T$. Al sustituir \mathbf{RF} por su aproximación de rango bajo en la ecuación matricial de radiosidad se obtiene la **ecuación de radiosidad de rango bajo** (ecuación RRB, presentada en (3.9)), donde la incógnita ya no es B sino una aproximación \tilde{B} .

$$(\mathbf{I} - \mathbf{U}_k \mathbf{V}_k^T) \tilde{B} = E \quad (3.9)$$

Esta nueva formulación del problema de radiosidad plantea un ahorro importante de memoria, debido a que la matriz densa y que ocupa una memoria de n^2 números de punto flotante es sustituida por dos matrices que en total ocupan una memoria de $2nk$ números de punto flotante. Esta supuesta ventaja puede no ser tal si no se dispone de un método eficaz para resolver la ecuación RRB. Como se mencionara, si para resolver la ecuación RRB se utiliza el algoritmo de eliminación gaussiana, es necesario realizar el producto $\mathbf{U}_k \mathbf{V}_k^T$ para generar la matriz de entrada del algoritmo. Esto implica que temporalmente se necesite utilizar memoria para almacenar n^2 números de punto flotante y que el costo computacional total sea de $O(n^3)$. Como consecuencia, se requieren métodos eficientes que saquen provecho de las nuevas estructuras matriciales para reducir la memoria utilizada y reducir los tiempos de resolución.

Otro tema importante a evaluar es la estimación del error que surge de sustituir la ecuación matricial de radiosidad por la ecuación RRB. Los valores de \tilde{B} pueden estar muy alejados de B , a pesar que la aproximación de $\mathbf{U}_k \mathbf{V}_k^T$ a \mathbf{RF} sea muy buena. Ambos temas son tratados a continuación.

3.4. Resolución de la ecuación de radiosidad de rango bajo

En esta sección se plantean dos metodologías para la resolución de la ecuación RRB: una metodología iterativa y otra directa. La metodología iterativa está basada en el uso de la serie de Neumann, tradicionalmente utilizada en el problema de radiosidad [15]. La metodología directa está basada en la fórmula de Sherman-Morrison-Woodbury, aspecto novedoso para el problema de radiosidad. Luego de calculada la solución, se realiza la estimación del error respecto a la solución de la ecuación matricial de radiosidad. A continuación se presentan dos extensiones a la resolución de la ecuación de radiosidad de rango bajo: se extiende la resolución al caso de la radiosidad en colores y al caso en que la geometría de la escena varía. Por último, se plantean algunas conclusiones de la radiosidad de rango bajo.

3.4.1. Resolución iterativa de \tilde{B}

La *serie de Neumann* es una generalización de la serie geométrica que es utilizada para el cálculo de la matriz $(\mathbf{I} - \mathbf{T})^{-1}$ [66]. Su expresión general es

$$(\mathbf{I} - \mathbf{T})^{-1} = \mathbf{I} + \mathbf{T} + \mathbf{T}^2 + \dots \quad (3.10)$$

Una condición suficiente de convergencia de la serie de Neumann es que el radio espectral de \mathbf{T} sea menor que uno. Esta condición es también suficiente para asegurar que $(\mathbf{I} - \mathbf{T})$ sea una matriz no singular [66].

La serie de Neumann puede ser utilizada de forma directa para la resolución de la ecuación RRB. A través de esta serie, la inversa de la matriz de la ecuación RRB se descompone en una sumatoria de matrices que se multiplican por el vector E para hallar el vector de radiosidad \tilde{B} , cuya expresión se presenta en la Ecuación (3.11).

$$\tilde{B} = (\mathbf{I} - \mathbf{U}_k \mathbf{V}_k^T)^{-1} E = \left(\mathbf{I} + \mathbf{U}_k \mathbf{V}_k^T + (\mathbf{U}_k \mathbf{V}_k^T)^2 + \dots \right) E \quad (3.11)$$

La serie de la Ecuación (3.11) se expresa de forma iterativa mediante la fórmula

$$\tilde{B}_{n+1} = E + \mathbf{U}_k \left(\mathbf{V}_k^T \tilde{B}_n \right) \text{ con } \tilde{B}_0 = E \quad (3.12)$$

ya que

$$\begin{aligned} \tilde{B}_1 &= (\mathbf{I} + \mathbf{U}_k \mathbf{V}_k^T) E \\ \tilde{B}_2 &= \left(\mathbf{I} + \mathbf{U}_k \mathbf{V}_k^T + (\mathbf{U}_k \mathbf{V}_k^T)^2 \right) E \\ &\vdots \\ \tilde{B}_\infty &= \left(\mathbf{I} + \mathbf{U}_k \mathbf{V}_k^T + (\mathbf{U}_k \mathbf{V}_k^T)^2 + \dots \right) E \end{aligned}$$

Los paréntesis en la Ecuación (3.12) indican la precedencia de las operaciones. De esta manera se evita la realización del producto matriz-matriz $\mathbf{U}_k \mathbf{V}_k^T$ (que genera una matriz de dimensión $n \times n$) y sólo hay productos del tipo matriz-vector. La complejidad de cada iteración es $O(nk)$.

La serie de la Ecuación (3.11) converge siempre que el radio espectral de $\mathbf{U}_k \mathbf{V}_k^T$ es menor que uno. Esta condición no está asegurada aunque \mathbf{RF} cumpla que su radio espectral es menor que uno. Si se impone la condición más estricta de que $\|\mathbf{RF}\|_2$ sea menor que uno se asegura también que $\|\mathbf{U}_k \mathbf{V}_k^T\|_2$ sea menor que uno, cuando las matrices se hallan con la descomposición SVD de \mathbf{RF} . De esta forma, ambos productos de matrices comparten el mismo máximo valor singular, que es el valor de la norma-2 [66] y que por lo tanto es menor que uno.

3.4.2. Resolución directa de \tilde{B}

La resolución iterativa de la ecuación RRB logra buenos ordenes de complejidad y de uso de memoria, pero no es la única alternativa de resolución. A partir de la ecuación RRB se deduce la expresión de la Ecuación (3.13).

$$\tilde{B} = (\mathbf{I} - \mathbf{U}_k \mathbf{V}_k^T)^{-1} E \quad (3.13)$$

A continuación se desarrolla una serie de razonamientos que concluyen en una expresión que permite calcular \tilde{B} de forma directa y con costos computacionales más bajos que los que implica el cálculo de la inversa de la matriz en la Ecuación (3.13). La expresión resultante se puede deducir a través de la fórmula de Sherman-Morrison-Woodbury [29], o a través de la serie de Neumann. La fórmula de Sherman-Morrison-Woodbury es una fórmula general que se debe simplificar para utilizarse en el problema de radiosidad. La serie de Neumann permite deducir la solución de la ecuación RRB de una forma más intuitiva. A continuación se desarrollan ambos razonamientos.

Fórmula de Sherman-Morrison-Woodbury en la resolución directa de \tilde{B}

La fórmula de *Sherman-Morrison-Woodbury* (SMW) se utiliza para el cálculo incremental de $(\mathbf{A} + \mathbf{U}\mathbf{V}^T)^{-1}$, cuando se conoce \mathbf{A}^{-1} y las matrices \mathbf{U} y \mathbf{V} tienen dimensión $n \times k$, por lo tanto su producto $\mathbf{U}\mathbf{V}^T$ es una matriz de rango máximo k .

La fórmula de SMW expresa que el cálculo de $(\mathbf{A} + \mathbf{U}\mathbf{V}^T)^{-1}$ se puede realizar con las operaciones que se presentan en la Ecuación (3.14) donde \mathbf{I}_k es una matriz identidad de dimensión $k \times k$.

$$(\mathbf{A} + \mathbf{U}\mathbf{V}^T)^{-1} = \mathbf{A}^{-1} - \left[\mathbf{A}^{-1}\mathbf{U} (\mathbf{I}_k + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1} \mathbf{V}^T \mathbf{A}^{-1} \right] \quad (3.14)$$

En la Ecuación (3.14), el término a la derecha de la igualdad es más simple de calcular que el término de la izquierda. La inversa de la izquierda tiene una complejidad $O(n^3)$ y el término de la derecha tiene una complejidad $O(n^2k)$.

La fórmula SMW puede aplicarse al cálculo de \tilde{B} en la Ecuación (3.13). Si en la Ecuación (3.14) la matriz \mathbf{A} se sustituye por \mathbf{I} y la matriz $\mathbf{U}\mathbf{V}^T$ se sustituye por $\mathbf{U}_k\mathbf{V}_k^T$ la Ecuación se simplifica, resultando la expresión de la Ecuación (3.15).

$$(\mathbf{I} - \mathbf{U}_k\mathbf{V}_k^T)^{-1} = \mathbf{I} + \left[\mathbf{U}_k (\mathbf{I}_k - \mathbf{V}_k^T\mathbf{U}_k)^{-1} \mathbf{V}_k^T \right] \quad (3.15)$$

La Ecuación (3.15) tiene una expresión aún más simple, que se presenta en la Ecuación (3.16), donde la matriz \mathbf{Y}_k tiene dimensión $n \times k$.

$$(\mathbf{I} - \mathbf{U}_k\mathbf{V}_k^T)^{-1} = \mathbf{I} - \mathbf{Y}_k\mathbf{V}_k^T, \text{ con } \mathbf{Y}_k = -\mathbf{U}_k (\mathbf{I}_k - \mathbf{V}_k^T\mathbf{U}_k)^{-1} \quad (3.16)$$

Serie de Neumann en la resolución directa de \tilde{B}

Si se aplica la serie de Neumann al cálculo de $(\mathbf{I} - \mathbf{U}_k\mathbf{V}_k^T)^{-1}$ se obtiene la serie de la Ecuación (3.17).

$$(\mathbf{I} - \mathbf{U}_k\mathbf{V}_k^T)^{-1} = \mathbf{I} + (\mathbf{U}_k\mathbf{V}_k^T) + (\mathbf{U}_k\mathbf{V}_k^T)^2 + \dots = \sum_{i=0}^{\infty} (\mathbf{U}_k\mathbf{V}_k^T)^i \quad (3.17)$$

La serie de la Ecuación (3.17) también se puede expresar mediante

$$\mathbf{I} + \mathbf{U}_k \left(\mathbf{I}_k + (\mathbf{V}_k^T\mathbf{U}_k) + (\mathbf{V}_k^T\mathbf{U}_k)^2 + \dots \right) \mathbf{V}_k^T = \mathbf{I} + \mathbf{U}_k \left(\sum_{i=0}^{\infty} (\mathbf{V}_k^T\mathbf{U}_k)^i \right) \mathbf{V}_k^T$$

donde \mathbf{I}_k es la matriz identidad de dimensión k y donde se puede aplicar nuevamente Neumann para eliminar la serie, resultando en la expresión de la Ecuación (3.18).

$$\mathbf{I} + \mathbf{U}_k \left(\sum_{i=0}^{\infty} (\mathbf{V}_k^T\mathbf{U}_k)^i \right) \mathbf{V}_k^T = \mathbf{I} + \mathbf{U}_k (\mathbf{I}_k - \mathbf{V}_k^T\mathbf{U}_k)^{-1} \mathbf{V}_k^T \quad (3.18)$$

Por lo tanto, a través de este razonamiento se llega a la igualdad presentada en la Ecuación (3.19), idéntica a la Ecuación (3.15).

$$(\mathbf{I} - \mathbf{U}_k \mathbf{V}_k^T)^{-1} = \mathbf{I} + \mathbf{U}_k (\mathbf{I}_k - \mathbf{V}_k^T \mathbf{U}_k)^{-1} \mathbf{V}_k^T \quad (3.19)$$

Esta igualdad se cumple independientemente de si las series intermedias utilizadas en la deducción convergen o no. La única restricción que debe cumplirse es que la matriz $(\mathbf{I}_k - \mathbf{V}_k^T \mathbf{U}_k)$ sea no singular [29].

Cálculo de \tilde{B}

Una vez hallada la Ecuación (3.19) puede aplicarse en la Ecuación (3.13) para simplificar el cálculo de \tilde{B} , obteniéndose la expresión de la Ecuación (3.20).

$$\tilde{B} = \left(\mathbf{I} + \mathbf{U}_k (\mathbf{I} - \mathbf{V}_k^T \mathbf{U}_k)^{-1} \mathbf{V}_k^T \right) E \quad (3.20)$$

Para un cálculo eficiente de \tilde{B} conviene operar con los paréntesis de la Ecuación (3.20), de lo contrario el costo se puede elevar a un uso temporal de memoria $O(n^2)$ y una complejidad $O(n^2k)$. Con la representación alternativa de la Ecuación (3.20) presentada en la Ecuación (3.21), el cálculo de \tilde{B} tiene una complejidad $O(nk^2)$ y precisa una memoria $O(nk)$.

$$\tilde{B} = E + \mathbf{U}_k \left((\mathbf{I} - \mathbf{V}_k^T \mathbf{U}_k)^{-1} (\mathbf{V}_k^T E) \right) \quad (3.21)$$

Cuando en una escena sólo varía la iluminación, en la ecuación de radiosidad sólo varía el término independiente E , por lo tanto la matriz \mathbf{Y}_k en la Ecuación (3.16) permanece constante y puede calcularse una única vez. En consecuencia, el cálculo de \tilde{B} se simplifica aun más, resultando en la **Ecuación RRB con geometría fija** (3.22) donde todas las operaciones incluidas en el cálculo de \mathbf{Y}_k se realizan una sola vez, obteniéndose una complejidad final $O(nk)$ y un consumo de memoria $O(nk)$.

$$\tilde{B} = E - \mathbf{Y}_k (\mathbf{V}_k^T E) \quad (3.22)$$

La expresión de la ecuación RRB con geometría fija resulta muy similar a la expresión de la Ecuación (3.12), utilizada para un paso de la iteración. Ambas expresiones tienen una complejidad similar (realizan dos productos matriz-vector de idénticas dimensiones y luego una suma vectorial). Cuando la geometría es constante conviene utilizar el método directo, debido a que tiene la misma complejidad que una única iteración del método iterativo. En cambio, cuando la geometría es variable, una vez conocido \mathbf{U}_k y \mathbf{V}_k el método directo puede no ser el más conveniente si el método iterativo converge en pocos pasos, debido a las $O(nk^2)$ operaciones que se requiere para calcular \mathbf{Y}_k . Debe estudiarse en cada caso cuál de los métodos es el más conveniente. Véase la sección 3.4.5.

3.4.3. Estimación del error relativo de \tilde{B}

La ecuación RRB se puede expresar como una versión de la ecuación matricial de radiosidad, pero con la matriz perturbada. En efecto, si ϵ_k es la perturbación resultante de sustituir $\mathbf{R}\mathbf{F}$ por $\mathbf{U}_k \mathbf{V}_k^T$, entonces $\epsilon_k = \mathbf{R}\mathbf{F} - \mathbf{U}_k \mathbf{V}_k^T$ y por lo tanto

$$E = (\mathbf{I} - \mathbf{R}\mathbf{F} + \epsilon_k) \tilde{B} = (\mathbf{I} - \mathbf{R}\mathbf{F} + (\mathbf{R}\mathbf{F} - \mathbf{U}_k \mathbf{V}_k^T)) \tilde{B} = (\mathbf{I} - \mathbf{U}_k \mathbf{V}_k^T) \tilde{B} \quad (3.23)$$

Según Golub y Van Loan (pags. 80 y 81 de [29]), en un sistema $\mathbf{A}x = b$ perturbado de forma paramétrica por δ tanto en su matriz como en su término independiente se cumple que

$$(\mathbf{A} + \delta\mathbf{G})x(\delta) = b + \delta f, \text{ con } x(0) = x$$

donde \mathbf{G} es una matriz de igual dimensión que \mathbf{A} y f es un vector de igual dimensión que b . La incógnita x es sensible a δ según la cota del error relativo de x , presentada en la Ecuación (3.24).

$$\frac{\|x(\delta) - x\|}{\|x\|} \leq |\delta| \|\mathbf{A}^{-1}\| \left(\frac{\|f\|}{\|x\|} + \|\mathbf{G}\| \right) + O(\delta^2) \quad (3.24)$$

Aplicando la expresión anterior a la Ecuación (3.23), se tiene que $\mathbf{A} = \mathbf{I} - \mathbf{R}\mathbf{F}$, $\delta = 1$, $\mathbf{G} = \epsilon_k$ y $f = 0$, resultando la expresión de la Ecuación (3.25).

$$\frac{\|\tilde{B} - B\|}{\|B\|} \leq \|(\mathbf{I} - \mathbf{R}\mathbf{F})^{-1}\| \|\epsilon_k\| \quad (3.25)$$

A su vez, como $\|\mathbf{R}\mathbf{F}\|$ es menor que uno, se puede acotar $\|(\mathbf{I} - \mathbf{R}\mathbf{F})^{-1}\|$ (véase el lema 2.3.3 en Golub y Van Loan [29]), dando por resultado la cota del error relativo de \tilde{B} , expresado en la Ecuación (3.26).

$$\frac{\|\tilde{B} - B\|}{\|B\|} \leq \frac{\|\epsilon_k\|}{1 - \|\mathbf{R}\mathbf{F}\|} \quad (3.26)$$

En las Ecuaciones (3.25) y (3.26), $\|\epsilon_k\|$ es la única variable que se puede ajustar con libertad para reducir la cota del error. La matriz $\mathbf{R}\mathbf{F}$ se deduce de los datos y es de esperar que su norma no varíe demasiado, independientemente de cómo haya sido construida. Si se utiliza norma-2, el menor valor posible de $\|\epsilon_k\|_2$ para un valor dado de k es el valor singular σ_{k+1} de $\mathbf{R}\mathbf{F}$. Este valor singular se alcanza sólo si \mathbf{U}_k y \mathbf{V}_k son las correspondientes matrices asociadas a la descomposición SVD.

Un caso más general, utilizado en el método de radiosidad de rango bajo, es aquel en que se emplea una matriz ortonormal \mathbf{M} diferente de \mathbf{V} . La matriz \mathbf{U} asociada no tiene por qué ser ortonormal y se calcula como $\mathbf{U} = \mathbf{R}\mathbf{F}\mathbf{M}$ debido a que $\mathbf{U}\mathbf{M}^T = \mathbf{R}\mathbf{F}\mathbf{M}\mathbf{M}^T = \mathbf{R}\mathbf{F}$. La matriz \mathbf{U}_k está compuesta por las k columnas con mayor norma-2 de la matriz \mathbf{U} . Para este caso el valor de $\|\epsilon_k\|$ está acotado por

$$\begin{aligned} \|\epsilon_k\| &= \|(\mathbf{I} - \mathbf{R}\mathbf{F}) - (\mathbf{I} - \mathbf{U}_k\mathbf{T}_k^T)\| \\ &= \|\mathbf{U}\mathbf{T}^T - \mathbf{U}_k\mathbf{T}_k^T\| \\ &= \left\| \sum_{j=1}^n U_j\mathbf{T}_j^T - \sum_{j=1}^k U_j\mathbf{T}_j^T \right\| \Rightarrow \\ \|\epsilon_k\| &= \left\| \sum_{\substack{j=k+1 \\ \|U_j\| \geq \|U_{j+1}\|}}^n U_j\mathbf{T}_j^T \right\| \leq \sum_{\substack{j=k+1 \\ \|U_j\| \geq \|U_{j+1}\|}}^n \|U_j\| \end{aligned} \quad (3.27)$$

Consideraciones finales sobre el error

Las cotas de error expresadas en esta sección pueden ser muy elevadas respecto a los valores experimentales hallados en las escenas. No es extraño que la estimación del error sea de un orden de magnitud superior al error real [22].

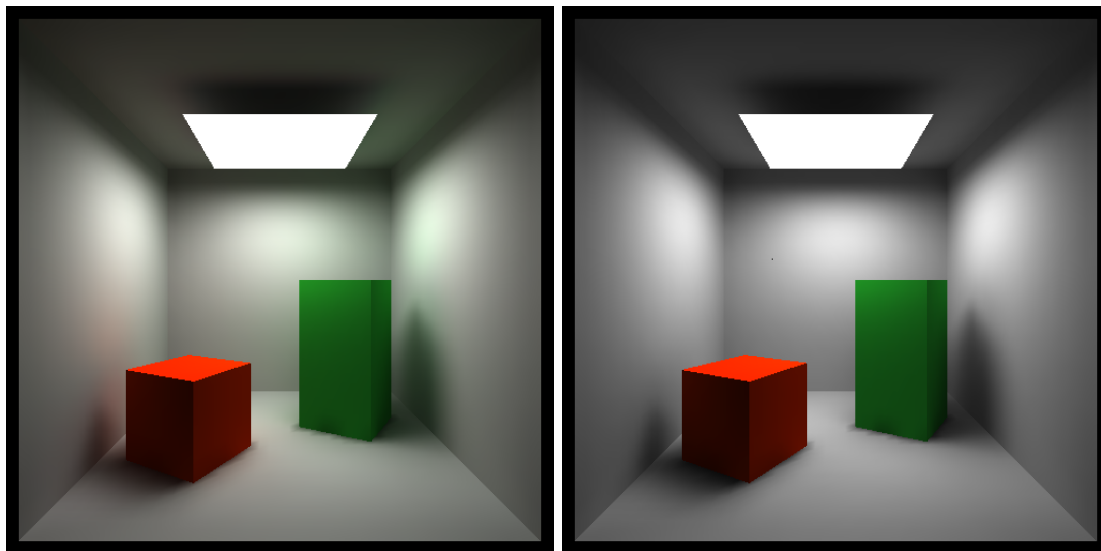
Una de las causas probables de este fenómeno es que en ninguna parte de los razonamientos anteriores se considera la restricción de que E y \tilde{B} no deben poseer valores negativos (dado que físicamente no existe la radiosidad negativa). Si se considera dicha restricción, las cotas superiores podrían tener valores menores. En efecto, en los casos testeados en experimentos no formalizados, las cotas superiores son alcanzadas para vectores E y B que contienen valores negativos.

Otra de las posibles causas de la diferencia entre la teoría y la práctica es que los vectores E que provocan la mayor diferencia entre B y \tilde{B} poseen configuraciones de parches emisores muy particulares, que no suelen darse en la práctica.

En conclusión, hallar buenas cotas de error para los métodos RRB es una tarea que requiere de mayores esfuerzos de investigación. Esta sección ha presentado tan solo una introducción a la temática.

3.4.4. Radiosidad en colores

Cuando las superficies emisoras y reflectoras de la escena están coloreadas, las soluciones realistas de la ecuación de radiosidad y de la ecuación matricial de radiosidad deben incluir el fenómeno de “sangrado” (*bleeding*), en donde la reflexión de la luz en un objeto coloreado “tiñe” del mismo color a los objetos cercanos. Por ejemplo, en la Figura 3.3 (a) se aprecia que la pared cercana al cubo rojo está también teñida de rojo. La Figura 3.3 (b) ha sido generada sin incluir el efecto de sangrado.



(a) Con sangrado.

(b) Sin sangrado.

Figura 3.3: Una misma escena con dos implementaciones de radiosidad en colores.

Para generar el efecto de sangrado se deben resolver tres sistemas lineales distintos, uno por cada color primario (*rojo, verde y azul*). En los tres sistemas lineales se comparte la matriz \mathbf{F} , que depende únicamente de la geometría de la escena, y se difiere en los otros componentes de la ecuación matricial de radiosidad (\mathbf{R} , E y B). Como se expresó en la sección 2.4.2, la matriz diagonal \mathbf{R} contiene el índice de reflexión de cada parche. A un parche i de color rojo puro, le corresponde el color $(1, 0, 0)$, y por lo tanto las matrices \mathbf{R} asociadas con cada color contienen en la diagonal los valores: $\mathbf{R}_r(i, i) = 1$, $\mathbf{R}_v(i, i) = 0$ y $\mathbf{R}_a(i, i) = 0$. Similares razonamientos se pueden realizar con el vector de emisión E , que se transforma en los vectores E_r, E_v y E_a , y con el vector de radiosidad B , que se transforma en los vectores B_r, B_v y B_a . Por lo tanto, la ecuación matricial de radiosidad se transforma en tres sistemas de ecuaciones independientes:

$$\begin{aligned} (\mathbf{I} - \mathbf{R}_r \mathbf{F}) B_r &= E_r \\ (\mathbf{I} - \mathbf{R}_v \mathbf{F}) B_v &= E_v \\ (\mathbf{I} - \mathbf{R}_a \mathbf{F}) B_a &= E_a \end{aligned}$$

Para simplificar, se unifican las tres ecuaciones con subíndices r, v y a en una sola con subíndice c (que significa “color”), y por lo tanto cada ecuación con subíndice c representa en realidad a tres ecuaciones, una por cada color.

$$(\mathbf{I} - \mathbf{R}_c \mathbf{F}) B_c = E_c \quad (3.28)$$

Siguiendo esta convención, la ecuación RRB se transforma en

$$(\mathbf{I} - \mathbf{R}_c \mathbf{U}_k \mathbf{V}_k^T) \tilde{B}_c = E_c \quad (3.29)$$

Para evitar el uso innecesario de memoria, conviene factorizar la matriz \mathbf{F} en lugar de las matrices $\mathbf{R}\mathbf{F}$ como ocurre en la ecuación RRB. Las matrices \mathbf{R}_c son diagonales, ocupando en memoria $3n$ números de punto flotante, en cambio las matrices $\mathbf{U}_{ck} = \mathbf{R}_c \mathbf{U}_k$ son densas, ocupando en total $3nk$ números de punto flotante.

Resolución iterativa y directa de la radiosidad en colores

Como consecuencia de mantener las matrices \mathbf{R}_c , deben reformularse las Ecuaciones (3.12), (3.20) y (3.21) para la resolución iterativa de la ecuación RRB y el cálculo de \tilde{B} , y la ecuación RRB con geometría fija para la resolución directa de la ecuación RRB.

Para el caso iterativo, la ecuación RRB se sustituye por las ecuaciones

$$\tilde{B}_{c,n+1} = E_c + \mathbf{R}_c \left(\mathbf{U}_k \left(\mathbf{V}_k^T \tilde{B}_{c,n} \right) \right) \text{ con } \tilde{B}_{c,0} = E_c \quad (3.30)$$

Para el caso directo, las Ecuaciones (3.20), (3.21) y la ecuación RRB con geometría fija se sustituyen respectivamente por las Ecuaciones (3.31), (3.32) y (3.33), donde $\mathbf{M}_c = (\mathbf{I} - \mathbf{V}_k^T \mathbf{R}_c \mathbf{U}_k)^{-1}$ son matrices $k \times k$, y las matrices \mathbf{R}_c son diagonales $n \times n$.

$$\tilde{B}_c = \left(\mathbf{I} + \mathbf{R}_c \mathbf{U}_k (\mathbf{I}_k - \mathbf{V}_k^T \mathbf{R}_c \mathbf{U}_k)^{-1} \mathbf{V}_k^T \right) E_c \quad (3.31)$$

$$\tilde{B}_c = E_c + \mathbf{R}_c \left(\mathbf{U}_k \left((\mathbf{I}_k - \mathbf{V}_k^T \mathbf{R}_c \mathbf{U}_k)^{-1} (\mathbf{V}_k^T E_c) \right) \right) \quad (3.32)$$

$$\tilde{B}_c = E_c + \mathbf{R}_c \left(\mathbf{U}_k (\mathbf{M}_c (\mathbf{V}_k^T E_c)) \right) \quad (3.33)$$

Las tres Ecuaciones (3.33) son las **ecuaciones RRB en colores con geometría fija** y se pueden aplicar en escenas donde la geometría no varía, dado que las matrices \mathbf{M}_c se calculan una única vez al principio del proceso. Estas ecuaciones tienen una complejidad $O(nk)$ y ocupan una memoria $O(nk)$. Cada Ecuación (3.33) tiene un costo computacional un poco superior a la del caso acromático expresado en la ecuación RRB con geometría fija, debido a que debe incluirse el producto matriz-vector que involucra a la matriz \mathbf{M}_c y que posee un costo $O(k^2)$. Por lo tanto, en comparación con el caso acromático, el costo de operar con las tres Ecuaciones (3.33) es poco más de tres veces superior. En cuanto al uso de memoria, la principal diferencia es la existencia de las matrices \mathbf{M}_c y \mathbf{R}_c y los vectores B_c y E_c en las Ecuaciones (3.33). Esto implica un incremento de $3k^2 + 6n$ números de punto flotante a almacenar en memoria, en comparación con su contraparte acromática, siendo un incremento relativamente bajo si $k \ll n$.

Como conclusión, todas las ecuaciones RRB que contemplan color tienen un costo computacional un poco mayor a tres veces el de sus contrapartes acromáticas, pero sólo implican un volumen de memoria levemente superior.

3.4.5. Ecuación RRB en escenas con objetos en movimiento

Si la escena cambia debido al movimiento de algunos objetos que pasan de un estado inicial I a un estado final F , cambian las distancias y otras relaciones entre algunos parches de la escena, principalmente los parches pertenecientes a los objetos que han sido movidos y los parches cercanos a la posición inicial y final de los objetos movidos. El movimiento de los objetos cambia la matriz \mathbf{F}_I que contiene los factores de forma de la escena en el estado I , generando la matriz \mathbf{F}_F que contiene los factores de forma de la escena en el estado F . El movimiento de los objetos probablemente altere también a la matriz $\mathbf{U}_I \mathbf{V}_I^T$ que haya sido generada como aproximación de rango bajo a \mathbf{F}_I , generando la matriz $\mathbf{U}_F \mathbf{V}_F^T$ que es la aproximación de rango bajo a la matriz \mathbf{F}_F .

Las modificaciones a las matrices \mathbf{F}_I y $\mathbf{U}_I \mathbf{V}_I^T$ no se dan de forma aleatoria, sino que se concentran en las filas y columnas relacionadas con los parches más influenciados por el movimiento de los objetos. Si la diferencia entre $\mathbf{U}_I \mathbf{V}_I^T$ y $\mathbf{U}_F \mathbf{V}_F^T$ (o entre \mathbf{F}_I y \mathbf{F}_F) se pudiese aproximar a una matriz de rango bajo $\mathbf{U}_\Delta \mathbf{V}_\Delta^T$ de forma que cumpla con las relaciones establecidas en la Ecuación (3.34), entonces la ecuación de radiosidad de rango bajo $(\mathbf{I} - \mathbf{U}_F \mathbf{V}_F^T) \tilde{B}_F = E$ que está asociada al estado final de la escena se puede aproximar a las ecuaciones que contienen matrices de rango bajo (3.35) y (3.36), donde $\hat{B}_F \approx \tilde{B}_F \approx B_F$.

$$\mathbf{U}_I \mathbf{V}_I^T + \mathbf{U}_\Delta \mathbf{V}_\Delta^T \approx \mathbf{U}_F \mathbf{V}_F^T \approx \mathbf{F}_F \approx \mathbf{F}_I + \mathbf{U}_\Delta \mathbf{V}_\Delta^T \quad (3.34)$$

$$(\mathbf{I} - \mathbf{U}_I \mathbf{V}_I^T - \mathbf{U}_\Delta \mathbf{V}_\Delta^T) \hat{B}_F = E \quad (3.35)$$

$$\hat{B}_F = (\mathbf{I} - \mathbf{U}_I \mathbf{V}_I^T - \mathbf{U}_\Delta \mathbf{V}_\Delta^T)^{-1} E \quad (3.36)$$

Estas ecuaciones se pueden resolver utilizando métodos iterativos o directos.

Resolución iterativa

Utilizando la Ecuación (3.12) y partiendo del radiosidad \tilde{B}_I calculada en el estado inicial I , se puede formular la resolución iterativa de la Ecuación (3.35) a través de la Ecuación (3.37), donde la matriz $[\mathbf{U}_I | \mathbf{U}_\Delta]$ contiene las matrices \mathbf{U}_I y \mathbf{U}_Δ , y la matriz $[\mathbf{V}_I | \mathbf{V}_\Delta]$ contiene las matrices \mathbf{V}_I y \mathbf{V}_Δ .

$$\hat{B}_F^{(n+1)} = E + [\mathbf{U}_I | \mathbf{U}_\Delta] ([\mathbf{V}_I | \mathbf{V}_\Delta]^T \hat{B}_F^{(n)}), \text{ con } \hat{B}_F^{(0)} = \tilde{B}_I \quad (3.37)$$

Cada iteración tiene una complejidad de orden $n(k + k_\Delta)$, en donde k es el rango de las matrices \mathbf{U}_I y \mathbf{V}_I^T y k_Δ es el rango de \mathbf{U}_Δ y \mathbf{V}_Δ^T .

Resolución directa

En la Ecuación (3.36), la inversa de la matriz puede ser hallada utilizando la fórmula de Sherman-Morrison-Woodbury, generando la Ecuación (3.38), donde \mathbf{Y} , \mathbf{W} y \mathbf{Z} se hallan respectivamente por las Ecuaciones (3.39, 3.40 y 3.41).

$$(\mathbf{I} - \mathbf{U}_I \mathbf{V}_I^T - \mathbf{U}_\Delta \mathbf{V}_\Delta^T)^{-1} = (\mathbf{I} - \mathbf{Y} \mathbf{V}_I^T - \mathbf{Z} \mathbf{W}^T) \quad (3.38)$$

$$\mathbf{Y} = -\mathbf{U}_I (\mathbf{I}_k - \mathbf{V}_I^T \mathbf{U}_I)^{-1} \quad (3.39)$$

$$\mathbf{W} = \mathbf{V}_\Delta - \mathbf{V}_I (\mathbf{Y}^T \mathbf{V}_\Delta) \quad (3.40)$$

$$\mathbf{Z} = -(\mathbf{I} - \mathbf{Y} \mathbf{V}_I^T) \mathbf{U}_\Delta (\mathbf{I}_{k_\Delta} - \mathbf{W}^T \mathbf{U}_\Delta)^{-1} \quad (3.41)$$

Al aplicar la Ecuación (3.38) en la Ecuación (3.36), se obtiene la ecuación

$$\hat{B}_F = (\mathbf{I} - \mathbf{Y} \mathbf{V}_I^T - \mathbf{Z} \mathbf{W}^T) E$$

cuya expresión más simple de calcular es

$$\hat{B}_F = E - \mathbf{Y} (\mathbf{V}_I^T E) - \mathbf{Z} (\mathbf{W}^T E) \quad (3.42)$$

Si se ha calculado previamente \tilde{B}_I (radiosidad de la escena en el estado inicial I), entonces se puede aprovechar ese resultado para acelerar el cálculo de \tilde{B}_F . Para ello se utiliza la ecuación RRB con geometría fija para eliminar algunos términos de la Ecuación (3.42), obteniéndose

$$\hat{B}_F = \tilde{B}_I - \mathbf{Z} (\mathbf{W}^T E) \quad (3.43)$$

La complejidad del cálculo de \tilde{B}_F a partir de la Ecuación (3.43) depende principalmente de la complejidad del cálculo de la matriz inversa contenida en el cálculo de \mathbf{Z} (véase la Ecuación (3.41)), que a su vez depende del cálculo de \mathbf{W} (Ecuación (3.40)). La complejidad total es $O(nkk_\Delta)$, con $n \gg k$ y $n \gg k_\Delta$.

3.4.6. Conclusiones sobre los métodos de radiosidad de rango bajo

Cuando la geometría de la escena es fija, se puede establecer que las ecuaciones RRB con geometría fija y RRB en colores con geometría fija son fuertes candidatas a ser utilizadas en cualquier algoritmo de radiosidad en tiempo real. El ahorro de memoria que implican, el orden de complejidad, así como las posibilidades de aplicación de técnicas de computación paralela para mejorar el desempeño del producto matriz-vector, las habilitan a ser utilizadas en estaciones de trabajo portátiles, con o sin tarjetas gráficas. Por otra parte, los procesos iterativos expuestos en las Ecuaciones (3.12) y (3.30) tienen buenos órdenes de complejidad y uso de memoria. Cuando la geometría es fija, las matrices \mathbf{U}_k , \mathbf{V}_k e \mathbf{Y}_k , necesarias para la utilización de las ecuaciones mencionadas pueden ser calculadas previamente, mejorando la eficiencia computacional del cálculo de radiosidad. En el Capítulo 4 se analizan experimentos realizados para evaluar la

pertinencia de la utilización de la ecuación RRB con geometría fija para el cálculo de radiosidad en tiempo real.

Cuando la geometría es variable, las dificultades principales se encuentran en la construcción de las matrices \mathbf{U}_Δ , \mathbf{V}_Δ , \mathbf{Z} y \mathbf{W} . Para el cálculo de las matrices \mathbf{U}_Δ y \mathbf{V}_Δ queda pendiente el desarrollo de metodologías que eviten el cálculo previo de las matrices \mathbf{U}_F y \mathbf{V}_F , o de la matriz \mathbf{F}_F . Si estos inconvenientes fueran superados, entonces los órdenes de complejidad de los métodos indican que si el método iterativo converge en unas pocas iteraciones entonces resulta ser más conveniente que el método directo. Luego, como ya se mencionara en la sección 3.2.5, el cálculo de las factorizaciones SVD, DFT y DWT es complejo y costoso en memoria, por lo tanto resulta necesaria la búsqueda de formas alternativas de construcción de matrices de rango bajo para la matriz de radiosidad. Esta tarea es abordada en las secciones 3.5 y 3.6.

3.5. Bases para el cálculo de nuevas aproximaciones de rango bajo de RF

En esta sección se desarrollan tres ideas centrales de la tesis respecto a una propuesta específica de generación de aproximaciones de rango bajo ($\mathbf{U}_k \mathbf{V}_k^T$) de la matriz \mathbf{RF} . Estas ideas son utilizadas en la sección 3.6 para el desarrollo de dos algoritmos eficientes para el cálculo de radiosidad.

La primera idea clave es el reconocimiento de que, en el problema de radiosidad, la coherencia espacial hace que la radiosidad reflejada por la escena sea poco sensible a las perturbaciones en la emisión de la escena. La segunda idea clave es la construcción de dos mallas para una misma escena y la definición de su interrelación: una malla de grano fino compuesta por elementos y otra malla de grano grueso compuesta por parches. Por último, la tercer idea clave es el análisis y síntesis de propiedades interesantes, a los efectos de la tesis, de los factores de forma de parche a elemento y de elemento a parche.

3.5.1. Influencia de la coherencia espacial en la emisión y la reflexión

En el problema de radiosidad, dada una emisión E y una escena se calcula una radiosidad B , que está compuesta por la emisión E más la reflexión C de la escena ($B = C + E$). Como E es un dato del problema, en el cálculo de la radiosidad B lo que interesa conocer es la reflexión C . En esta sección se muestra que debido a la coherencia espacial, el cálculo de C es poco sensible a las variaciones de E . Esta propiedad puede ser aprovechada para generar buenos valores de C partiendo de una emisión E poco precisa representada en una malla de grano grueso. En la sección 3.6 se aprovecha esta propiedad para el cálculo de aproximaciones de rango bajo de la matriz \mathbf{RF} .

Existe una relación biunívoca entre la emisión y la reflexión de una escena, dado que por cada vector E de emisión, existe un único vector C de reflexión posible. En la práctica, esta relación biunívoca teórica resulta difícil de vivenciar. La coherencia espacial lleva a que emisiones E_1 y E_2 que sean sensiblemente diferentes se correspondan con reflexiones C_1 y C_2 muy parecidas (véase la Figura 3.4), prácticamente indetectables por un ojo o una cámara. Aunque las diferencias sean detectables, también es necesario establecer cuál emisión corresponde a cuál reflexión.

Matemáticamente, la relación entre C y E se define a partir de la ecuación matricial de radiosidad. Dada la relación existente entre B , C y E , se puede sustituir B en la ecuación matricial de radiosidad, obteniéndose $(\mathbf{I} - \mathbf{RF})(C + E) = E$. Una ecuación equivalente es la Ecuación (3.44), la **ecuación matricial de la radiosidad reflejada**.

$$(\mathbf{I} - \mathbf{RF})C = \mathbf{RFE} \quad (3.44)$$

En la Ecuación (3.44), hallar C conociendo E es un problema similar al problema de hallar B en la ecuación matricial de radiosidad. Los sistemas lineales tienen la misma matriz $(\mathbf{I} - \mathbf{RF})$ y difieren en que el término independiente de la Ecuación (3.44) es el vector \mathbf{RFE} y en la ecuación matricial de radiosidad es el vector E . Como ya se analizó en la propiedad 14 de la sección 2.4.2, estos problemas suelen estar bien condicionados.

Si se utiliza la Ecuación (3.44) para el cálculo de E a partir de C , el problema no está bien condicionado. En la Ecuación (3.44), la matriz del sistema lineal es \mathbf{RF} , que tiene un rango numérico bajo. Al ser un problema mal condicionado, es muy sensible a las variaciones del término independiente: pequeñas variaciones de C (y por lo tanto de $(\mathbf{I} - \mathbf{RF})C$) provocan grandes variaciones de E . El mal condicionamiento de \mathbf{RF} (provocado por la coherencia espacial de la escena) es la causa principal de la dificultad existente en establecer una correspondencia biunívoca entre C y E .

Existe una diferencia notoria entre la reflexión C real y la percibida por el ojo, por varios motivos: porque la pupila regula la intensidad de luz percibida, por existir una sensibilidad bastante limitada a la percepción de intensidades luminosas, porque el ojo está más orientado a detectar diferencias de intensidad y no tanto valores absolutos de intensidad luminosa, etc. [71]. Algunos de estos problemas son compartidos con las cámaras, escáneres y otros periféricos de entrada de información gráfica. Dados estos errores de percepción, existe una diferencia entre la reflexión C percibida y la real. En la Figura 3.4 se aprecia la correspondencia entre dos vectores E y C con valores reales E_R y C_R (puntos oscuros). Los valores de C percibidos por el usuario tienen un margen de error que se expresa por el área circundante. Dentro de esa área existen otros vectores C alternativos C_{P_1} y C_{P_2} que son indistinguibles de C por el usuario y que determinan emisiones E_{P_1} y E_{P_2} notoriamente diferentes a la real. Por lo tanto, al usuario le resulta prácticamente imposible establecer relaciones biunívocas claras entre ambas variables.

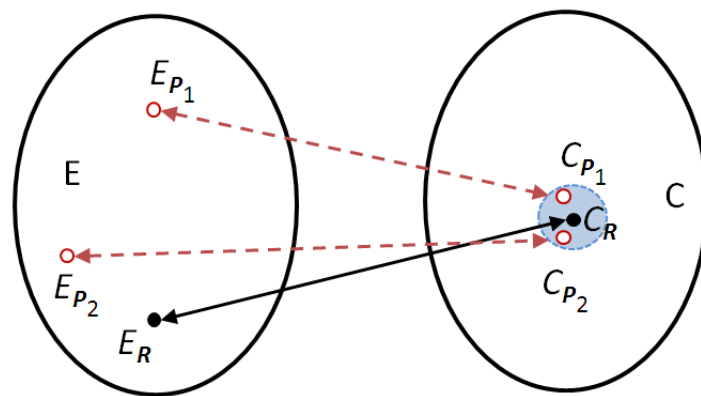


Figura 3.4: Pequeñas variaciones en C implican grandes variaciones en E .

La Figura 3.5 ejemplifica la insensibilidad de la reflexión con respecto a perturbaciones en la emisión. Se muestran dos emisiones E_1 y E_2 y sus correspondientes reflexiones C_1 y C_2 . Las dos emisiones son claramente diferentes entre sí, donde la emisión E_1 cubre un área mayor y posee una geometría más sencilla que la de la emisión E_2 . En cambio las reflexiones C_1 y C_2 son muy parecidas más allá de pequeñas diferencias. Si se intercambian las Figuras 3.5 (b) y (d) el resultado es indiferente para el usuario.

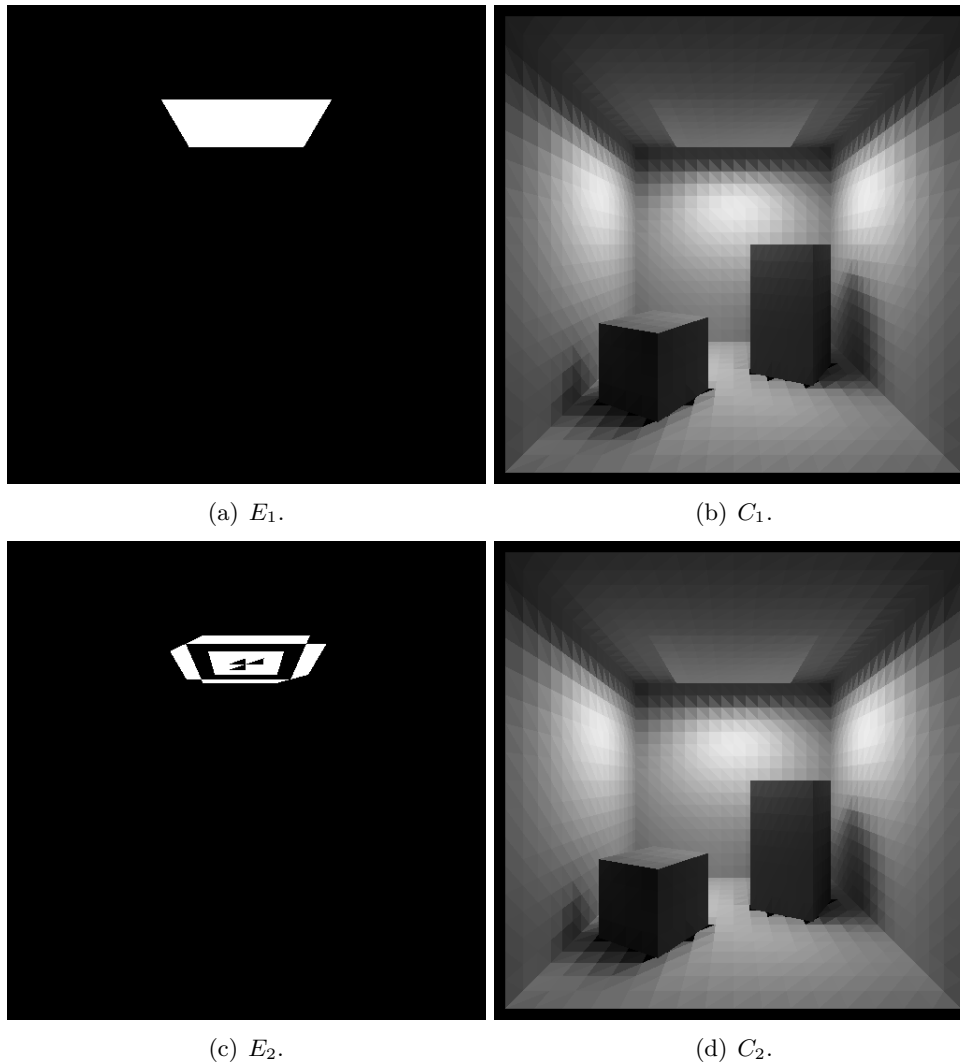


Figura 3.5: Dos E sensiblemente diferentes generan dos C prácticamente iguales.

La Figura 3.6 ejemplifica la extremada sensibilidad de la emisión respecto a cambios pequeños en la reflexión. Se muestran dos reflexiones que surgen de una misma reflexión C mostrada con distinta precisión (16 dígitos de precisión en (a) y 3 dígitos de precisión en (c)). Las diferencias entre ambas son irreproducibles por ningún dispositivo gráfico y aunque lo fueran, serían imperceptibles por el ojo humano. En cambio, las emisiones asociadas son extremadamente diferentes, generándose en el segundo caso radiosidades de emisión con valores negativos.

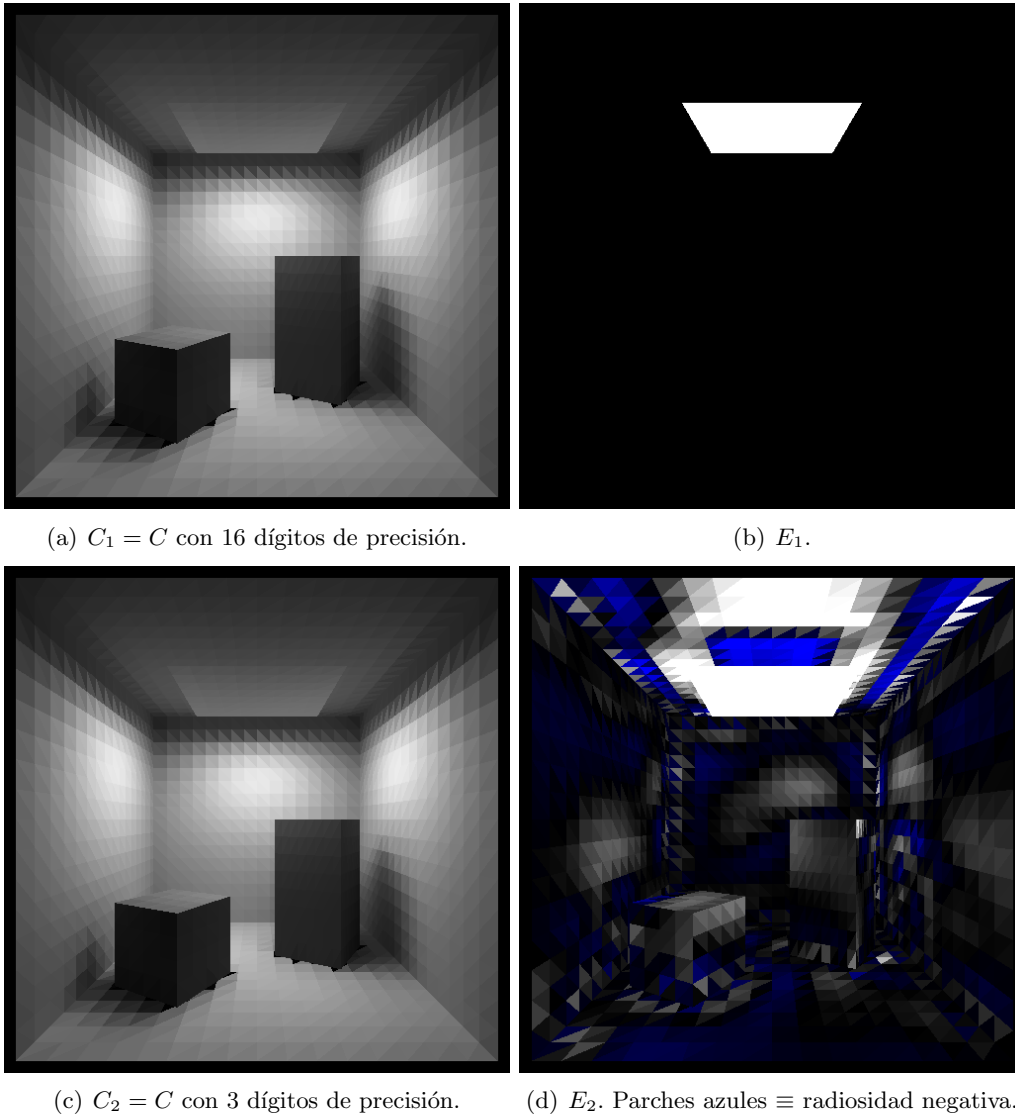


Figura 3.6: Dos reflexiones muy parecidas generan emisiones extremadamente distintas.

La dificultad de determinar cuál es la relación entre la emisión y la reflexión en radiosidad puede ser aprovechada para generar aproximaciones de rango bajo a la matriz \mathbf{F} , donde la geometría de la emisión E es simplificada a través del uso en la escena de una malla de grano grueso, pero donde la geometría de la reflexión C y de la radiosidad B es calculada con una malla de grano fino.

En lo que resta de la sección 3.5 y en la sección 3.6 se desarrolla una propuesta basada en la coherencia espacial para la generación de matrices \mathbf{U}_k y \mathbf{V}_k , cuyo producto genera una aproximación de rango bajo a la matriz \mathbf{F} . En la sección 4.2 del capítulo de análisis experimental se muestran los resultados de algunas experiencias realizadas con distintos vectores de emisión y reflexión, donde se compara la diferencia en norma entre ellos y se muestran escenas generadas, entre otros resultados.

3.5.2. Malla gruesa y malla fina de una escena

Para el desarrollo de la nueva propuesta de aproximación de \mathbf{RF} por matrices de rango bajo, se parte de la generación de dos mallas para la misma escena: una **malla de grano grueso** (mg), compuesta por k parches grandes, y una **malla de grano fino** (mf), compuesta por n parches pequeños denominados **elementos**. La cantidad de elementos y parches de las mallas determinan las dimensiones n y k de las matrices. Los elementos y parches cumplen que:

1. Los elementos de mf no se superponen. Lo mismo ocurre para parches de mg .
2. Cada elemento de mf está completamente contenido en un único parche de mg .
3. Cada parche de mg está cubierto por completo por elementos de mf de igual área.

Se dice que dos mallas mg y mf son **compatibles** si cumplen con las propiedades anteriores. La forma de los parches y de los elementos puede ser muy variada, dependiendo de la técnica utilizada para su generación. Por ejemplo, un conjunto de elementos puede formar un rostro humano y la unión de esos elementos conformar un único parche. Por lo tanto, la forma de cada parche mg puede ser muy irregular, existiendo incluso la posibilidad de armar parches a partir de elementos no conexos.

3.5.3. Factores de forma de parche a elemento y de elemento a parche

En esta sección se describen los conceptos básicos utilizados en la construcción de las matrices \mathbf{U}_k y \mathbf{V}_k .

Como se expresara en la sección 3.3.1, la existencia de coherencia espacial en la escena está relacionada con que haya pares de filas parecidas en \mathbf{F} . Si las filas son parecidas ($\mathbf{F}_{i_1j} \approx \mathbf{F}_{i_2j} \forall j$) y las áreas de los elementos asociados son iguales ($A_{i_1} = A_{i_2}$), como \mathbf{AF} es simétrica (propiedad 10 de la sección 2.4.2), las columnas de \mathbf{F} también son parecidas entre sí, es decir se cumple que

$$A_j \mathbf{F}_{ji_1} = A_{i_1} \mathbf{F}_{i_1j} \approx A_{i_2} \mathbf{F}_{i_2j} = A_j \mathbf{F}_{ji_2} \forall j, \text{ por lo tanto } \mathbf{F}_{ji_1} \approx \mathbf{F}_{ji_2} \forall j$$

Cuando las mallas son compatibles, los elementos pertenecientes a un mismo parche tienen áreas iguales y si poseen filas parecidas en la matriz \mathbf{F} (la cercanía existente entre los elementos posibilita la coherencia espacial), se concluye que las columnas de la matriz \mathbf{F} pertenecientes a elementos de un mismo parche también son parecidas.

Para ahorrar espacio en memoria, en lugar de almacenar un conjunto de columnas parecidas entre sí asociadas a los elementos de un parche se puede almacenar una única *columna representante* del parche. La columna representante puede ser el promedio del conjunto o ser la mejor columna del conjunto, seleccionada según algún criterio.

Cada columna representante expresa los factores de forma entre cada elemento de mf de la escena y un parche de mg . La técnica desarrollada calcula k columnas representantes (una por parche) de forma tal que cada columna de \mathbf{F} se parezca al menos a una de ellas. Estas columnas constituyen la matriz \mathbf{U}_k , cuya dimensión es $n \times k$ por tener tantas filas como elementos mf y columnas como parches mg tiene la escena. Cada columna de la matriz \mathbf{U}_k está asociada a un parche mg .

Al disponer de la matriz \mathbf{U}_k , se puede construir una matriz $\tilde{\mathbf{F}}$ parecida a \mathbf{F} , duplicando cada columna p de \mathbf{U}_k en las columnas de $\tilde{\mathbf{F}}$ asociadas a los elementos de p . Para lograr este objetivo, se construye una matriz \mathbf{V}_k de forma que el producto $\mathbf{U}_k \mathbf{V}_k^T$ genere

la matriz $\tilde{\mathbf{F}}$. Por construcción, la matriz \mathbf{V}_k es dispersa. En la Figura 3.7 se ejemplifica el razonamiento. Se muestran dos columnas densas de \mathbf{U}_k y las correspondientes dos filas de \mathbf{V}_k^T . Las celdas no nulas en las filas de \mathbf{V}_k^T contienen unos. Cada columna de \mathbf{V}_k^T contiene una única celda distinta de cero (que vale uno). En el ejemplo de la figura, las dos filas de \mathbf{V}_k^T contienen cuatro unos cada una. En el producto $\mathbf{U}_k \mathbf{V}_k^T$ se observa cómo cada columna de \mathbf{U}_k se repite cuatro veces en $\tilde{\mathbf{F}}$.

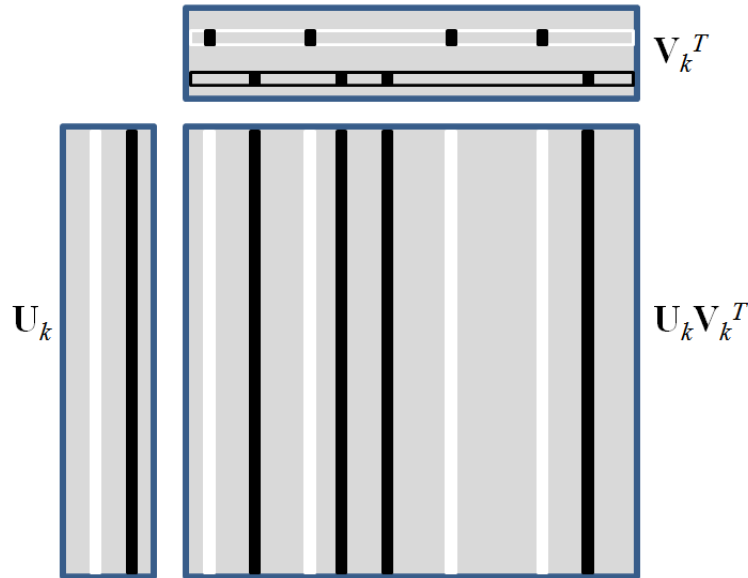


Figura 3.7: Producto matricial de dos columnas de \mathbf{U}_k por dos filas de una matriz \mathbf{V}_k^T dispersa. Cada columna de \mathbf{U}_k se replica cuatro veces en la matriz producto.

Cálculo de los factores de forma de parche a elemento

A partir de las Ecuaciones (2.6) y (2.30) se puede deducir que el factor de forma entre un parche y un elemento está definido por la expresión en la Ecuación (3.45), donde A_p es el área del parche p , θ_p es el ángulo entre la normal a dp y la dirección hacia de , θ_e es el ángulo entre la normal a de y la dirección hacia dp , r es la distancia entre de y dp , y V_{pe} indica la visibilidad entre de y dp (véase la Figura 3.8).

$$\mathbf{F}_{pe} = \frac{1}{A_p} \int_p \int_e \frac{\cos \theta_p \cos \theta_e}{\pi r^2} V_{pe} de dp \quad (3.45)$$

Dado que las áreas de los elementos del parche p son disjuntas, se pueden realizar las siguientes operaciones presentadas en la Ecuación (3.46) donde c_p es la cantidad de elementos pertenecientes al parche p y e_i^p es el i -ésimo elemento del parche p . A partir de este razonamiento se concluye que el factor de forma entre un parche p y un elemento e es igual al promedio ponderado por el área, de los factores de forma entre todos los elementos constituyentes del parche p y el elemento e .

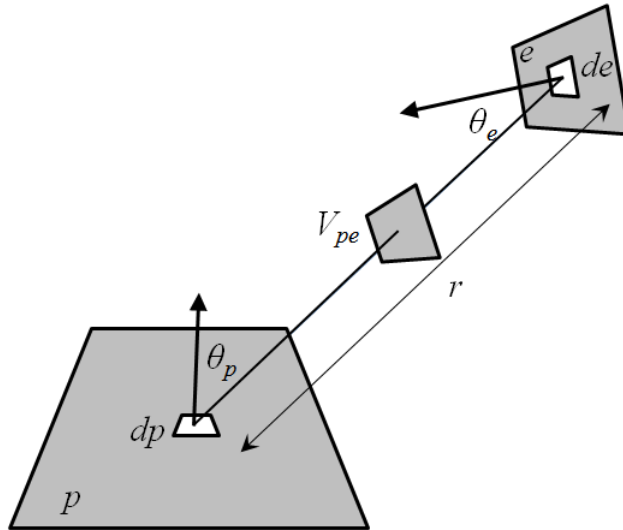


Figura 3.8: Elementos que determinan el factor de forma entre dos áreas.

$$\begin{aligned}
 \mathbf{F}_{pe} &= \frac{1}{A_p} \int_{\bigcup_{i=1}^{c_p} e_i^p} \int_e \frac{\cos \theta_p \cos \theta_e}{\pi r^2} V_{pe} de dp \\
 &= \frac{1}{A_p} \sum_{i=1}^{c_p} \int_{e_i^p} \int_e \frac{\cos \theta_{e_i^p} \cos \theta_e}{\pi r^2} V_{e_i^p e} de de_i^p \\
 &= \frac{1}{A_p} \sum_{i=1}^{c_p} A_{e_i^p} \mathbf{F}_{e_i^p e} \quad (3.46)
 \end{aligned}$$

Por otra parte, dado que las áreas $A_{e_i^p}$ de todos los elementos del parche p son iguales entre sí en mallas mg y mf compatibles, la fórmula se simplifica aun más, resultando

$$\mathbf{F}_{pe} = \frac{1}{c_p} \sum_{i=1}^{c_p} \mathbf{F}_{e_i^p e} \quad (3.47)$$

De esta manera, se determina que el factor de forma entre un parche p y un elemento e es igual al promedio de los factores de forma entre los elementos contenidos en p y el elemento e .

Cálculo de los factores de forma de elemento a parche

Los factores de forma \mathbf{F}_{ep} entre un elemento e y un parche p se calculan como la suma de los factores de forma $\mathbf{F}_{ee_i^p}$ entre el elemento e y los elementos e_i^p pertenecientes al parche p . Como $\mathbf{A}\mathbf{F}$ es una matriz simétrica, se puede expresar \mathbf{F}_{ij} mediante la Ecuación (3.48) para todo elemento (i, j) de la escena.

$$\mathbf{F}_{ij} = \frac{A_j}{A_i} \mathbf{F}_{ji} \quad (3.48)$$

En particular, para el elemento e y el parche p se puede establecer que

$$\mathbf{F}_{ep} = \frac{A_p}{A_e} \mathbf{F}_{pe} \quad (3.49)$$

por lo tanto, aplicando la Ecuación (3.48) en la Ecuación (3.46), se deduce

$$\mathbf{F}_{pe} = \frac{1}{A_p} \sum_{i=1}^{c_p} A_{e_i^p} \mathbf{F}_{e_i^p e} = \frac{1}{A_p} \sum_{i=1}^{c_p} A_{e_i^p} \frac{A_e}{A_{e_i^p}} \mathbf{F}_{ee_i^p} = \frac{A_e}{A_p} \sum_{i=1}^{c_p} \mathbf{F}_{ee_i^p}$$

y combinando las dos ecuaciones previas, se concluye que

$$\mathbf{F}_{ep} = \sum_{i=1}^{c_p} \mathbf{F}_{ee_i^p} \quad (3.50)$$

De esta manera se determina que el factor de forma entre un elemento e y un parche p es igual a la suma de los factores de forma entre el elemento e y cada uno de los elementos e_i^p contenidos en el parche p .

Parches con coherencia espacial

Un parche p de mg es un **parche con coherencia espacial** si todos los elementos pertenecientes a p observan prácticamente la misma escena, es decir, que si se toma un elemento e cualquiera perteneciente a mf se cumple siempre que todos los factores de forma $\mathbf{F}_{ee_j^p}$ son parecidos entre sí. Esto implica que podría comprimirse la información de los factores de forma, guardando un único número por elemento y parche. En efecto, la Ecuación (3.51) muestra que al conocer los \mathbf{F}_{ep} para un parche p y $\forall e$, se conocen de manera aproximada todos los $\mathbf{F}_{ee_j^p}$, donde los elementos e_j^p son aquellos pertenecientes a p .

$$\frac{\mathbf{F}_{ep}}{c_p} = \frac{\sum_{i=1}^{c_p} \mathbf{F}_{ee_i^p}}{c_p} \approx \mathbf{F}_{ee_j^p}, \forall e \text{ y } \forall e_j^p \in p \quad (3.51)$$

Si esta propiedad se extendiese a todos los parches de mg , podría reducirse sustancialmente la información a almacenar para generar una aproximación a la matriz \mathbf{F} de factores de forma entre elementos. En lugar de almacenar los n^2 números de punto flotante correspondientes a los $\mathbf{F}_{e_i e_j}$ $\forall i$ y $\forall j$, alcanza con almacenar los nk números de punto flotante correspondientes a los \mathbf{F}_{ep} $\forall e$ y $\forall p$.

Mallas mf y mg con coherencia espacial

Se define a dos mallas mf y mg como **mallas con coherencia espacial** si ambas mallas son compatibles y si todos los parches de mg son parches con coherencia espacial. El que todos los parches posean coherencia espacial implica que existen relaciones iguales a las expresadas en (3.51) entre todo elemento de mf y todo parche de mg . Cuando las mallas mf y mg son mallas con coherencia espacial, es posible construir una matriz $\tilde{\mathbf{F}}$ aproximada a \mathbf{F} si se dispone únicamente de los factores de forma \mathbf{F}_{ep} . Esta posibilidad permite ahorrar memoria, debido a que existen $n \times n$ factores de forma de elemento a elemento y existen sólo $n \times k$ factores de forma de elemento a parche.

En la sección 3.6.1 se desarrollan estos conceptos para la construcción de dos matrices \mathbf{U}_k y \mathbf{V}_k cuyo producto sea una aproximación a \mathbf{F} cuando hay mallas con coherencia espacial. Luego, en la sección (3.6.3) se desarrolla un algoritmo para la construcción de mallas con coherencia espacial. Las matrices \mathbf{U}_k y \mathbf{V}_k construidas poseen dimensión $n \times k$, donde n es la cantidad de elementos de mf y k la cantidad de parches de mg . La matriz \mathbf{V}_k es dispersa y la matriz \mathbf{U}_k es densa y contiene todos los factores de forma de elemento a parche \mathbf{F}_{ep} .

La ventaja de que la matriz \mathbf{V}_k sea dispersa es que su utilización ahorra memoria, ahorro que se mantiene cuando se la utiliza en el cálculo de \tilde{B} al emplear las ecuaciones directas: (3.21), (3.22), (3.32) y (3.33), o iterativas: (3.12) y (3.30).

3.6. Algoritmos 2MF y 2MCE

En esta sección se exponen dos algoritmos que generan un nuevo tipo de aproximaciones $\mathbf{U}_k \mathbf{V}_k^T$ a la matriz \mathbf{F} . Estas aproximaciones se componen de una matriz \mathbf{U}_k , cuyas columnas están fuertemente vinculadas a las columnas de \mathbf{F} (de forma similar a lo que ocurre en la descomposición CUR), y de una matriz \mathbf{V}_k que entre sus propiedades se destaca el ser ortonormal (como ocurre en las aproximaciones basadas en SVD, DFT y DWT) y ser dispersa. Los algoritmos desarrollados utilizan una memoria temporal que varía según el caso entre $O(nk)$ y $O(n^2)$, y poseen una complejidad que también varían entre $O(nk)$ y $O(n^2)$.

Se proponen dos algoritmos de cálculo:

1. El algoritmo 2MF, que *presupone* que mf y mg son mallas con coherencia espacial.
2. El algoritmo 2MCE, que *genera* dos mallas mf y mg con coherencia espacial.

Ambos métodos generan matrices \mathbf{V}_k dispersas, permitiendo un ahorro de memoria que posibilita la realización de aproximaciones competitivas con aquellas basadas en la descomposición SVD.

3.6.1. Algoritmo 2MF: existencia de mallas con coherencia espacial

En esta sección se explica el algoritmo **2MF** (**2 Mallas Fijas**) en el que, partiendo de dos mallas con coherencia espacial se generan dos matrices \mathbf{U}_k y \mathbf{V}_k de dimensión $n \times k$, cuyo producto $\mathbf{U}_k \mathbf{V}_k^T$ es una aproximación de rango bajo de la matriz \mathbf{F} .

Como se expresó en la sección anterior, si se conocen todos los factores de forma \mathbf{F}_{ep} entonces, a partir de la Ecuación (3.51) es posible construir una buena aproximación de \mathbf{F} a través del producto $\mathbf{U}_k \mathbf{V}_k^T$ con \mathbf{V}_k dispersa (véase la Figura 3.7).

Al asumir que mf y mg son mallas con coherencia espacial no es necesario modificarlas para construir aproximaciones de rango bajo de \mathbf{F} . La evaluación para determinar si son mallas con coherencia espacial se trata en la sección 3.6.2 y el cálculo de mallas mf y mg que cumplan con la coherencia espacial se trata en la sección 3.6.3.

Asumiendo que, para cualquier par e y p , todos los factores de forma $\mathbf{F}_{ee_i^p}$ son parecidos entre sí, basándose en la Ecuación (3.51) se puede establecer la expresión de la Ecuación (3.52) para todo par de elementos e_i y e_j . La función $P(e)$ devuelve el parche al que pertenece cualquier elemento e .

$$\mathbf{F}_{e_i e_j} \approx \frac{\mathbf{F}_{e_i P(e_j)}}{c_{P(e_j)}} = \tilde{\mathbf{F}}_{e_i e_j} \quad (3.52)$$

Por lo tanto, si se conocen todos los factores de forma \mathbf{F}_{ep} , con la expresión (3.52) se puede construir la matriz $\tilde{\mathbf{F}}$ que es aproximada a \mathbf{F} y con rango k .

La construcción de la matriz $\tilde{\mathbf{F}}$ también puede ser realizada construyendo dos matrices \mathbf{U}_k y \mathbf{V}_k , de dimensión $n \times k$, siendo n el número de elementos y k el número de parches, y calculando el producto $\mathbf{U}_k \mathbf{V}_k^T$. A continuación se definen ambas matrices y luego se muestra que su producto es una aproximación a la matriz \mathbf{F} .

La matriz \mathbf{U}_k contiene en cada celda (e, p) el valor del factor de forma \mathbf{F}_{ep}/c_p :

$$\mathbf{U}_k(e, p) = \mathbf{F}_{ep}/c_p \quad (3.53)$$

La matriz \mathbf{V}_k es una matriz dispersa que pasa la emisión de la malla fina a la malla gruesa. Se construye de forma que cada fila asociada al elemento e tiene solamente un uno en las columna $P(e)$ (otra forma de definir a \mathbf{V}_k consiste en establecer que cada columna p sólo tiene unos en las filas asociadas a los elementos e_i^p).

$$\mathbf{V}_k(P(e), e) = 1 \quad (3.54)$$

El resto de las celdas valen cero. Dado que cada elemento pertenece a un único parche, cada fila de \mathbf{V}_k contiene un único uno.

Con estas definiciones, el producto $\mathbf{U}_k \mathbf{V}_k^T = \tilde{\mathbf{F}}$ es similar a \mathbf{F} , ya que para todo $\tilde{\mathbf{F}}_{e_i e_j}$ se cumple que

$$\tilde{\mathbf{F}}_{e_i e_j} = \sum_{p=1}^m \mathbf{U}_k(e_i, p) \mathbf{V}_k^T(p, e_j) \quad (3.55)$$

En todos los términos de la sumatoria $\mathbf{V}_k^T(p, e_j)$ es nulo salvo cuando $p = P(e_j)$, donde vale uno por definición de \mathbf{V}_k . La sumatoria se reduce a un único término, permitiendo obtener (basándose en las Ecuaciones (3.53), (3.54) y (3.52)):

$$\tilde{\mathbf{F}}_{e_i e_j} = \mathbf{U}_k(e_i, P(e_j)) \mathbf{V}_k^T(P(e_j), e_j) = \frac{\mathbf{F}_{e_i P(e_j)}}{c_{P(e_j)}} \mathbf{1} \approx \mathbf{F}_{e_i e_j} \quad (3.56)$$

Por lo tanto, el producto $\mathbf{U}_k \mathbf{V}_k^T$ de las matrices calculadas en esta sección permite construir una matriz $\tilde{\mathbf{F}}$ aproximada a \mathbf{F} .

Como se observa en la Ecuación (3.53), la matriz \mathbf{U}_k se construye directamente a partir de la matriz de dimensión $n \times k$ con los factores de forma \mathbf{F}_{ep} . La formulación para calcular estos factores de forma se expone en el siguiente punto.

Cálculo de \mathbf{F}_{ep}

El algoritmo del hemicubo (véase la Sección 2.4.1) se centra en un elemento e_i y calcula los factores de forma $\mathbf{F}_{e_i e_j} \forall e_j$. Este algoritmo puede ser utilizado como base para la construcción de los factores de forma \mathbf{F}_{ep} entre elementos y parches. Para la generación de estos factores de forma existen dos metodologías claramente distinguibles:

1. Para cada elemento e se realizan las siguientes operaciones:
 - se calculan todos los factores de forma $\mathbf{F}_{e e_i^p}$. Los elementos e_i^p son todos los elementos de la malla mf , representados de forma de poderlos asociar al parche al que pertenecen.
 - se realiza el cálculo de \mathbf{F}_{ep} utilizando la Ecuación (3.50).

Como ventaja destacable de esta metodología se tiene la exactitud teórica del cálculo de los \mathbf{F}_{ep} debida a la Ecuación (3.50) -si no se tienen en cuenta los errores de redondeo y el *aliasing* del algoritmo del hemicubo-. Entre las desventajas se destaca la lentitud de la metodología, debido a que implica el cálculo de n^2 factores de forma. El cálculo de los n^2 factores de forma no implica que deban estar almacenados. De hecho, sólo n factores de formas calculados con el algoritmo del hemicubo se encuentran en memoria al mismo tiempo.

2. Para cada parche p se realizan las siguientes operaciones:

- se calculan los factores de forma \mathbf{F}_{pe} para todos los elementos e de mf .
- se realiza el cálculo de \mathbf{F}_{ep} utilizando la Ecuación (3.49).

Como ventaja destacable de esta metodología se tiene la velocidad de cómputo, debido a que implica el cálculo de sólo nk factores de forma. Son sólo n los factores de formas calculados con el algoritmo del hemicubo que se almacenan en memoria al mismo tiempo. Como desventaja de la metodología se destaca su relativa imprecisión, debido a que el algoritmo del hemicubo puede cometer errores mayores a los deseados en el cálculo de los factores de forma \mathbf{F}_{pe} , cuando el parche p es grande.

Rango del cálculo de \mathbf{U}_k y \mathbf{V}_k

Como se observara en el punto anterior, la complejidad del cálculo de la matriz \mathbf{U}_k varía entre nk y n^2 operaciones, dependiendo de la metodología empleada, pero en ambos casos la memoria utilizada tiene $O(nk)$ debido al tamaño de la matriz final.

El cálculo de la matriz \mathbf{V}_k tal como está expresado en la Ecuación (3.54) tiene una complejidad $O(n)$ y utiliza también una memoria $O(n)$.

3.6.2. Evaluación de la existencia de mallas con coherencia espacial

La construcción de las matrices \mathbf{U}_k y \mathbf{V}_k depende de cómo se distribuyen los n elementos de mf en los k parches de mg . Además de cumplir con las tres propiedades de la sección 3.5.2, la ecuación (3.51) indica que es necesario la existencia de mallas con coherencia espacial.

Para que estas propiedades se verifiquen, en general basta con tomar parches relativamente pequeños respecto a la escena. Debido a la coherencia espacial, es común que los valores de los factores de forma sean cercanos en los elementos de cada parche. Lamentablemente, la coherencia espacial no siempre se cumple. Un caso donde no se cumple ocurre cuando hay discontinuidad en las superficies, como la existente entre la pared y el piso de una habitación. En estas superficies se da el caso en que cuanto más cercano a la pared se encuentre un punto del piso, su radiosidad está influida por áreas más pequeñas de la pared. Esta propiedad de las superficies puede provocar singularidades en los factores de forma obteniéndose como resultado la pérdida la coherencia espacial. Por ejemplo, en la Figura 3.9 (a) se observa que la influencia que ejerce el área A5 sobre la radiosidad de P5, es la misma que ejerce A1 sobre P1, por compartir el mismo ángulo sólido. Dado que el área A1 es 256 veces más grande que A5, una variación en la radiosidad emitida por A5 influirá mucho más en el punto P5 que del punto P1, contradiciendo la propiedad de coherencia espacial.

Para que estos fenómenos no influyan demasiado en la radiosidad final calculada para la escena, se deben construir mallas con parches cada vez más pequeños a medida que se acercan a la discontinuidad problemática (véase la Figura 3.9 (b)). Este tipo de mallas fue desarrollado por Hanrahan et al. en su trabajo sobre radiosidad jerárquica [38]. Como regla práctica, se define un área mínima por debajo de la cual no se construyen parches, a pesar que la geometría indique lo contrario. De cualquier manera, los errores que se pueden cometer al aplicar esta regla práctica tienen poca influencia en el resultado global de la radiosidad de la escena.

En esta subsección se muestra cómo pueden ser evaluados los parches para decidir sobre la conveniencia de dividirlos, y se determina un algoritmo que utiliza la evaluación de parches para generar una malla mg que cubra toda la escena.

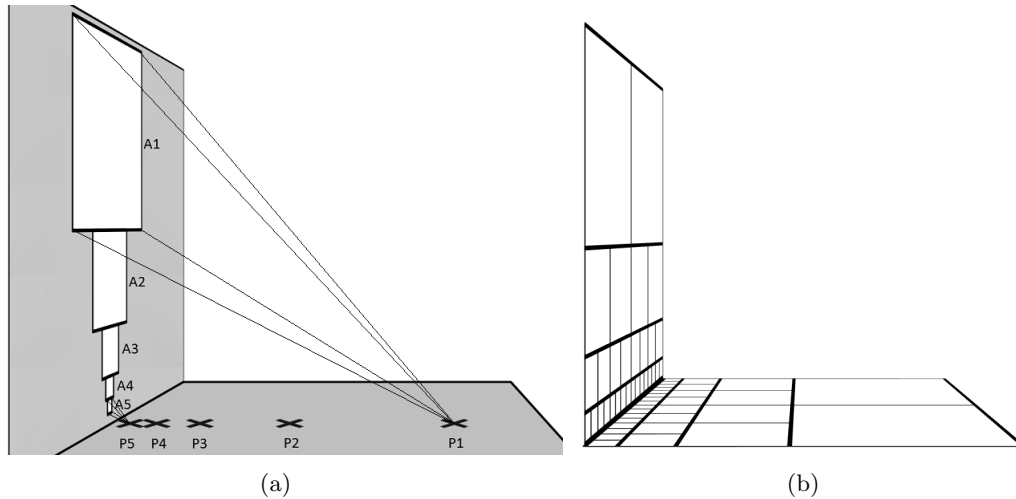


Figura 3.9: Entre la pared y el piso se pierde coherencia espacial, forzando a generar mallas especiales.

Tipos de evaluación.

Para saber si se requiere dividir un parche p es necesario evaluar si la Ecuación (3.51) se cumple para todos los factores de forma $\mathbf{F}_{ee_i^p}$ entre todos los elementos e de la escena y todos los elementos e_i^p incluidos en p . Para ello, dado un parche p es necesario hallar n valores \mathbf{F}_{ep} (uno por cada elemento e) realizando las sumatorias (3.50) y luego, con alguna métrica, medir la distancia de los $\mathbf{F}_{ee_i^p}$ al correspondiente promedio $\frac{\mathbf{F}_{ep}}{c_p}$. Si esta distancia está siempre por debajo de un umbral preestablecido, entonces se cumple la Ecuación (3.51) y no es necesario subdividir al parche p .

La evaluación de cada parche p puede realizarse de forma completa o parcial respecto a la cantidad de elementos utilizados, y estricta o amplia considerando la distribución de los factores de forma.

Respecto a la cantidad de elementos utilizados, la evaluación de parches puede ser:

- **evaluación completa** es aquella que utiliza todos los valores $\mathbf{F}_{ee_i^p}$, $\forall e$ y $\forall e_i^p$ del parche p .
- **evaluación parcial** es aquella que trabaja con una muestra significativa de elementos e y/o con una muestra significativa de elementos e_i^p del parche p .

Considerando la distribución de los valores de los factores de forma, la evaluación de cada parche puede ser:

- **evaluación estricta** es aquella en que basta que algún $\mathbf{F}_{e_j e_i^p}$ tenga un valor alejado de $\frac{\mathbf{F}_{ep}}{c_p}$ más allá de cierto umbral, para que se subdivida el parche.
- **evaluación amplia** es aquella en que la subdivisión del parche se determina a partir de una métrica global única que involucra a elementos e de la escena y elementos e_i^p del parche p , y donde se observa prioritariamente la relación de toda la escena con el parche sobre la de cada elemento de la escena con el parche.

Las combinaciones dos a dos de estas clasificaciones (completa y estricta, parcial y estricta, completa y amplia, parcial y amplia), agrupan a las evaluaciones en cuatro categorías.

En cuanto a la complejidad de cada categoría, resulta claro que evaluar todos los parches de una escena con una metodología completa y estricta es más costoso que evaluarlos a través de una metodología parcial y amplia. La primera evaluación implica el cálculo de n^2 factores de forma, y realizar al menos n^2 operaciones con esos factores de forma. La segunda evaluación posibilita operar con una fracción representativa de los factores de forma.

En cuanto a la calidad de los resultados, argumentos estadísticos permiten asegurar que utilizando una evaluación parcial y amplia que tome muestras representativas y trabaje con métricas adecuadas, se pueden obtener resultados similares a los de una evaluación completa y estricta.

Implementación de los tipos de evaluación.

Para realizar una evaluación completa de cada parche p se utiliza la Ecuación (3.50) para calcular los factores de forma \mathbf{F}_{ep} .

- *Una posible evaluación estricta y completa del parche p* está basada en el cálculo del máximo de las desviaciones estándar de los factores de forma $\mathbf{F}_{ee_i^p}$ agrupados por cada elemento e de la escena

$$eval_{ec}(p) = \max_{\forall e \in escena} \sqrt{\frac{1}{c_p} \sum_{i=1}^{c_p} \left(\frac{\mathbf{F}_{ep}}{c_p} - \mathbf{F}_{ee_i^p} \right)^2} \quad (3.57)$$

En esta evaluación, una desviación estándar mayor al umbral en un único elemento e alcanza para dividir un parche.

- *Una posible evaluación amplia y completa del parche p* está basada en el cálculo del promedio de las desviaciones estándar de los factores de forma $\mathbf{F}_{ee_i^p}$ agrupados por cada elemento e de la escena

$$eval_{ac}(p) = \frac{1}{n} \sum_{\forall e \in escena} \left(\sqrt{\frac{1}{c_p} \sum_{i=1}^{c_p} \left(\frac{\mathbf{F}_{ep}}{c_p} - \mathbf{F}_{ee_i^p} \right)^2} \right) \quad (3.58)$$

En esta evaluación, tienen que existir un número considerable de desviaciones estándar mayores al umbral para que el resultado final también lo sea.

Si en lugar de evaluación completa se realiza una evaluación parcial entonces, en vez de utilizar la Ecuación (3.50) para calcular \mathbf{F}_{ep} , se pueden tomar muestras de m_p elementos por parche y aplicar la Ecuación (3.59) para calcular una aproximación $\tilde{\mathbf{F}}_{ep}$.

$$\tilde{\mathbf{F}}_{ep} = \frac{c_p}{m_p} \sum_{m=1}^{m_p} \mathbf{F}_{ee_{im}^p} \quad (3.59)$$

La Ecuación (3.59) es utilizada en el cálculo de las siguientes evaluaciones parciales:

- Una posible evaluación estricta y parcial del parche p está basada en el cálculo del máximo de las desviaciones estándar de los factores de forma $\mathbf{F}_{ee_i^p}$ considerando una muestra m_{escena} de los elementos de la escena y una muestra m_p de los elementos del parche

$$eval_{ep}(p) = \max_{\forall e \in m_{escena}} \sqrt{\frac{1}{|m_p|} \sum_{\forall e_i^p \in m_p} \left(\frac{\tilde{\mathbf{F}}_{ep}}{c_p} - \mathbf{F}_{ee_i^p} \right)^2} \quad (3.60)$$

- Una posible evaluación amplia y parcial del parche p está basada en el cálculo del promedio de las desviaciones estándar de los factores de forma $\mathbf{F}_{ee_i^p}$ considerando una muestra m_{escena} de los elementos de la escena y una muestra m_p de los elementos del parche, debiendo estar uniformemente distribuidos los elementos de la escena y los del parche.

$$eval_{ap}(p) = \frac{1}{|m_{escena}|} \sum_{\forall e \in m_{escena}} \left(\sqrt{\frac{1}{|m_p|} \sum_{\forall e_i^p \in m_p} \left(\frac{\tilde{\mathbf{F}}_{ep}}{c_p} - \mathbf{F}_{ee_i^p} \right)^2} \right) \quad (3.61)$$

El análisis de las fórmulas precedentes muestra que se pueden reducir las n^2 operaciones de la Ecuación (3.57) (dado que $\sum_{p=1}^k c_p = n$), ya que la implementación de (3.61) requiere $k|m_p||m_{escena}|$ operaciones, que es una fracción de n^2 si $k \ll n$, $|m_p| \ll c_p$ y $|m_{escena}| \ll n$.

3.6.3. Algoritmo 2MCE: generación de mallas con coherencia espacial

En esta sección se explica el algoritmo **2MCE** (**2** Mallas con **C**oherencia **E**spacial), que incluye un método recursivo. En este algoritmo, dada una escena se generan dos mallas de diferente nivel (una fina y otra gruesa) con coherencia espacial. Luego de generadas ambas mallas se está en condiciones de generar dos matrices \mathbf{U}_k y \mathbf{V}_k , cuyo producto $\mathbf{U}_k \mathbf{V}_k^T$ es una aproximación de rango bajo a la matriz \mathbf{F} .

El algoritmo parte de una malla inicial de grano grueso mg inicial, compuesta por parches de aproximadamente el mismo tamaño. La malla inicial es lo suficientemente gruesa como para que no haya necesidad de generar parches más grandes a partir de la unión de los parches iniciales, pero sí puede haber necesidad de dividir a algunos de ellos para lograr que las mallas tengan coherencia espacial.

La malla de grano fino mf se construye subdividiendo la malla inicial mg , para que ambas mallas sean compatibles entre sí (véase la sección 3.5.2). Todos los parches de la malla inicial mg se dividen en igual cantidad de elementos y todos los elementos dentro

de un parche son de igual tamaño. La malla mf se considera lo suficientemente fina como para que no exista la necesidad de subdividir a ninguno de sus elementos.

El algoritmo 2MCE mantiene fija a la malla mf y modifica la malla mg dividiendo aquellos parches cuya coherencia espacial se encuentra por encima de un umbral preestablecido.

En el algoritmo 2MCE se realiza una evaluación amplia y parcial de los parches (véase la Ecuación (3.61)). En la implementación utilizada en análisis experimental presentada en el Capítulo 4 las muestras son de dimensiones $|m_{escena}| = n$ y $|m_p| = 4$. Se utiliza una evaluación amplia debido a que las evaluaciones estrictas pueden terminar generando muchos parches (incrementando el valor de k) y generar radiosidades de calidad mayor a la deseable (con el consiguiente costo computacional). El valor de $|m_{escena}|$ es n porque para el cálculo de \mathbf{U}_k se requiere disponer de \mathbf{F}_{ep} para todos los elementos de la escena. En la implementación desarrollada del algoritmo 2MCE se trabaja con $|m_p| = 4$ como resultado de la experiencia y debido a la forma en que se dividen los parches (en un número de subparches potencia de cuatro).

Generación del nivel inicial de parches

El primer paso del algoritmo consiste en la generación de un nivel inicial (nivel cero) de parches que cubran toda la escena. Se parte de una escena compuesta por triángulos, que pueden ser muy grandes para ser considerados como parches. Por ejemplo, bastan dos triángulos para definir la pared de una habitación (véase la Figura 3.10 (a)). Por lo tanto, para obtener resultados realistas es necesario subdividirlos hasta que cada triángulo resultante tenga un tamaño relativamente pequeño y sea útil para las técnicas de radiosidad (véanse las Figuras 3.10 (b) y (c)). No habrá parches más grandes que estos triángulos y probablemente muchos no pasen la evaluación de coherencia espacial, por lo que será necesario subdividirlos.

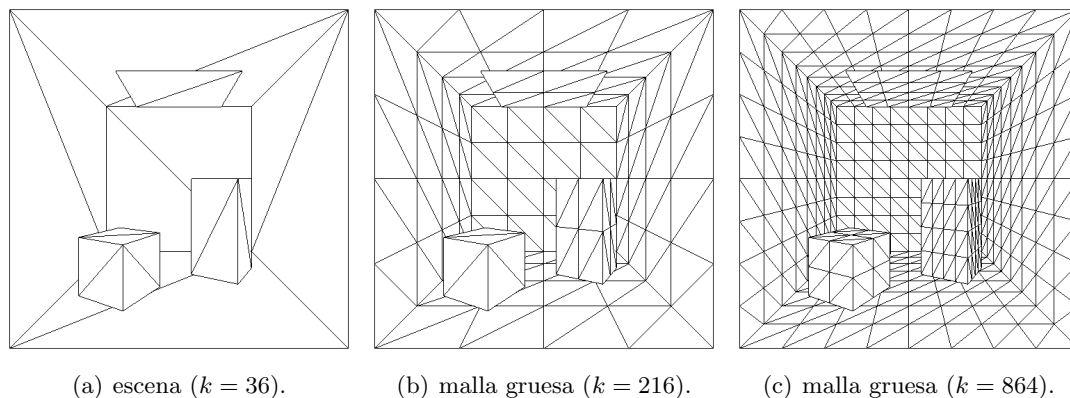


Figura 3.10: En (a) los triángulos de la escena son muy grandes. En (b) y (c) se representa la misma escena con mallas gruesas de distintas densidades.

Conviene que los parches generados tengan tamaños similares entre sí. En tal caso, si cada parche tiene igual cantidad de elementos y éstos son de igual área, todos los elementos de la escena también tendrán tamaños similares.

Generación de una estructura jerárquica de parches y elementos

Una vez definido el nivel cero de parches se está en condiciones de generar la malla de elementos. Como se observó en la sección 3.5.2, para que las mallas de parches mg y de elementos mf sean compatibles los elementos que cubren un parche tienen que tener igual área, no deben superponerse entre sí y deben ser parte de un único parche. En el método propuesto, para lograr la compatibilidad entre las mallas de elementos y parches se subdividen los parches de nivel cero en cuatro triángulos iguales, éstos en cuatro nuevamente y así sucesivamente hasta llegar a un tamaño de triángulo adecuado considerando la calidad de la imagen, el uso de memoria, etc. (véase la Figura 3.11).

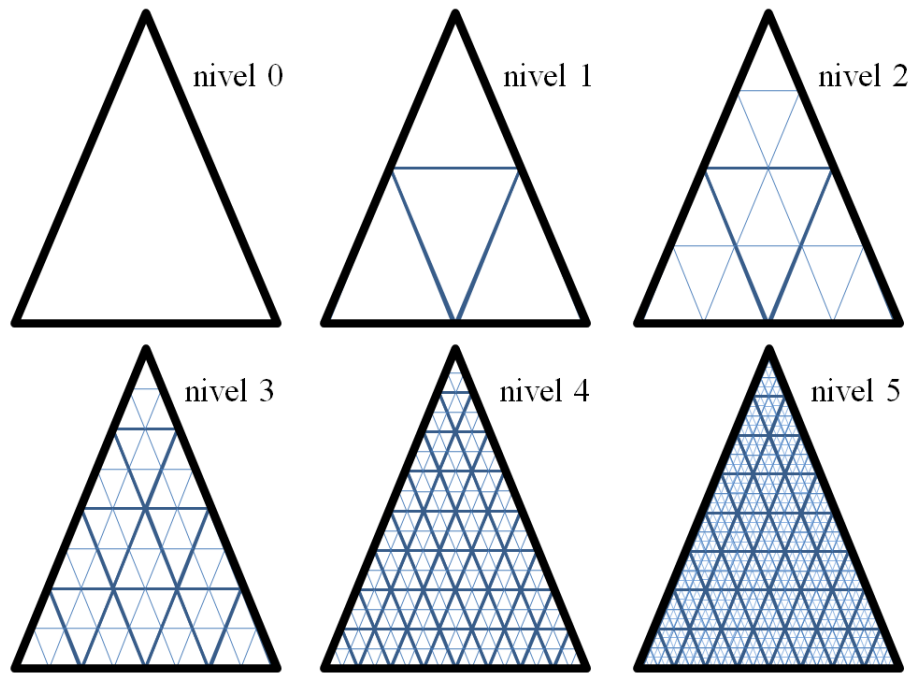


Figura 3.11: Cinco niveles de subdivisión de un parche.

A medida que se generan los triángulos se crea una estructura jerárquica compuesta por árboles, donde se almacenan todos los triángulos que se van obteniendo. Los parches de nivel cero son los nodos raíz de los árboles en la estructura jerárquica. El parche p está asociado al p -ésimo árbol, del cual es raíz. El resto de los nodos de cada árbol almacenan todos los triángulos obtenidos en las subdivisiones del parche correspondiente. Un triángulo que se divide en cuatro triángulos se representa en el árbol como un nodo con cuatro hijos. Las hojas del árbol p son todos los elementos e_i^p que cubren la superficie del parche p (y que es nodo raíz de ese árbol) (véase la Figura 3.12).

Así como el nivel cero de triángulos es una versión inicial de la malla gruesa mg de la escena, todas las hojas de la estructura jerárquica representan la malla fina mf que contiene todos los elementos de la escena. Las dos mallas mg y mf pueden ser suficientes para generar resultados de radiosidad satisfactorios en escenas variadas, como se muestra en los trabajos del autor [22, 23] donde se trabaja con este tipo de mallas sin evaluar los parches.

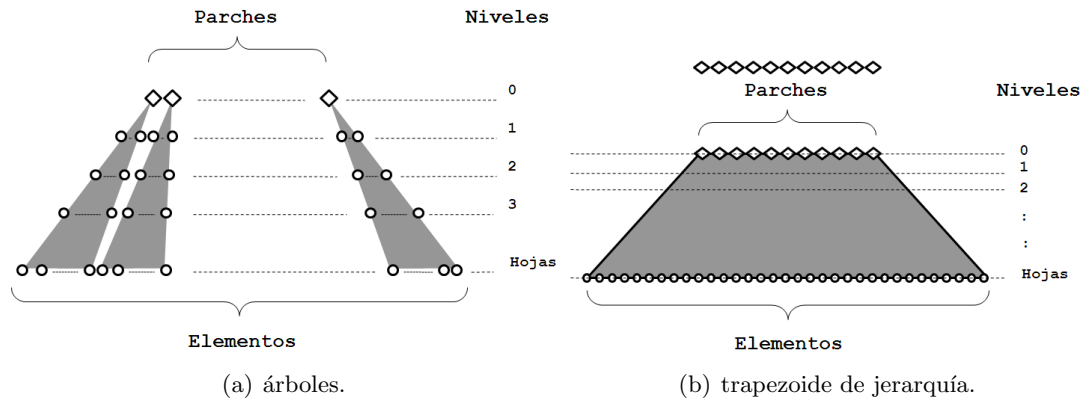


Figura 3.12: Dos representaciones de la estructura jerárquica. En (a) cada parche nivel cero genera un árbol. En (b) sólo se aprecia la jerarquía.

La cantidad de nodos de la estructura jerárquica es de $O(\frac{4}{3}(n - i))$, donde n es el número de elementos de la malla fina (y el número de hojas de la estructura), e i es el número de nodos del nivel inicial de la estructura.

Como se verá en el siguiente punto, la estructura jerárquica es conveniente debido a que facilita la tarea de saber cuál es el parche al que pertenece cada elemento, a que simplifica la tarea de dividir un parche en recoger información de los hijos de un nodo y a que facilita toda la algoritmia necesaria para la construcción de \mathbf{U}_k y \mathbf{V}_k .

El cuadro de la Figura 3.13 presenta la estructura de datos requerida para implementar el algoritmo de generación de la estructura jerárquica de parches y elementos.

```

nodo:
  polígono
  conjunto de punteros a nodo: hijos
  puntero a nodo: padre
  entero: nivel
  booleano: tieneparches
  vector de  $n$  números de punto flotante:  $ff$  (factores de forma)
  entero:  $maxnivel$ ,  $nivelhojas$ 
  número de punto flotante:  $\epsilon$ 
  conjunto de punteros a nodo: árboles, parches

```

Figura 3.13: Estructura de datos para implementar el método jerárquico.

Generación final de la malla mg

En la estructura de datos utilizada en el algoritmo 2MCE (véase la Figura 3.13), cada nodo tiene una variable booleana *tieneparches*, que indica si ese nodo o algún descendiente es parche. En un principio está en Falso. Para simplificar el funcionamiento del algoritmo y mantener la forma triangular de los parches, si el hijo de un nodo es parche o contiene parches, entonces los tres hijos restantes pasan a ser parches.

Un esquema del algoritmo 2MCE se presenta en el Algoritmo 1.

Algoritmo 1 Algoritmo 2MCE.

```

1 parches=∅
2 Defino maxnivel
3 Defino umbral
4 para cada (nodo ∈ árboles / nodo.nivel==0) hacer
5   HalloParches(nodo)
6 fin
  
```

Los pasos del algoritmo 2MCE involucran:

- Inicializar *parches* como una lista vacía de punteros a nodo.
- Determinar *maxnivel*, que define hasta qué nivel se desciende en el árbol para generar los parches. Como se observara en la sección 3.6.2 y especialmente en la Figura 3.9, donde hay discontinuidad en las normales de las superficies se requieren parches infinitesimales para salvar la coherencia espacial de estos, pero también es cierto que el costo de tener muchos parches pequeños no compensa el efecto final en la radiosidad generada. Por lo tanto, resulta necesario establecer un tope de nivel máximo en el que puede estar ubicado un parche.
- Determinar el *umbral* que no debe superar la evaluación de parche (véase la sección 3.6.2).
- Invocar a la rutina HalloParche por cada *nodo* nivel cero que exista en *árboles*.

La rutina HalloParches es el componente principal del algoritmo 2MCE. Su entrada es un *nodo* de la estructura jerárquica *árboles* y su salida afecta la estructura jerárquica y otras variables globales. El objetivo de HalloParches es hallar los parches de *mg* que son hijos de un nodo. La rutina es recursiva: comienza en la raíz de todos los árboles (nivel cero), baja hasta el máximo nivel (*maxnivel*), y luego sube evaluando los nodos y asignándolos como parches si corresponde.

La rutina HalloParches determina qué nodos descendientes de *nodo* son parches o contienen parches. Para cumplir con este objetivo, el algoritmo realiza las siguientes operaciones:

- Como primer tarea, dado *nodo* el algoritmo se invoca a sí mismo para determinar si sus *hijos* son parches o contienen parches. Si algún hijo es o contiene parches, entonces el resto de sus hijos que no sean ni contengan parches pasan a ser parches y son incluidos en la lista *parches* (véase la Figura 3.14 (a)).
- Si ningún hijo del *nodo* es o contiene parches, entonces los hijos tienen coherencia espacial. Resta saber si el propio *nodo* cumple con la coherencia espacial, por lo tanto se evalúa la coherencia espacial de *nodo* (determinando si la Ecuación (3.61) evaluada en el *nodo* es menor que el *umbral*). Si la coherencia espacial de *nodo* es superior al *umbral*, entonces *nodo* no cumple con la coherencia espacial y no puede ser parche, ni tampoco sus ancestros. Como sus *hijos* sí cumplen la coherencia espacial, ellos determinan el mayor nivel que cumple la coherencia espacial y por lo tanto se los designa como parches (véase la Figura 3.14 (b)).

- Por completitud, para cubrir toda la escena con parches los nodos de nivel cero que no tienen hijos con parches y que tienen buena coherencia espacial se los designa como parches (véase la Figura 3.14 (c)). Con este ítem se asegura que los parches cubran toda la escena.

- En los nodos de nivel *maxnivel* se calculan los factores de forma utilizando hemicubos u otra técnica. Estos factores de forma son necesarios para evaluar la coherencia espacial de sus padres, utilizando la ecuación (3.61). No corresponde verificar si sus hijos tienen parches (porque no hay parches en el nivel siguiente) ni corresponde evaluar la coherencia espacial (porque el parche no será subdividido en caso que no pase la evaluación).

Un esquema de la rutina `HalloParches` se presenta en el Algoritmo 2.

Algoritmo 2 Algoritmo recursivo `HalloParches`.

```

1 HalloParches(nodo)
2 si el nivel del nodo < maxnivel
3   para cada hijo del nodo hacer
4     HalloParches(hijo)
5   fin
6 hallar ff del nodo a partir de los ff de los hijos (Ec. (3.50))
7 si  $\exists$  hijos del nodo que posean parches (tieneparches == True)
8   asignar tieneparches = True en el resto de los hijos del nodo
9   incluir en parches al resto de los hijos del nodo
10  asignar tieneparches = True en el nodo
11 sino si la coherencia espacial del nodo > umbral (Ec. (3.61))
12   poner tieneparches = True a todos los hijos del nodo
13   incluir en parches a todos los hijos del nodo
14 fin
15 si el nivel del nodo = 0 y ningún hijo tiene parches
16   asignar tieneparches = True en el nodo
17   incluir en parches al nodo
18 fin
19 sino si el nivel del nodo = maxnivel
20 hallar ff del nodo (por ej. utilizando hemicubo)
21   tieneparches = False en el nodo
22 fin

```

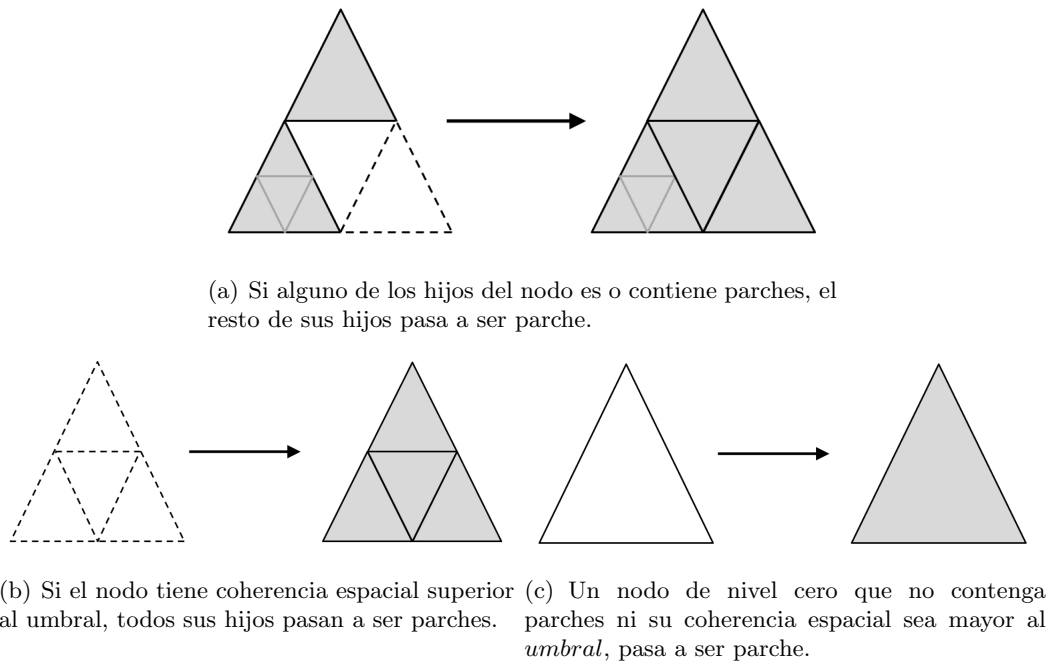


Figura 3.14: Acciones que se realizan en cada nodo de la estructura jerárquica.

Con el algoritmo 2MCE se cubre toda la escena con parches. En la Figura 3.15 se muestra cómo podrían quedar distribuidos los parches de la escena luego del algoritmo. Se observa que los parches se distribuyen entre el nivel cero y *maxnivel* y que cubren toda la escena, ya que todos los elementos están debajo de algún parche.

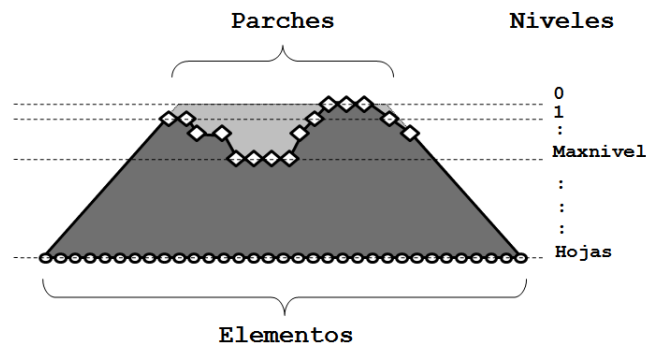


Figura 3.15: Los parches se distribuyen entre el nivel cero y *maxnivel*. Estos nodos así como sus ancestros (que están en la zona de color gris claro) pertenecen a la categoría *tiene parches*.

En la Figura 3.16 se muestran dos representaciones de los parches de la escena. La Figura 3.16 (a) fue generada utilizando *maxnivel* = 1 por lo que incluye sólo dos niveles de parches. La Figura 3.16 (b) fue generada utilizando *maxnivel* = 2 por lo que incluye tres niveles de parches. Se puede observar que allí donde el piso se encuentra con la pared o con los objetos se acumulan parches pequeños (ubicados en *maxnivel*) y cómo en superficies más aisladas los parches son más grandes, tendiendo al nivel cero.

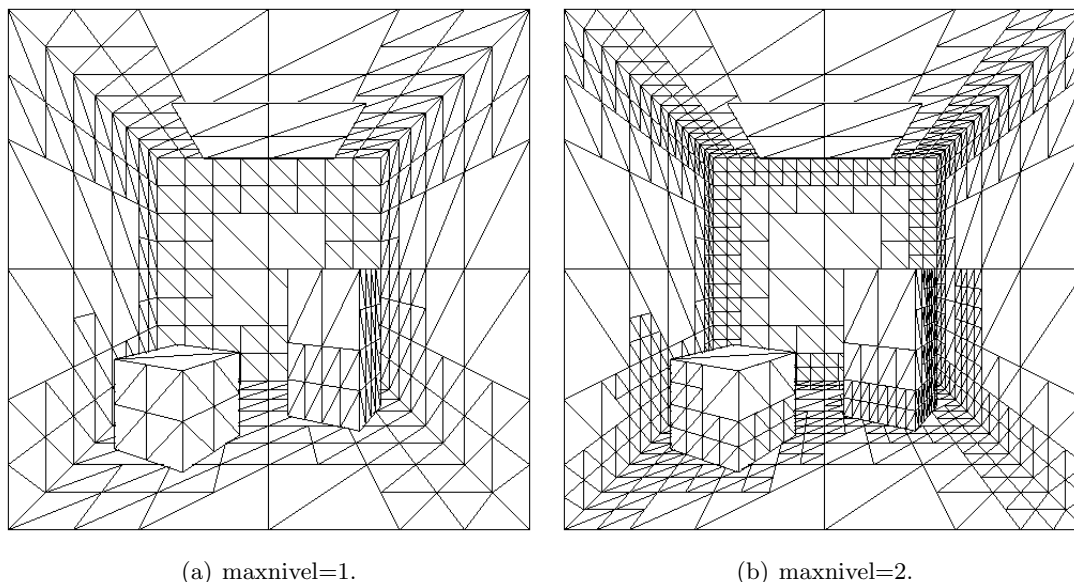


Figura 3.16: Parches generados al aplicar el algoritmo 2MCE

La recorrida y cálculo de todos los factores de forma $\mathbf{F}_{e,nodo}$ entre elemento y nodo, tiene una complejidad máxima de $n^{\frac{4i}{3}} \cdot 4^{maxnivel}$, donde n es el número de hojas en la estructura, i es el número de nodos en el nivel inicial y $\frac{4i}{3} \cdot 4^{maxnivel}$ es el número de nodos entre el nivel inicial y $maxnivel$ inclusive, si cada nodo tiene cuatro hijos. Para que se cumpla que $k \ll n$ y entonces sea conveniente utilizar el algoritmo, $maxnivel$ debe estar dos o más niveles por encima del nivel de las hojas. En esta situación se cumple que $\frac{4}{3} \frac{n}{16} = \frac{1}{12}n \geq \frac{4i}{3} \cdot 4^{maxnivel}$ y la complejidad del algoritmo 2MCE está acotada por $\frac{1}{12}n^2$. Esta cota puede ser elevada en comparación con la complejidad del algoritmo en la práctica, ya que corresponde al peor caso, en que todos los parches están en el máximo nivel. Aun así es sensiblemente inferior a la del cálculo de SVD. Como ejemplo, para una matriz con $n = 3500$, el algoritmo 2MCE es diez veces más rápido que SVD.

Construcción de \mathbf{U}_k y \mathbf{V}_k

Luego de ejecutar el algoritmo 2MCE, dado un elemento e se puede saber a qué parche p pertenece, recorriendo sus ancestros hasta encontrar uno que pertenezca a $parches$. Como se observó en la sección 3.6.1, conociendo el parche al que pertenece cada elemento, se puede construir \mathbf{V}_k (véase la Ecuación (3.54)), que es utilizado para transformar la emisión de malla fina en emisión de malla gruesa (véanse las Figuras 4.4 y 4.5 del Capítulo 4). Por otra parte, dado que por cada parche p de la escena se tienen almacenados los correspondientes \mathbf{F}_{ep} , aplicando la Ecuación (3.53) se construye \mathbf{U}_k .

El cálculo de la matriz \mathbf{U}_k tiene complejidad $O(nk)$ y el cálculo de la matriz \mathbf{V}_k tiene complejidad $O(n)$. Ambas matrices se calculan mediante el algoritmo cuyo esquema se presenta en el Algoritmo 3.

Algoritmo 3 Algoritmo para construir \mathbf{U} y \mathbf{V}

```

1  $n$  = cantidad de elementos
2  $k$  = cantidad de parches
3 Generar la matriz  $\mathbf{U}$   $n \times k$  con ceros
4 para cada nodo que sea parche hacer
5     En  $\mathbf{U}$  poner en la columna del (índice del parche) los (ff del nodo)/ $c_{\text{parche}}$ 
6 fin
7 Generar la matriz  $\mathbf{V}$   $n \times k$  con ceros
8 para cada nodo que sea hoja en árboles hacer
9     Hallar el ancestro del nodo que sea parche
10     $\mathbf{V}((\text{índice de nodo}), (\text{índice del parche ancestro})) = 1$ 
11 fin

```

3.7. Conclusiones del capítulo

En este capítulo se planteó el ecuación de radiosidad de rango bajo (RRB). Se explicaron sus bases teóricas y las consecuencias de sustituir la matriz \mathbf{RF} por una matriz de rango bajo. Utilizando esta sustitución se observó que se facilita el cálculo de la radiosidad con un error acotado, lográndose órdenes competitivos para el cálculo en tiempo real si la geometría es estática. Cuando la geometría varía, el problema principal se encuentra en el cálculo en tiempo real de las matrices de rango bajo cercanas al producto matricial \mathbf{RF} . Se desarrollaron los conceptos necesarios para calcular nuevas aproximaciones de rango bajo que permitan mejorar la eficiencia computacional del método. Por último, se desarrolló una metodología basada en la utilización de dos niveles de malla de la escena: una de grano grueso compuesta por parches, y otra de grano fino compuesta por elementos.

En base a los conceptos presentados, se desarrollaron algoritmos para radiosidad de rango bajo que poseen en general una complejidad de $O(nk)$ y un uso de memoria $O(nk)$, donde k es la cantidad de parches de la malla gruesa y n la cantidad de elementos de la malla fina. Estos órdenes son muy inferiores a aquellos que resultan de las descomposiciones SVD, DFT y DWT, y son similares a los de CUR.

En el Capítulo 4 se analizan y comparan diversos resultados de este capítulo.

Capítulo 4

Análisis experimental

4.1. Introducción

En este capítulo se presentan los principales resultados del análisis experimental, relativos a la resolución del problema de radiosidad utilizando matrices de rango bajo.

En primer término se exponen algunos resultados que brindan una primer validación a los desarrollos realizados sobre la relación entre la emisión y la reflexión de la escena. El capítulo continúa realizando comparaciones del error de la radiosidad generada por diversas implementaciones del algoritmo de radiosidad de rango bajo, donde se compara la solución exacta con las soluciones que surgen de la utilización de aproximaciones de rango bajo calculadas a partir de las descomposiciones SVD, CUR, 2MF y 2MCE. Luego, prosigue con la comparación de los tiempos de ejecución de la ecuación RRB con geometría fija en CPU y GPU. Por último, para el caso en que la escena contiene objetos en movimiento, se evalúa la validez de la hipótesis que considera que las modificaciones de la matriz \mathbf{F} pueden aproximarse por matrices de rango bajo.

4.2. Relación entre la emisión y la reflexión

En esta sección se presentan análisis experimentales con el objetivo de probar la validez de los conceptos desarrollados en la sección 3.5.1 sobre las relaciones existentes entre la emisión E y la reflexión C . Para obtener resultados concluyentes se deberán realizar cálculos con escenas de geometría diversa. En particular, se deberá establecer en qué tipo de escenas es más factible que se cumplan las hipótesis planteadas para la matriz de factores de forma \mathbf{F} y por lo tanto, cuándo es conveniente utilizar los métodos desarrollados en este trabajo. Los resultados de esta sección son útiles para evaluar la insensibilidad de la radiosidad de la escena respecto a la emisión inicial, así como la efectividad de la nueva factorización en sacar provecho de esa característica para reducir el volumen de información contenida en las matrices \mathbf{U}_k y \mathbf{V}_k .

4.2.1. Evaluación de la sensibilidad entre la emisión y la reflexión

En un primer paso, se realizan pruebas de sensibilidad de la reflexión ante perturbaciones en la emisión para una geometría con 896 parches (véase la Figura 4.1).

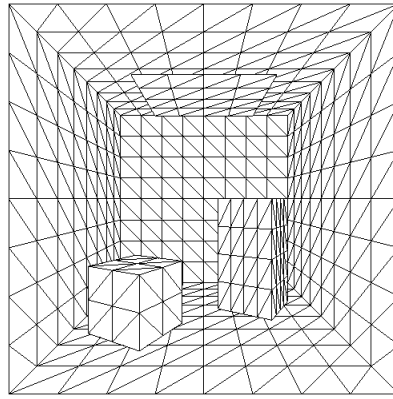


Figura 4.1: Esquema de una escena con 896 parches.

Para realizar las pruebas de sensibilidad se toman dos configuraciones de parches emisores (E_1 y E_2) que son notoriamente diferentes (véanse las Figuras 4.2 (a) y (b)).

En la configuración mostrada en la Figura 4.2 (a) los parches emisores cubren parcialmente un cuadrado superior de la escena, y en la configuración mostrada en la Figura 4.2 (b) los parches emisores cubren totalmente el cuadrado superior. Las dos configuraciones de parches emisores pueden verse como dos representaciones de la misma iluminación pero con diferente nivel de granularidad. En la Figura 4.1 hay 32 triángulos en el cuadrado emisor, con los que se puede generar la emisión representada en la Figura 4.2 (a). Si, en cambio, el cuadrado emisor no tiene divisiones, entonces sólo se puede generar la emisión representada en la Figura 4.2 (b).

En ambas configuraciones de emisión se iguala la potencia lumínica. Como la radiosidad es la potencia lumínica por unidad de área, para igualar la potencia emisora se debe cumplir la igualdad $E_1^T A = E_2^T A$, donde A es un vector columna con las áreas de los parches. Luego se calculan los valores de radiosidad reflejada por los parches en ambas escenas (que equivale al cálculo de C_1 y C_2 con $C_i = B_i - E_i$). Posteriormente se calculan las diferencias relativas (en norma) $\|E_1 - E_2\|_2 / \|E_1\|_2$ y $\|C_1 - C_2\|_2 / \|C_1\|_2$. Por último, se generan las imágenes correspondientes a la reflexión de los parches (véanse las Figuras 4.2 (c) y (d)) y la diferencia entre ellas (véase la Figura 4.2 (e)). La Figura 4.2 (f) es la representación de las reflexiones con interpolación de Gouraud [35].

Del análisis de los resultados se extrae que, en la escena utilizada y con emisiones testeadas, la distancia entre ambas emisiones es mucho mayor a la distancia de las reflexiones resultantes.

$$\frac{\|E_1 - E_2\|_2}{\|E_1\|_2} = 0,7746 \quad (4.1)$$

$$\frac{\|C_1 - C_2\|_2}{\|C_1\|_2} = 0,0332 \quad (4.2)$$

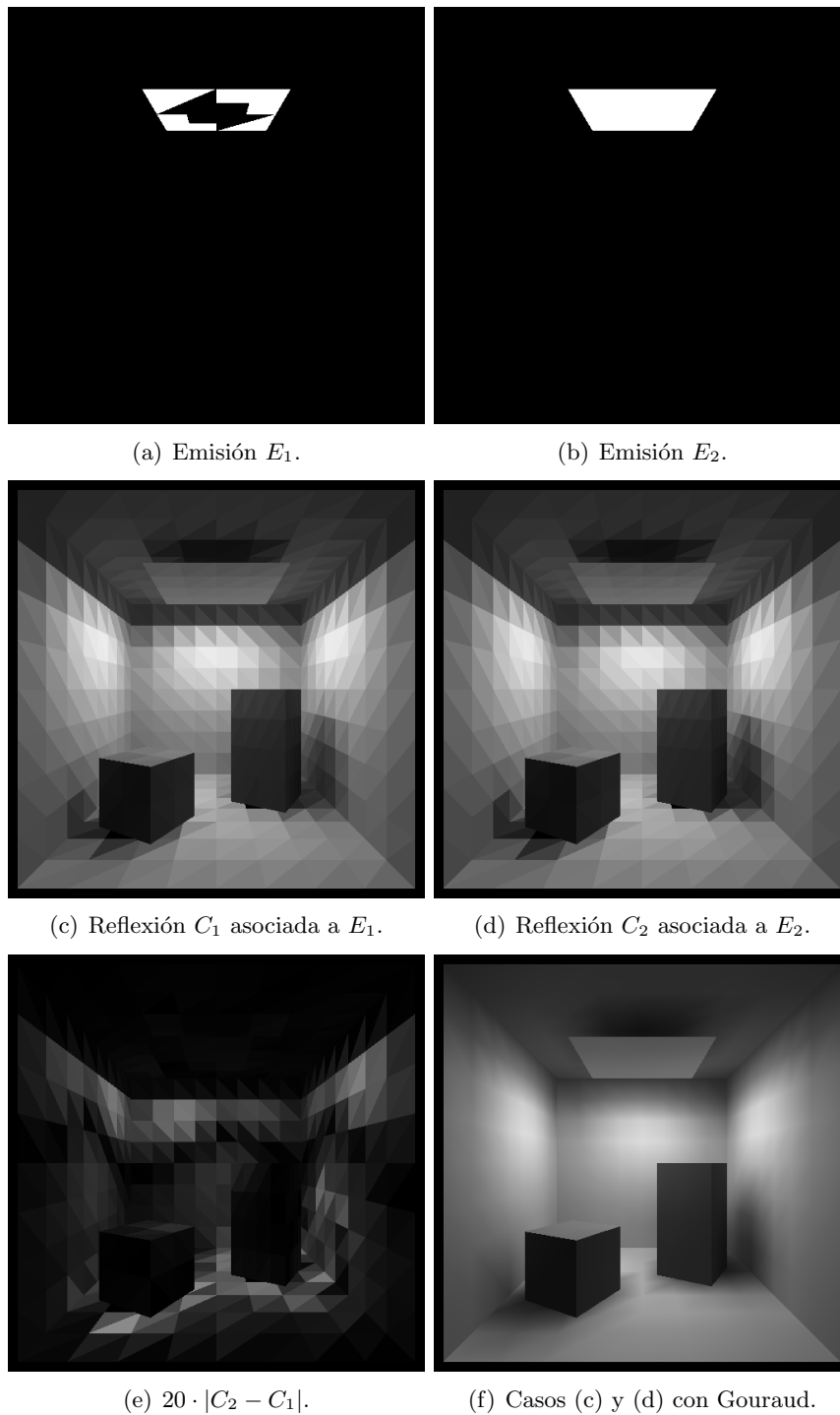


Figura 4.2: Diferentes imágenes que muestran el comportamiento de la reflexión de los parches ante emisiones diferentes. En (a) y (b) se muestran las emisiones, en (c) y (d) las reflexiones correspondientes, en (e) se muestra la diferencia entre ambas reflexiones amplificadas por un factor de 20, y en (f) cómo se ven las reflexiones si son interpoladas con Gouraud (por ser muy parecidas sólo se muestra una de las interpolaciones).

Se observa que la diferencia relativa de las emisiones genera diferencias unas 20 veces menores en las reflexiones. Desde punto de vista perceptual, una persona puede percibir claramente las diferencias las emisiones E_1 y E_2 (Figuras 4.2 (a) y (b)), pero le resulta extremadamente difícil percibir diferencias entre las reflexiones C_1 y C_2 (Figuras 4.2 (c) y (d)), cuyas diferencias ($|C_1 - C_2|$) se muestran, amplificadas por un factor de 20, en la Figura 4.2 (e). Por lo tanto, dos emisiones sensiblemente diferentes pueden generar dos conjuntos de reflexiones muy parecidas.

El problema inverso es aun más interesante. Si se parte de un vector de reflexión C y se halla el vector de emisión E a partir de la ecuación matricial de la radiosidad reflejada, se observa que el resultado es muy sensible a la cantidad de dígitos de precisión de C . Por ejemplo, para llegar a E con un dígito de precisión se necesita conocer C con nueve dígitos de precisión. Esto se debe a que el número de condición de la matriz \mathbf{RF} , que es la matriz del sistema que permite hallar E a partir de C , es del orden de 10^9 . Cuando C se genera a partir de un dispositivo -como la captura de pantalla, o una cámara en una escena real- no se tienen 9 dígitos de precisión y por lo tanto resulta imposible calcular la ubicación correcta de los parches emisores, o la configuración del vector E .

Si, con el fin de lograr mayor precisión, se generan más parches a partir de la subdivisión de parches (por ejemplo, subdividiendo en los parches de la Figura 4.1), la coherencia espacial hará que se produzca el efecto inverso, debido a que las filas de \mathbf{F} se parecerán entre sí aun más, incrementando el número de condición de la matriz \mathbf{RF} y dificultando aún más el cálculo preciso de E . Otro resultado notable es que los vectores E hallados a partir de C pueden tener muchos valores negativos grandes. Este resultado es inconsistente con la realidad, ya que la radiosidad sólo puede ser no negativa. En conclusión, resulta complejo establecer la correspondencia entre las emisiones y las reflexiones de una escena.

Resultados físicamente más realistas se obtienen si se agrega la restricción de que el E calculado a partir de C debe ser un vector con valores no negativos. En este caso se deben aplicar técnicas de optimización lineal. En experimentos no formalizados realizados con sistemas lineales de 896 variables, el tiempo empleado para el cálculo de E con restricciones resultó ser unas mil veces superior al empleado en la resolución del sistema lineal sin restricciones.

El ejemplo desarrollado muestra cómo el cambio de la emisión E de los parches implica cambios mucho menores en la radiosidad C reflejada por los parches. A su vez, se observan las dificultades que conlleva deducir la emisión E de una escena a partir de valores de reflexión C relevados de la escena.

La propuesta de aproximación de la matriz \mathbf{RF} utilizando dos mallas con coherencia espacial implica la simplificación de la geometría de emisión en la escena, dado que se la aproxima por una emisión de grano grueso que utiliza menos parches. Los resultados presentados en esta sección indican que esta simplificación es conveniente debido a que tiene resultados poco perceptibles en el cálculo de la reflexión C y la radiosidad $B = E + C$.

Se deben realizar estudios exhaustivos con un abanico amplio de escenas y configuraciones de parches emisores para relevar problemas que puedan surgir de la simplificación de la emisión. Esta investigación se propone como una de las líneas de trabajo futuro.

4.2.2. RRB basada en mallas con coherencia espacial

En esta sección se ejemplifica el funcionamiento del método de radiosidad de rango bajo cuando la aproximación a la matriz \mathbf{RF} surge de los algoritmos 2MF o 2MCE. Se observa que la emisión E definida en una malla fina es reemplazada por otra aproximada que está definida sobre una malla gruesa compatible con la malla fina. Luego de realizado el reemplazo se calcula la radiosidad de la escena. Para la implementación del ejemplo, se parte de una escena sobre la que se genera una malla de grano fino con $n = 13312$ elementos y una malla de grano grueso compatible con la anterior que contiene únicamente $k = 559$ parches (véanse las Figuras 4.3 (a) y (b)). Como consecuencia, se genera una aproximación con matrices \mathbf{U}_k y \mathbf{V}_k de dimensión 13312×559 .

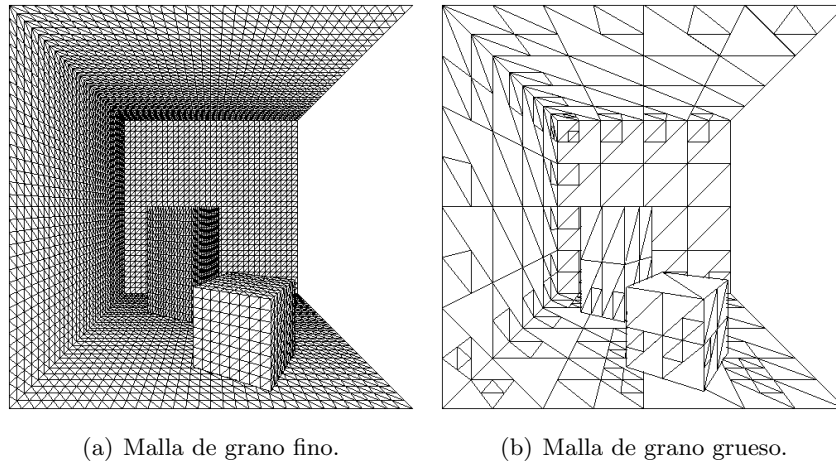


Figura 4.3: Mallas de grano fino y grano grueso utilizados en el cálculo de la aproximación $\mathbf{U}_k \mathbf{V}_k^T$ de la matriz \mathbf{RF} .

En primera instancia se analiza el método directo de radiosidad expresado en la ecuación RRB con geometría fija y luego se realiza un análisis similar con el método iterativo desarrollado en la Ecuación (3.12).

Análisis del método directo de radiosidad

En la ecuación RRB con geometría fija, la primera operación a realizar es el producto $E^{(k)} = \mathbf{V}_k^T E^{(n)}$ que transforma la emisión $E^{(n)}$ de malla fina a otra emisión de malla gruesa $E^{(k)}$. El vector $E^{(k)}$ tiene dimensión $k = 559$ y cada valor $E^{(k)}(i)$ posee el valor de emisión del i -ésimo parche de la malla de grano grueso. La emisión $E^{(n)}$ inicial es representada en la Figura 4.4 (a). Como en la figura sólo se observan los parches emisores, no se aprecian sombras ni reflejos. Sólo se aprecia un óvalo formado por muchos parches emisores que se encuentran en la pared izquierda de la escena. La emisión $E^{(k)}$ resultante del producto matricial se muestra en la Figura 4.4 (b).

El segundo producto matriz-vector de la ecuación RRB con geometría fija, es el que se realiza con la operación $-\mathbf{Y}_k E^{(k)}$, que genera el vector \tilde{C} con la reflexión final aproximada de todos los parches en malla fina. El resultado se observa en la Figura 4.4 (c). Finalmente se calcula la radiosidad aproximada \tilde{B} al realizar la suma $E + \tilde{C}$ (véase la Figura 4.4 (d)).

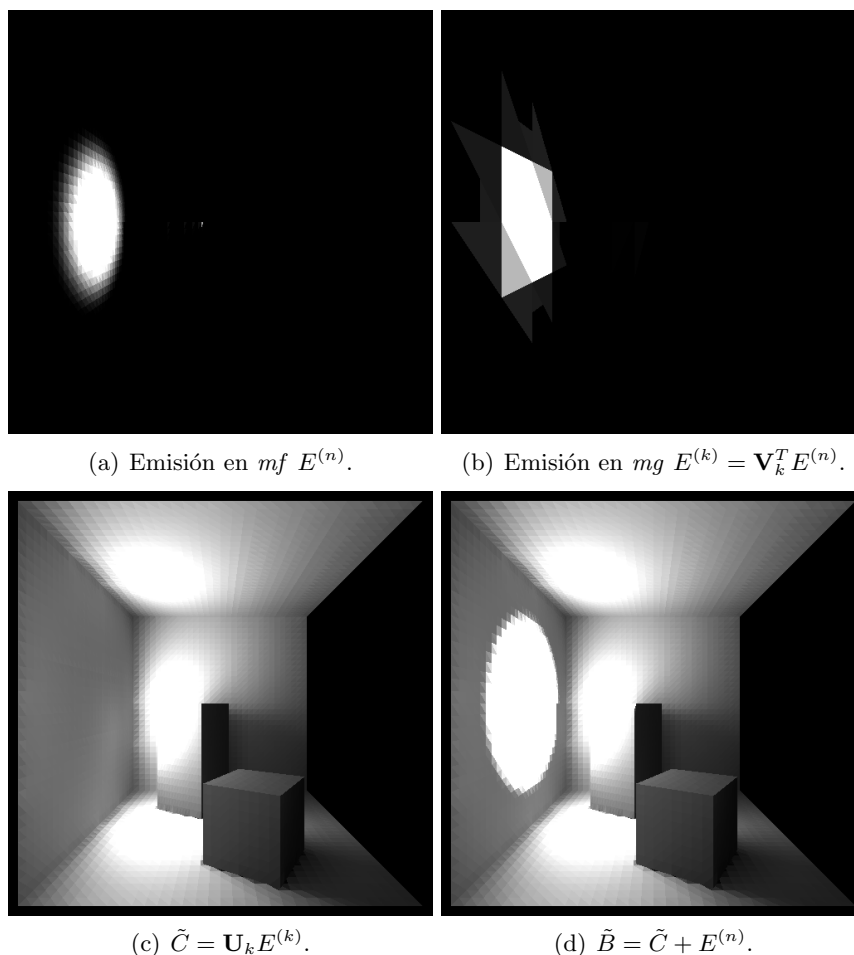


Figura 4.4: Etapas del método directo de radiosidad basado en la ecuación RRB con geometría fija, con la aproximación a **RF** basada en mallas con coherencia espacial.

Análisis del método iterativo de radiosidad

Se analiza cada paso del método desarrollado en la Ecuación (3.12), estudiando el reemplazo de la emisión en malla fina por la de malla gruesa. Se toma como base la Ecuación (3.12) mostrándose algunos pasos de la iteración.

La primera iteración comienza con la asignación de la emisión E^n como el valor de radiosidad inicial $\tilde{B}_0^{(n)}$. Luego se realiza el producto $\tilde{B}_0^{(k)} = \mathbf{V}_k^T \tilde{B}_0^{(n)}$, con el que se transforma la radiosidad en la malla fina a una equivalente en malla gruesa. Las Figuras 4.5 (a) y (b) son representativas de estas operaciones. El resto de la iteración consiste en el cálculo $\tilde{B}_1^{(n)} = E^{(n)} + \mathbf{U}_k \tilde{B}_0^{(k)}$, donde se obtiene la primer aproximación a la radiosidad de la escena (Figura 4.5 (c)).

Las siguientes iteraciones consisten en repetir el proceso mencionado, para calcular en cada paso la radiosidad $\tilde{B}_{i+1}^{(n)}$ partiendo de $\tilde{B}_i^{(n)}$ (Figuras 4.5 (d) a (g)). Las iteraciones convergen a la solución de la ecuación de radiosidad de rango bajo. Las Figuras 4.5 (h) e (i) muestran la solución final de radiosidad para la malla gruesa y fina.

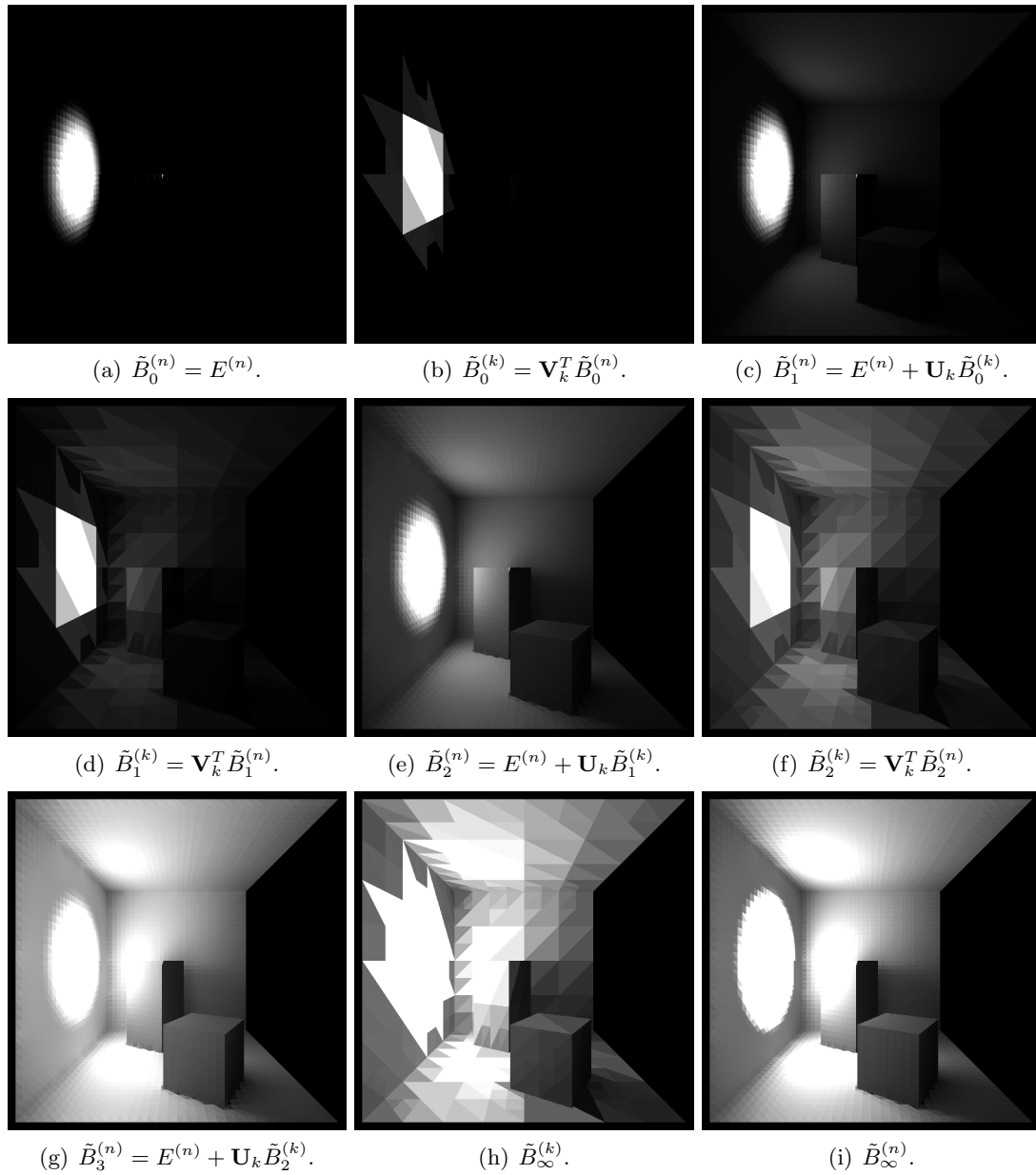


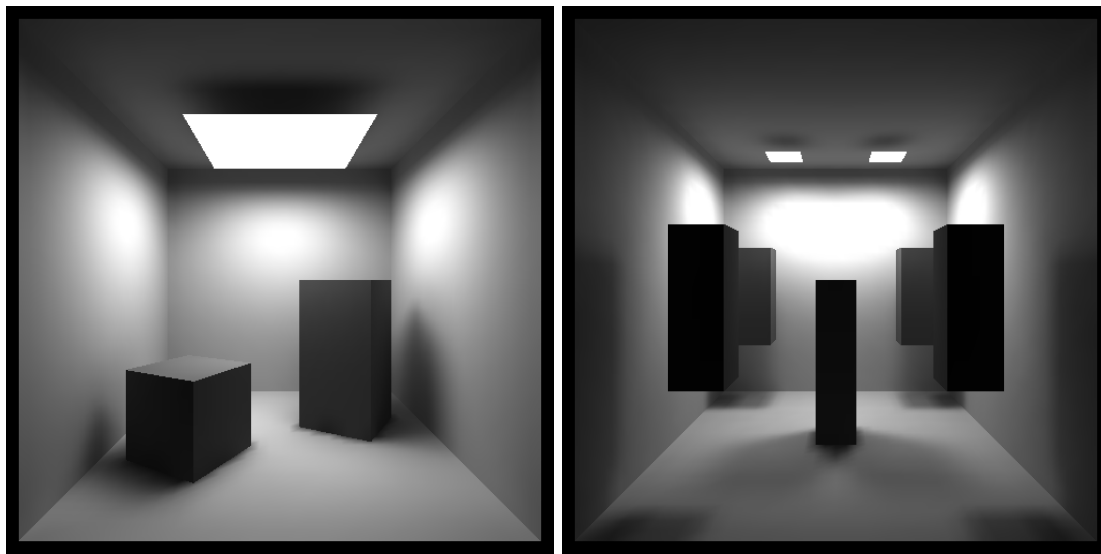
Figura 4.5: Etapas del método iterativo de radiosity. En (a) se muestra la emisión inicial y en (b) se realiza el pasaje de la emisión de malla fina a malla gruesa. En (c) se culmina la primera iteración. De (d) a (g) se realizan los pasos dos y tres de la iteración. En (h) e (i) se muestra el resultado final en malla gruesa y malla fina respectivamente.

4.3. Comparación de las factorizaciones

Para comprobar la eficiencia de los algoritmos 2MF y 2MCE, se los compara con las descomposiciones SVD y CUR estudiando su capacidad para hallar aproximaciones de bajo rango de la matriz \mathbf{F} y el error generado cuando las aproximaciones son utilizadas para calcular la radiosity \tilde{B} . Se utiliza la descomposición SVD debido a que genera las mejores aproximaciones a una matriz para un rango dado y se utiliza la descomposición

CUR porque tiene un orden de complejidad similar al de los algoritmos 2MF y 2MCE dado que las tres metodologías relevan sólo algunas filas de la matriz para sus cálculos.

Para realizar la comparación se generaron las escenas Cb1 y Cb2 (véanse las Figuras 4.6 (a) y (b)).



(a) Escena Cb1.

(b) Escena Cb2.

Figura 4.6: Escenas generadas para evaluar las factorizaciones de la matriz \mathbf{F} .

En las escenas Cb1 y Cb2 se trabajó con una cantidad de parches relativamente pequeña (3584 parches en Cb1 y 3312 parches en Cb2) para posibilitar especialmente la realización de la factorización SVD, dado que los costos en tiempo y memoria en escenas con más parches pueden ser prohibitivos.

Para cada escena se generó la ecuación matricial del problema de radiosidad con la finalidad de hallar soluciones de buena precisión, que a los efectos del análisis experimental se consideran exactas. Para cada matriz \mathbf{F} (de dimensión $n \times n$) generada a partir de las escenas se calcularon las siguientes aproximaciones:

- factorización generada utilizando el algoritmo 2MF,
- factorización generada utilizando el algoritmo 2MCE,
- factorización generada basada en la descomposición en valores singulares,
- factorización generada a partir de la descomposición CUR por el método de máximo volumen.

El objetivo principal de la comparación entre factorizaciones es evaluar la conveniencia de las factorizaciones (o aproximaciones) basadas en los algoritmos 2MF y 2MCE con respecto a otras factorizaciones para sustituir a la matriz \mathbf{F} del problema de radiosidad, permitiendo así ahorrar memoria y hallar buenas aproximaciones a la radiosidad B . Esta comparación se realiza de dos maneras: evaluando el error de los \tilde{B} hallados cuando las factorizaciones tienen igual rango, y comparando \tilde{B} cuando las factorizaciones tienen igual cantidad de información. El cálculo de \tilde{B} se realiza a partir de un mismo conjunto de emisiones E .

Cuando se comparan las factorizaciones con igual rango k , cada factorización calcula matrices \mathbf{U}_k y \mathbf{V}_k de rango k (de dimensión $n \times k$). En la comparación de factorizaciones con igual cantidad de información, para cada factorización se halla un producto de matrices cuya representación en memoria tenga aproximadamente la misma cantidad de números de punto flotante.

En el caso de la comparación para igual cantidad de información, la dimensión k de las matrices \mathbf{U}_k y \mathbf{V}_k puede variar, según la factorización utilizada. Por ejemplo, en una matriz dispersa se considera información significativa aquella que indica sólo los valores y la ubicación de los términos distintos de cero. Como las factorizaciones SVD y CUR contienen matrices densas y dado que en las factorizaciones 2MF y 2MCE hay una matriz dispersa, el manejo de cantidades iguales de información implica que la cantidad de columnas (o rango) de las matrices \mathbf{U}_k y \mathbf{V}_k será menor en SVD y CUR que en 2ME y 2MCE. En las factorizaciones SVD y CUR el rango de las matrices debe ser apenas mayor a la mitad de las columnas empleadas en las factorizaciones 2MF y 2MCE. Por lo tanto, en la comparación para igual cantidad de información, el rango de la aproximación a \mathbf{F} basado en las factorizaciones SVD y CUR es aproximadamente la mitad del rango de las factorizaciones 2MF y 2MCE.

En base a estos conceptos, se generaron las factorizaciones indicadas por “Sí” en los cuadros correspondientes de la Tabla 4.1.

| Modelo | n | Rango | | Factorizaciones | | | |
|--------|------|-----------------|-----|-----------------|-----|---------------|---------------|
| | | | | SVD | CUR | 2MF | 2MCE |
| Cb1 | 3584 | $(\approx k/2)$ | 114 | Sí | Sí | No | No |
| | | k | 224 | Sí | Sí | Sí | Sí |
| | | $(\approx k/2)$ | 450 | Sí | Sí | No | No |
| | | k | 896 | Sí | Sí | Sí | Sí |
| Cb2 | 3312 | $(\approx k/2)$ | 116 | Sí | Sí | No | No |
| | | k | 228 | Sí | Sí | Sí | Sí |
| | | $(\approx k/2)$ | 416 | Sí | Sí | No | No |
| | | k | 828 | Sí | Sí | Sí | Sí |

Tabla 4.1: Factorizaciones realizadas para evaluar la conveniencia de 2MF y 2MCE con respecto a SVD y CUR.

La Tabla 4.1 indica cuáles factorizaciones fueron evaluadas considerando distintos rangos (valores de k) y diferentes metodologías (SVD, CUR, 2MF y 2MCE) para las escenas Cb1 y Cb2. Se realizaron factorizaciones con todas las metodologías para los rangos 224 y 896 en la escena Cb1, y para los rangos 228 y 828 en la escena Cb2. Estas factorizaciones permiten comparar los errores en el cálculo de \tilde{B} para todas las factorizaciones cuando éstas comparten un mismo rango. Además, se realizaron factorizaciones en SVD y CUR para los rangos 114 y 450 en la escena Cb1, y para los rangos 116 y 416 en la escena Cb2. Estas factorizaciones son utilizadas para realizar comparaciones de los errores en el cálculo de \tilde{B} cuando las factorizaciones comparten aproximadamente el mismo volumen de información. Como se mencionara en el párrafo anterior, a igual volumen de información el rango de las factorizaciones SVD y CUR debe ser aproxi-

madamente la mitad del rango de las factorizaciones 2ME y 2MCE. Por lo tanto, para estudiar los errores entre factorizaciones con igual volumen de información para la escena Cb1 se comparan entre sí SVD y CUR con rango 114 y 2MF y 2MCE con rango 224, y también se comparan entre sí SVD y CUR con rango 450 y 2MF y 2MCE con rango 896. Un argumento similar se utiliza para las comparaciones de la escena Cb2, con los rangos correspondientes.

Los valores de k elegidos para las escenas Cb1 y Cb2 surgen de las factorizaciones 2MF y 2MCE, donde las mallas utilizadas deben ser compatibles (para el caso 2MCE también deben ser mallas con coherencia espacial) y por lo tanto, donde los valores de k son un resultado de las factorizaciones. En consecuencia, en primer instancia se realizaron las factorizaciones 2MF y 2MCE, que determinaron los rangos k , y con esos valores se calcularon las factorizaciones SVD y CUR de rango k y rango ($\approx k/2$).

4.3.1. Generación de las factorizaciones surgidas de 2MF y 2MCE

En la factorización 2MF la cantidad de parches de la malla de grano fino es por lo general un múltiplo de la cantidad de parches de la malla de grano grueso. En los experimentos realizados se utilizaron valores de k iguales a $\frac{n}{4}$ y a $\frac{n}{16}$. Luego de fijar estos valores de k , se estuvo en condiciones de hallar las otras factorizaciones.

Para el cálculo de una factorización 2MCE se parte de una malla gruesa relativamente uniforme y otra fina tal que ambas sean compatibles entre sí, y se decide qué parches de la malla gruesa conviene dividir para lograr que ambas mallas tengan coherencia espacial. Para el proceso de evaluación de parches se utiliza alguna de las métricas señaladas en la sección 3.6.2. El parche se divide si la métrica da un valor que se encuentra dentro de un intervalo determinado, con el objetivo de que el valor de k sea el mismo que el del algoritmo 2MF. Para lograrlo, se realiza un proceso iterativo donde se ajusta el intervalo para llegar al valor de k preestablecido.

4.3.2. Generación de las factorizaciones SVD

Como se trabaja con matrices \mathbf{F} pequeñas, para calcular la factorización SVD se utilizó el comando `svd` de Matlab, que utiliza la biblioteca LAPACK [1].

El comando `[U,D,V]=svd(F)` genera tres matrices resultantes de la factorización y de dimensión $n \times n$. Para hallar aproximaciones a \mathbf{F} de rango k se toman los k mayores valores singulares y las k primeras columnas de \mathbf{U} y \mathbf{V} , mediante los comandos

$$U_{svdk} = U(:, 1:k); \quad D_{svdk} = D(1:k, 1:k); \quad V_{svdk} = V(:, 1:k);$$

Por último, se calcula el producto $U_{svdk} = U_{svdk} * D_{svdk}$; para reducir la factorización a dos matrices densas.

4.3.3. Generación de las factorizaciones CUR

Como se trabaja con matrices relativamente pequeñas, se toma la matriz \mathbf{F} y se aplica el algoritmo de volumen máximo de Goreinov y Tyrtyshnikov [32], para obtener la matriz que maximiza el determinante de la matriz \mathbf{W} .

4.3.4. Generación de las emisiones de prueba

El vector E en el sistema lineal debe tener valores no negativos, por contener la emisión de radiosidad de todos los parches. Se podrían realizar experimentos generando un conjunto aleatorio de vectores de prueba E_j que se empleen en los cálculos de \tilde{B}_j y también en el cálculo exacto de B_j . El problema de este enfoque es que un vector aleatorio de emisión genera reflexiones poco intuitivas al observador humano y por lo tanto de difícil comprensión. Para evitar este problema se generaron 70 vectores de emisión para cada escena, equivalentes a la iluminación con una linterna de un recorrido a través de la escena.

Para la generación de los vectores de emisión se utiliza una proyección de perspectiva donde el centro de proyección atraviesa la escena. Se relevan los parches de emisión proyectados sobre el plano de proyección y se cuenta la cantidad de píxeles que ocupa cada parche en la proyección. La cantidad de píxeles correspondientes a cada parche se encuentra en relación directa con la potencia luminosa que recibe cada parche de la linterna. En la Figura 4.7 se muestran algunas de las imágenes generadas.

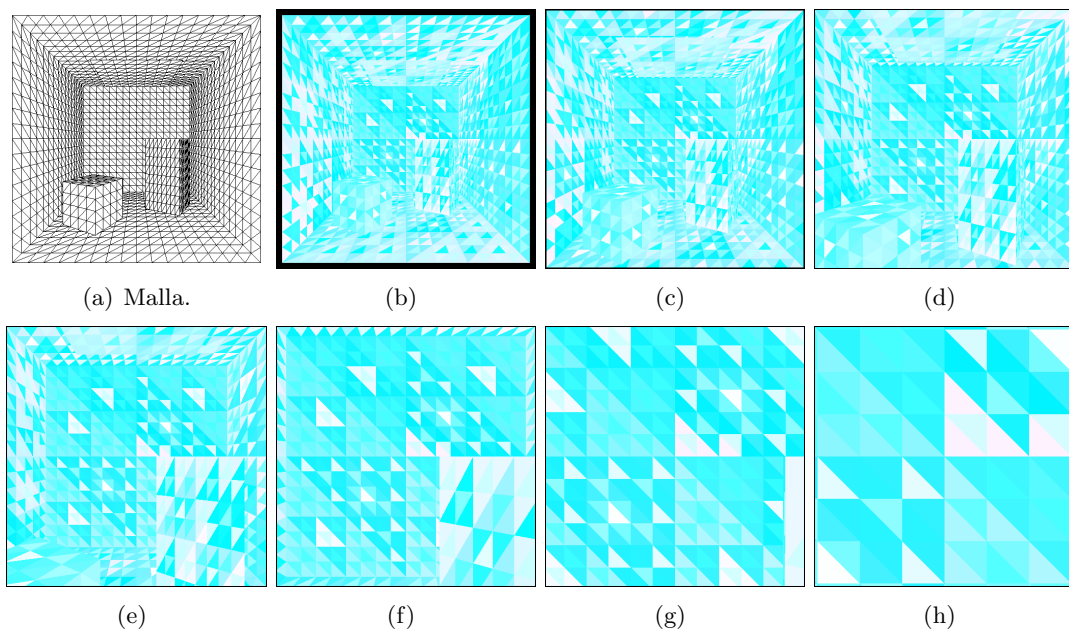


Figura 4.7: En (a) se muestra una vista de la escena. El resto de las figuras muestran distintas vistas de la escena, durante un recorrido que atraviesa la escena. Cada parche tiene un color único que permite generar un vector E por imagen.

En la Figura 4.7 cada imagen representa lo que se ve desde una luz *spot* que ilumina una escena, y en ellas se observa un conjunto de parches iluminados, donde cada parche está representado con un color identificador. Un algoritmo cuenta los píxeles de cada color que se ven en la imagen. La cantidad de píxeles en cada color está en relación directa con el valor de la emisión E de sus parches. Por ejemplo, si el color c_1 se encuentra en 100 píxeles y el color c_2 se encuentra en 20 píxeles, entonces el parche asociado a c_1 emite cinco veces más luz que el parche asociado a c_2 .

4.3.5. Cálculo del error relativo de radiosidad

Para cada factorización f y para cada vector de emisión E_j se calcula un vector de radiosidad \tilde{B}_{fj} , utilizando la Ecuación (3.21). Los vectores \tilde{B}_{fj} se comparan con la solución exacta para cada emisión B_j , calculándose el error relativo en norma Frobenius

$$\epsilon_{fj} = \frac{\|\tilde{B}_{fj} - B_j\|_{Fro}}{\|B_j\|_{Fro}} \quad (4.3)$$

Como métrica para evaluar el error relativo de cada factorización se emplea el promedio de los errores asociados a cada vector de emisión

$$\epsilon_f = \frac{\sum_j \frac{\|B_{fj} - B_j\|_{Fro}}{\|B_j\|_{Fro}}}{|j|} \quad (4.4)$$

4.3.6. Resultados obtenidos

La Tabla 4.2 presenta los resultados obtenidos en el cálculo de los errores relativos de radiosidad al aplicar cada factorización a las dos escenas Cb1 y Cb2, calculando el promedio de los errores para los vectores de emisión.

| Modelo | n | Rango | Factorizaciones | | | | |
|--------|------|-----------------|-----------------|--------|--------|-------------------|-------------------|
| | | | SVD | CUR | 2MF | 2MCE | |
| Cb1 | 3584 | $(\approx k/2)$ | 114 | 0,1001 | 0,3394 | 0,0291 | 0,0374 |
| | | k | 224 | 0,0351 | 0,3166 | 0,0291 | 0,0374 |
| | | $(\approx k/2)$ | 450 | 0,0189 | 0,2703 | 0,0112 | 0,0179 |
| | | k | 896 | 0,0024 | 0,1613 | 0,0112 | 0,0179 |
| Cb2 | 3312 | $(\approx k/2)$ | 116 | 0,032 | 0,2495 | 0,027 | 0,0642 |
| | | k | 228 | 0,0244 | 0,214 | 0,027 | 0,0642 |
| | | $(\approx k/2)$ | 416 | 0,0085 | 0,1577 | 0,0096 | 0,0168 |
| | | k | 828 | 0,0012 | 0,0305 | 0,0096 | 0,0168 |

Tabla 4.2: Promedio de los errores de radiosidad para cada factorización en norma de Frobenius.

Para cada una de las factorizaciones estudiadas, el error relativo disminuye al aumentar el rango k de la aproximación a \mathbf{F} . Estos resultados son coherentes con el concepto de que aumentando el rango de las aproximaciones a \mathbf{F} aumenta la información disponible sobre \mathbf{F} y se pueden construir aproximaciones más cercanas.

Comparando las factorizaciones se aprecia que, salvo en un caso, para un mismo k la factorización SVD siempre tiene los menores valores de error. El caso anómalo ocurre en la escena Cb1, donde la factorización 2NF tiene menor error cuando $k = 224$.

En la Tabla 4.3 se evalúa la distancia en norma Frobenius entre la matriz \mathbf{F} y las diversas aproximaciones de rango bajo calculadas con las distintas factorizaciones f para diferentes rangos k y escenas s : $\|\mathbf{F} - \mathbf{U}_{fks} \mathbf{V}_{fks}^T\|_{Fro}$.

| Modelo | n | Rango | | Factorizaciones | | | |
|--------|------|-----------------|-----|-----------------|--------|----------------|----------------|
| | | | | SVD | CUR | 2MF | 2MCE |
| Cb1 | 3584 | $(\approx k/2)$ | 450 | 1,8249 | 3,6363 | N/C | N/C |
| | | k | 896 | 0,3492 | 1,5053 | 9,2625 | 10,205 |
| Cb2 | 3312 | $(\approx k/2)$ | 416 | 0,8182 | 2,3975 | N/C | N/C |
| | | k | 828 | 0,1574 | 2,338 | 3,5710 | 2,6535 |

Tabla 4.3: Distancia de las factorizaciones a las matrices \mathbf{F} evaluadas.

La Tabla 4.3 muestra que para un mismo rango k la mejor aproximación es la surgida de la descomposición SVD, seguida de la CUR. Las factorizaciones 2MF y 2MCE intercambian la tercera y cuarta posición en las escenas Cb1 y Cb2 para los k utilizados (896 y 828 respectivamente). Cuando se comparan las factorizaciones con un volumen de información similar, las distancias de las aproximaciones a la matriz \mathbf{F} mantienen el mismo orden, a pesar que el rango de las aproximaciones SVD y CUR es aproximadamente la mitad de los rangos de las aproximaciones 2MF y 2MCE.

Resulta extraño que las factorizaciones CUR generen mejores aproximaciones a \mathbf{F} que las que se generan con 2MF y 2MCE, pero que al utilizar estas aproximaciones en la ecuación de radiosidad de rango bajo, los \tilde{B} asociados a CUR posean mayor error que los asociados a 2MF y 2MCE. Se podría suponer que en un sistema lineal aproximado donde el error se encuentra en la matriz, cuanto menor es el error en los datos también es menor el error de la solución. Sin embargo, esta suposición no se cumple para estos casos. Se requiere un mayor análisis de las factorizaciones desarrolladas para sacar conclusiones sobre este comportamiento y establecer por qué no se cumplen los supuestos mencionados.

La Tabla 4.4 muestra el cociente de los errores relativos del cálculo de \tilde{B} de las factorizaciones CUR, 2MF y 2MCE dividido el error relativo de la factorización SVD.

| Escena | n | Rango | | Factorizaciones | | | |
|--------|------|-----------------|-----|-----------------|----------------|----------------|----------------|
| | | | | SVD | CUR | 2MF | 2MCE |
| Cb1 | 3584 | $(\approx k/2)$ | 114 | 1 | 3,3906 | N/C | N/C |
| | | k | 224 | N/C | N/C | 0,2907 | 0,3736 |
| | | $(\approx k/2)$ | 450 | 1 | 14,302 | N/C | N/C |
| | | k | 896 | N/C | N/C | 0,5926 | 0,9471 |
| Cb2 | 3312 | $(\approx k/2)$ | 116 | 1 | 7,7969 | N/C | N/C |
| | | k | 228 | N/C | N/C | 0,8438 | 2,0063 |
| | | $(\approx k/2)$ | 416 | 1 | 18,553 | N/C | N/C |
| | | k | 828 | N/C | N/C | 1,1294 | 1,9765 |

Tabla 4.4: Comparación de factorizaciones con aproximadamente la misma cantidad de información para una misma escena. Se comparan los errores relativos del cálculo de B en las factorizaciones CUR, 2MF y 2MCE con respecto al error en la factorización SVD con rango $\approx \frac{k}{2}$.

En la Tabla 4.4, en las factorizaciones comparadas se mantiene aproximadamente la misma cantidad de información. La comparación muestra los siguientes resultados: la factorización CUR genera errores superiores a los de la factorización SVD (el cociente de errores varía entre 3,4 y 18,6); la factorización 2MF genera errores comparables a los de la factorización SVD (el cociente de error varía entre 0,3 y 1,13); la factorización 2MCE genera errores comparables a los de la factorización SVD pero con una variabilidad un poco mayor (el cociente de error varía entre 0,37 y 2). En tres de los cuatro casos la factorización 2MF tiene errores relativos menores a los de la factorización SVD. En cuanto a la factorización 2MCE, dos de los cuatro casos poseen valores inferiores a los de la factorización SVD.

Es importante destacar que cuando se trabaja con volúmenes similares de información, el uso de la factorización 2MF para el cálculo de \tilde{B} puede obtener mejores resultados que los obtenidos cuando se utiliza la factorización SVD. Esto sucede aunque la aproximación 2MF esté más alejada de \mathbf{F} que la factorización SVD. Las Tablas 4.3 y 4.4 brindan resultados concluyentes para el caso en que $k = 896$ (y se utiliza 2MF) y $k = 450$ (y se utiliza SVD).

4.4. Evaluación de eficiencia computacional en CPU y GPU

Esta sección compara la eficiencia computacional del cálculo de radiosidad utilizando la ecuación RRB con geometría fija mediante una implementación secuencial de CPU y una versión paralela en GPU, y utilizando como software versiones de la biblioteca BLAS [51] adaptadas a cada arquitectura. Los principales objetivos del análisis experimental que se reportan en esta sección son: experimentar con una arquitectura de hardware novedosa que se relaciona directamente con el procesamiento de gráficos (GPU) y la evaluación de un algoritmo paralelo para analizar la mejora de desempeño en procura de lograr el objetivo de procesar las imágenes en tiempo real.

Las factorizaciones se asumen conocidas y en los experimentos no se evalúa la precisión de los resultados ni se consideran las ventajas de trabajar con la matriz \mathbf{V}_k dispersa.

4.4.1. Hardware y Software

El hardware utilizado es un computador con procesador Pentium Dual-Core CPU E5200 de 2.50GHz, con 2 GB de memoria RAM. El computador tiene instalada una tarjeta gráfica NVIDIA 9800 GTX+, con 512Mb de memoria. El software utilizado consiste en dos versiones de la biblioteca BLAS desarrollada para ejecutar sobre CPU y GPU.

Para el cálculo de la ecuación RRB con geometría fija se realizan dos llamadas consecutivas a la subrutina XGEMV de BLAS, que computa el producto matriz-vector. La versión para CPU utiliza la implementación de BLAS realizada por el proyecto ATLAS [72], optimizada para el equipo utilizado. El código implementado tiene la estructura que se presenta en el Algoritmo 4.

Algoritmo 4 Cálculo de \tilde{B} con BLAS en la CPU.

Cargar las matrices \mathbf{Y}_k y \mathbf{V}_k , y los vectores E en la memoria RAM

for all i de 1 a $CantE$ **do**

 Tomar un vector E

 Aplicar XGEMV dos veces para obtener el vector \tilde{B}

 Almacenar \tilde{B} en la memoria RAM

end for

En el Algoritmo 4 se cargan las matrices \mathbf{Y}_k y \mathbf{V}_k y los vectores E en la memoria RAM del equipo. Luego se ejecuta un loop tantas veces como vectores E haya, donde se calcula el \tilde{B} asociado a cada E . Una vez que se calcula un vector B , éste es almacenado en la memoria RAM.

La versión para GPU utiliza la implementación de BLAS desarrollada sobre arquitectura CUDA denominada CUBLAS [58] versión 2.0. El código implementado tiene la estructura que se presenta en el Algoritmo 5.

Algoritmo 5 Cálculo de \tilde{B} con CUBLAS en la GPU.

Cargar las matrices \mathbf{Y}_k y \mathbf{V}_k , y los vectores E en la memoria RAM

Pasar las matrices \mathbf{Y}_k y \mathbf{V}_k a la memoria de la GPU

for all i de 1 a $CantE$ **do**

 Pasar un vector E a la memoria de la GPU

 Aplicar XGEMV dos veces en la GPU para obtener el vector \tilde{B}

 Enviar el vector \tilde{B} de la memoria de la GPU a la memoria RAM

end for

En el Algoritmo 5 se cargan las matrices \mathbf{Y}_k y \mathbf{V}_k de rango bajo y los vectores E en la memoria RAM. Luego se pasan las matrices a la memoria de la GPU. A continuación se ejecuta, tantas veces como vectores E existan, un proceso que transfiera el vector E a la GPU y aplicar XGEMV para calcular el producto matriz-vector. Por último, el vector \tilde{B} resultante de los cálculos es transferido a la memoria de la CPU.

La principal diferencia entre la versión secuencial y paralela del cálculo de \tilde{B} es que la implementación de XGEMV es paralela y permite sacar provecho de los múltiples recursos de cómputo de la arquitectura GPU. El ciclo que procesa los vectores de emisión no se paraleliza para contemplar el caso de cómputo en tiempo real, donde los vectores E no son conocidos de antemano. La estrategia que aplique paralelismo a los vectores de emisión puede ser válida en otros contextos donde los vectores E se dispongan previamente, por ejemplo en el caso de generación de imágenes para una película animada.

El tiempo promedio de cálculo de \tilde{B} se determina evaluando el tiempo insumido por el ciclo que procesa los vectores E y normalizándolo respecto a la cantidad de iteraciones realizadas. Tanto la versión para CPU como la versión para GPU utilizan precisión simple en los cálculos.

4.4.2. Casos de prueba

En el diseño de los casos de prueba para evaluar los tiempos de ejecución se tomó una geometría de la escena similar a una Cornell Box y se construyeron mallas finas con distintas cantidades de elementos n . Luego se realizó la factorización 2MF de la matriz \mathbf{RF} para distintos valores de k , que son submúltiplos de n .

La ecuación RRB con geometría fija tiene una complejidad $O(nk)$ debido a que se compone de dos productos matriz-vector con complejidades $O(nk)$ y de una suma de dos vectores con una complejidad $O(n)$.

Se escogieron valores de n y k de modo que cumplan con las siguientes premisas:

- Que el producto nk abarque un rango amplio, contemplando así distintos niveles de exigencia del hardware.
- Que las combinaciones de pares n y k se puedan agrupar en conjuntos de igual producto. Con esta premisa se pretende brindar una primera aproximación a la evaluación de si hay diferencias notables en los tiempos de ejecución dentro de una misma complejidad teórica.
- Que las mismas matrices puedan ser evaluadas en los distintos equipamientos utilizados, sin que haya limitaciones de memoria ni de capacidad para realizar las operaciones expresadas en la ecuación RRB con geometría fija.

Los valores elegidos de n y k se muestran en la Tabla 4.5.

| $n \backslash k$ | 216 | 864 | 3.456 |
|------------------|-----------|-----------|-----------|
| 3.456 | <i>Sí</i> | <i>Sí</i> | <i>Sí</i> |
| 13.824 | <i>Sí</i> | <i>Sí</i> | <i>Sí</i> |
| 55.296 | <i>Sí</i> | <i>Sí</i> | <i>No</i> |
| 221.184 | <i>Sí</i> | <i>No</i> | <i>No</i> |

Tabla 4.5: Combinaciones posibles de n y k consideradas en los experimentos.

En la Tabla 4.5 se observa que ciertas combinaciones no fueron consideradas en los experimentos debido a problemas en la memoria de la tarjeta gráfica. Por ejemplo, para almacenar una matriz en simple precisión con $n = 221184$ y $k = 864$ se requiere una memoria de unos 730 Mbytes. Este valor es superior al disponible en la tarjeta gráfica.

Las diagonales compuestas por celdas rellenas con el mismo color gris son aquellas que comparten el mismo producto nk . La celda con bordes punteados indica un caso especial, en el que la factorización no aporta ninguna ventaja en memoria, pero se la consideró para así completar dos diagonales con tres celdas.

4.4.3. Resultados de eficiencia computacional

La Tabla 4.6 describe los diferentes casos de prueba generados. Por cada caso se presenta el total de parches, el total de elementos, el producto nk y la memoria usada en \mathbf{U}_k y \mathbf{V}_k .

| Casos | Total de Elementos (n) | Total de Parches (k) | Flotantes por matriz (nk) en millones | Elementos por Parche ($d=n/k$) | Espacio en Memoria GPU ($U_k + V_k$) en Mbytes |
|-------|----------------------------|--------------------------|-------------------------------------------|----------------------------------|--------------------------------------------------|
| 1 | 3.456 | 216 | 0,7 | 16 | 5,7 |
| 2 | 3.456 | 864 | 3,0 | 4 | 22,8 |
| 3 | 13.824 | 216 | 3,0 | 64 | 22,8 |
| 4 | 3.456 | 3.456 | 11,9 | 1 | 91,1 |
| 5 | 13.824 | 864 | 11,9 | 16 | 91,1 |
| 6 | 55.296 | 216 | 11,9 | 256 | 91,1 |
| 7 | 13.824 | 3.456 | 47,8 | 4 | 364,5 |
| 8 | 55.296 | 864 | 47,8 | 64 | 364,5 |
| 9 | 221.184 | 216 | 47,8 | 1.024 | 364,5 |

Tabla 4.6: Números de punto flotante y memoria utilizada en cada uno de los casos considerados en los experimentos.

La Tabla 4.7 presenta los tiempos obtenidos al evaluar las versiones para CPU y para GPU del cálculo de \tilde{B} sobre los casos de prueba estudiados. Se reportan los tiempos del cálculo de \tilde{B} en CPU y GPU, los tiempos de transferencia de E y \tilde{B} hacia y desde la GPU y el speedup de la GPU respecto a la CPU. El *speedup* evalúa la aceleración al utilizar un algoritmo paralelo, calculando el cociente entre el tiempo de ejecución de la versión secuencial y la versión paralela [25].

| Casos | Total de Elementos (n) | Total de Parches (k) | (a) Tiempo de cálculo de \tilde{B} en CPU en ms | (b) Tiempo de cálculo de \tilde{B} en GPU en ms | (c) Tiempo de transf. de E y \tilde{B} a GPU en ms | (b+c) Tiempo total de GPU en ms | (a)/(b+c) speedup CPU/GPU |
|-------|----------------------------|--------------------------|------------------------------------------------------|------------------------------------------------------|-----------------------------------------------------------|------------------------------------|------------------------------|
| 1 | 3.456 | 216 | 1,27 | 0,69 | 0,04 | 0,73 | 1,7 |
| 2 | 3.456 | 864 | 4,64 | 1,07 | 0,03 | 1,10 | 4,2 |
| 3 | 13.824 | 216 | 5,23 | 2,51 | 0,11 | 2,62 | 2,0 |
| 4 | 3.456 | 3.456 | 14,53 | 3,61 | 0,04 | 3,65 | 4,0 |
| 5 | 13.824 | 864 | 20,14 | 3,61 | 0,11 | 3,72 | 5,4 |
| 6 | 55.296 | 216 | 22,70 | 9,74 | 0,37 | 10,11 | 2,2 |
| 7 | 13.824 | 3.456 | 72,77 | 13,29 | 0,11 | 13,40 | 5,4 |
| 8 | 55.296 | 864 | 88,37 | 14,19 | 0,43 | 14,62 | 6,0 |
| 9 | 221.184 | 216 | 86,99 | 41,33 | 1,51 | 42,84 | 2,0 |

Tabla 4.7: Tiempos de ejecución para el cálculo de \tilde{B} .

Observando los resultados de la Tabla 4.7 se pueden establecer las siguientes conclusiones preliminares:

- El speedup CPU/GPU varía entre 1,7 y 6.
- El tiempo de cálculo de \tilde{B} en CPU es proporcional al producto nk . Los tiempos de ejecución son similares para cada franja nk , y entre franjas consecutivas los tiempos cambian de escala por un factor de entre tres y cinco.
- El tiempo de cálculo de \tilde{B} en GPU es más irregular. Existen dos grupos de casos:

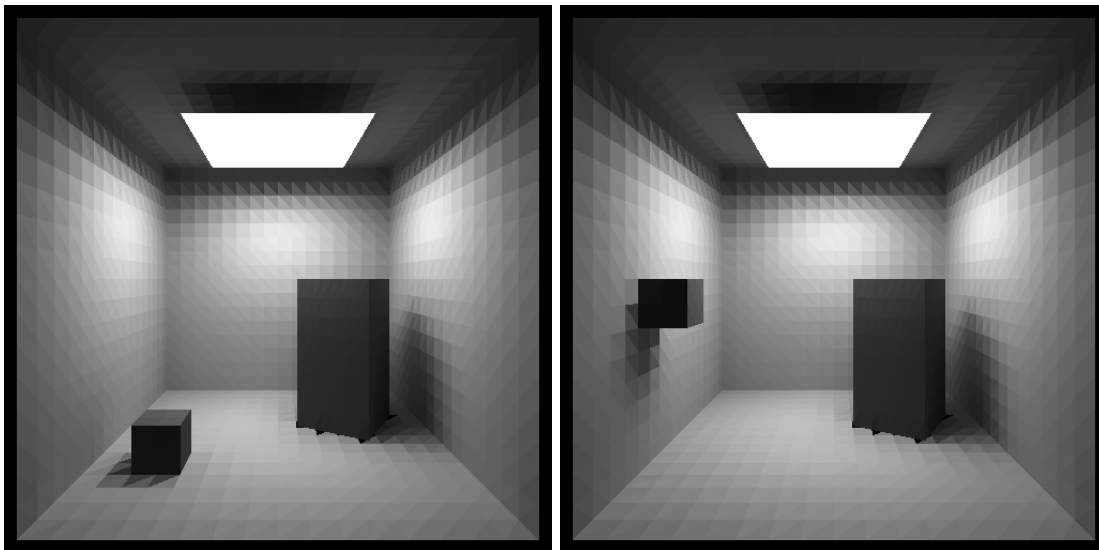
aquellos en que $k = 216$ y aquellos en que $k \neq 216$. Con la información disponible, se observa que los tiempos de ejecución son proporcionales a nk .

Para el caso en que $k = 216$, el speedup es unas tres veces inferior al otro grupo. Esta singularidad fue observada en los trabajos de Igual et al. [43] y Barrachina et al. [5], quienes plantean que las dimensiones matriciales deben ser múltiplo de 32 debido a temas relacionados con la arquitectura de las tarjetas gráficas. De todos los n y k considerados, sólo 216 no es múltiplo de 32.

- Los tiempos de transferencia de E y \tilde{B} son proporcionales a n .
- El tiempo de cálculo de \tilde{B} en la GPU es entre 10 y 120 veces mayor que el tiempo de transferencia de E y \tilde{B} .
- El cálculo y la transferencia de \tilde{B} en la versión implementada en GPU se puede realizar a 20 imágenes por segundo, y a más de 60 imágenes por segundo si no se considera el caso 9, que tiene 216 parches. Estos valores son similares a los empleados por las aplicaciones interactivas [71].

4.5. Escenas con objetos en movimiento

En esta sección se presentan resultados para el caso en que las escenas incluyen el movimiento de un objeto. El análisis experimental se realizó sobre una escena basada en una Cornell box que contiene un pequeño cubo, además de otros objetos, y se procede a mover el cubo de una posición inicial a otra posición final. Ese movimiento provoca cambios en la matriz \mathbf{F} de factores de forma, así como en sus aproximaciones de rango bajo. Imágenes con el objeto en su posición inicial y final se pueden apreciar en las Figuras 4.8 (a) y (b).



(a) Imagen con el cubo en su posición inicial.

(b) Imagen con el cubo en su posición final.

Figura 4.8: Una escena donde un cubo se mueve de una posición inicial a otra final.

A la escena con el objeto en su posición inicial le corresponde la matriz de factores de forma \mathbf{F}_I y a la escena con el objeto en su posición final le corresponde la matriz \mathbf{F}_F . La matriz diferencia es $\Delta\mathbf{F} = \mathbf{F}_F - \mathbf{F}_I$.

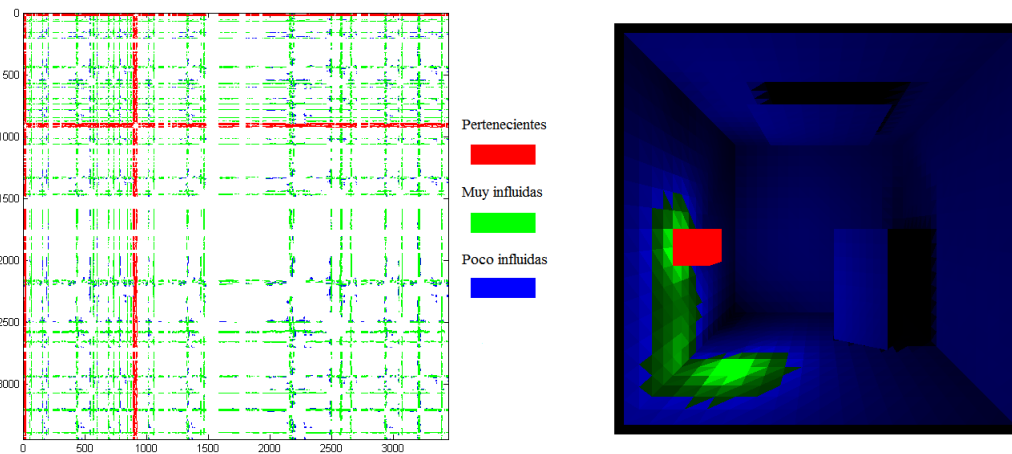
Esta tesis no propone ninguna técnica eficiente para hallar los cambios en la matriz \mathbf{F} al mover el objeto. Esta tarea se propone como trabajo futuro. En la propuesta desarrollada en esta sección, se parte de las celdas de \mathbf{F} que fueron modificadas por el movimiento del objeto y se propone una técnica para el cálculo de la iluminación a partir de la selección de aquellas filas y columnas con cambios más significativos.

Al mover un objeto en la escena, las celdas de la matriz \mathbf{F} (o $\mathbf{R}\mathbf{F}$) se pueden catalogar en cuatro tipos:

1. aquellas filas y columnas asociadas a los parches del objeto en movimiento;
2. aquellas filas y columnas asociadas a parches cercanos a la posición inicial o final del objeto, cuya visibilidad cambia debido al movimiento del objeto (muchos factores de forma de esas filas y columnas se ven afectados);
3. aquellas celdas aisladas asociadas a pares de parches cuya visibilidad es alterada por el movimiento del objeto;
4. las celdas que no sufrieron ningún cambio.

Cualquier heurística que pretenda encontrar en tiempo real las celdas que cambian a raíz del movimiento de los objetos depende en gran medida de que las celdas cambiadas sean un porcentaje bajo del total de las celdas de la matriz.

En la Figura 4.9 (a) se cataloga en rojo, verde, azul y blanco a las celdas de la matriz $\Delta\mathbf{F}$ asociadas a los tipos 1, 2, 3 y 4 respectivamente. Sólo el 3.5% de las celdas sufre modificación cuando se mueve el objeto. En la Figura 4.9 (b) se aprecia la escena con el objeto en su posición final y los parches catalogados según pertenezcan al objeto (en rojo), según hayan sido muy afectados por el movimiento del objeto (en verde), poco afectados (en azul) o afectados (negro).



(a) Celdas de $\Delta\mathbf{F}$ catalogadas según el tipo de modificación de sus celdas. (b) Parches de la escena catalogadas según el tipo de modificación de sus celdas.

Figura 4.9: La matriz $\Delta\mathbf{F}$ y los parches de la escena con el cubo en su posición final.

4.5.1. Definición de $\Delta(\mathbf{RF})$ y $\Delta(\widetilde{\mathbf{RF}})$

En la sección anterior se definió la matriz $\Delta\mathbf{F}$ como la resta de dos matrices de factores de forma \mathbf{F}_F y \mathbf{F}_I . Si a esta matriz se la multiplica por \mathbf{R} se obtiene una matriz que se define como $\Delta(\mathbf{RF}) = \mathbf{RF}_F - \mathbf{RF}_I$.

El cálculo de $\Delta(\mathbf{RF})$ es costoso dado que se debe disponer de las matrices de factores de forma. Si se sustituyen las matrices \mathbf{F}_F y \mathbf{F}_I por aproximaciones de rango bajo $\mathbf{U}_F\mathbf{V}_F^T$ y $\mathbf{U}_I\mathbf{V}_I^T$, se puede generar otra matriz $\Delta(\widetilde{\mathbf{RF}}) = \mathbf{RU}_F\mathbf{V}_F^T - \mathbf{RU}_I\mathbf{V}_I^T$. Como resultado se obtiene una matriz de rango bajo. Si el rango de las matrices \mathbf{U}_F , \mathbf{V}_F , \mathbf{U}_I y \mathbf{V}_I es k , y el rango de la matriz diagonal \mathbf{R} es n , el rango de la matriz resultante es como máximo $2k$. Este rango máximo es mayor al esperado, porque si la matriz de factores de forma de una escena pudo ser reducida a una matriz de rango k , es de esperar que una pequeña modificación de ésta pueda ser representada con una matriz de rango mucho menor. En la Figura 4.10 se observa la matriz $\Delta(\widetilde{\mathbf{RF}})$ obtenida a través de la resta de dos factorizaciones 2MF ($\mathbf{RU}_{2MF,F}\mathbf{V}_{2MF,F}^T - \mathbf{RU}_{2MF,I}\mathbf{V}_{2MF,I}^T$), con $k = 860$ en ambos casos.

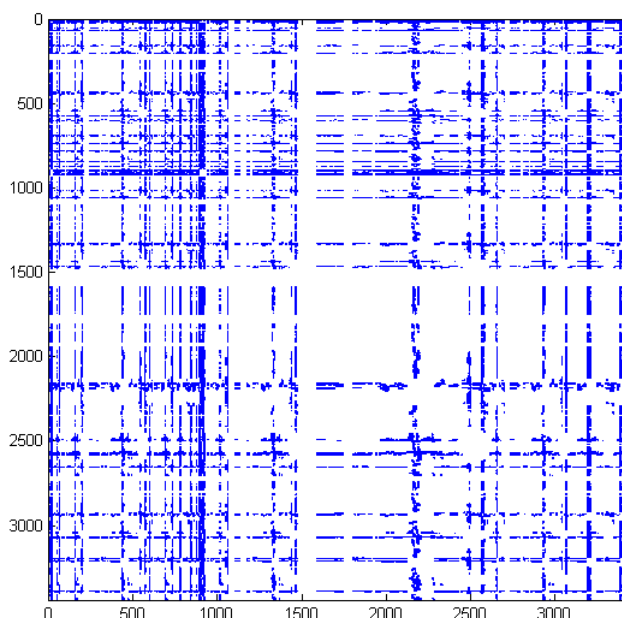


Figura 4.10: Matriz $\Delta(\widetilde{\mathbf{RF}})$ que surge de la resta de dos aproximaciones de rango $k = 860$ $\mathbf{RU}_F\mathbf{V}_F^T - \mathbf{RU}_I\mathbf{V}_I^T$.

Por lo tanto, independientemente de si se utiliza $\Delta(\mathbf{RF})$ o $\Delta(\widetilde{\mathbf{RF}})$, estas matrices tendrán algunas filas y columnas con muchos elementos distintos de cero (las asociadas a las celdas de tipos 1 y 2) y muchas otras filas y columnas con pocos elementos distintos de cero (las asociadas a las celdas de tipo 3 y 4). Una matriz es la aproximación de la otra y en consecuencia se las considerará en las próximas subsecciones como la misma matriz, y se las agrupará bajo la expresión $\Delta(\mathbf{RF})$, independientemente de cual sea su origen.

4.5.2. Reducción del rango de $\Delta(\mathbf{RF})$

En esta subsección se expone un método que aproxima a la matriz $\Delta(\mathbf{RF})$ por otra que sólo incluye las celdas de tipos 1 y 2 según la clasificación presentada previamente. Este procedimiento tiene sentido dado que las celdas tipo 1 y 2 contienen a la mayoría de las celdas distintas de cero de la matriz $\Delta(\mathbf{RF})$, y estas celdas se agrupan convenientemente en unas pocas filas y columnas (véase la Figura 4.9 (a)). Por lo tanto, se parte de la matriz $\Delta(\mathbf{RF})$ e independientemente de cómo haya sido generada, se pueden crear dos nuevas matrices \mathbf{U}_Δ y \mathbf{V}_Δ tal que su producto contenga las filas y columnas de mayor norma de la matriz $\Delta(\mathbf{RF})$.

El Algoritmo 6 presenta el pseudocódigo del método propuesto para el cálculo aproximado de $\Delta(\mathbf{RF})$.

Algoritmo 6 Cálculo de \mathbf{U}_Δ y \mathbf{V}_Δ .

```

----- Cálculo de  $\mathbf{U}_{\Delta c}$  y  $\mathbf{V}_{\Delta c}$  -----
1  Calcular índices  $I$  de ubicación de las  $k_\Delta$  mayores columnas de  $\Delta\mathbf{F}$ 
2  Hallar  $\mathbf{U}_{\Delta c}$  con las columnas  $I$  de  $\Delta\mathbf{F}$ 
3  Crear la matriz  $\mathbf{V}_{\Delta c}$  de dimensión  $n \times k_\Delta$  e inicializarla con ceros
4  Para cada  $i$  entre 1 y  $k_\Delta$  hacer
5       $\mathbf{V}_{\Delta c}(i, I(i)) = 1$ 
6  fin

----- Cálculo de  $\mathbf{U}_{\Delta r}$  y  $\mathbf{V}_{\Delta r}$  -----
7  Realizar  $\Delta\mathbf{F} = \Delta\mathbf{F} - \mathbf{U}_{\Delta c}\mathbf{V}_{\Delta c}^T$ 
8  Calcular índices  $I$  de ubicación de las  $k_\Delta$  mayores filas de  $\Delta\mathbf{F}$ 
9  Hallar  $\mathbf{V}_{\Delta r}$  con las filas  $I$  de  $\Delta\mathbf{F}$ 
10 Crear la matriz  $\mathbf{U}_{\Delta r}$  de dimensión  $n \times k_\Delta$  e inicializarla con ceros
11 Para cada  $i$  entre 1 y  $k_\Delta$  hacer
12      $\mathbf{U}_{\Delta r}(i, I(i)) = 1$ 
13 fin

----- Cálculo de  $\mathbf{U}_\Delta$  y  $\mathbf{V}_\Delta$  -----
14 Hallar  $\mathbf{U}_\Delta$  concatenando las columnas de  $\mathbf{U}_{\Delta c}$  y  $\mathbf{U}_{\Delta r}$ 
15 Hallar  $\mathbf{V}_\Delta$  concatenando las columnas de  $\mathbf{V}_{\Delta c}$  y  $\mathbf{V}_{\Delta r}$ 

```

En el Algoritmo 6, en primer instancia se generan las matrices $\mathbf{U}_{\Delta c}$ y $\mathbf{V}_{\Delta c}$, cuyo producto contiene a las columnas de mayor norma de la matriz $\Delta\mathbf{F}$. A continuación se eligen las k_Δ columnas con mayor norma y se almacenan en $\mathbf{U}_{\Delta c}$. La matriz $\mathbf{V}_{\Delta c}$ es dispersa y almacena información sobre la posición que ocupa cada columna de $\mathbf{U}_{\Delta c}$ en la matriz original $\Delta(\mathbf{RF})$. En la Figura 4.11 (a), se observa un ejemplo de las matrices $\mathbf{U}_{\Delta c}$ y $\mathbf{V}_{\Delta c}$. En segunda instancia se calculan las matrices $\mathbf{U}_{\Delta r}$ y $\mathbf{V}_{\Delta r}$, cuyo producto contiene a las filas de mayor norma de la matriz $\Delta\mathbf{F} - \mathbf{U}_{\Delta c}\mathbf{V}_{\Delta c}^T$ (véase la Figura 4.11 (b)). Por último, se agrupan las matrices halladas obteniéndose \mathbf{U}_Δ y \mathbf{V}_Δ , cuyo producto genera la aproximación de rango k_Δ a la matriz $\Delta\mathbf{RF}$ (véase la Figura 4.11 (c)), de acuerdo a la expresión de la Ecuación (4.5).

$$\widetilde{\Delta\mathbf{RF}} = \mathbf{U}_{\Delta c}\mathbf{V}_{\Delta c}^T + \mathbf{U}_{\Delta r}\mathbf{V}_{\Delta r}^T = (\mathbf{U}_{\Delta c}|\mathbf{U}_{\Delta r})(\mathbf{V}_{\Delta c}|\mathbf{V}_{\Delta r})^T = \mathbf{U}_\Delta\mathbf{V}_\Delta^T \quad (4.5)$$

En el ejemplo de la Figura 4.11, la matriz $\Delta\mathbf{RF}$ tiene 420722 elementos distintos de cero y de ellos 358908 elementos (el 85 %) son incluidos en $\widetilde{\Delta\mathbf{RF}}$ para un $k_\Delta = 200$.

La elección de k_Δ depende de la memoria disponible y de la precisión buscada. En la próxima subsección se analiza cómo el no contemplar a todos los elementos distintos de cero de $\Delta\mathbf{RF}$ puede alterar sensiblemente a las sombras generadas por los objetos, en especial cuando son sombras alargadas.

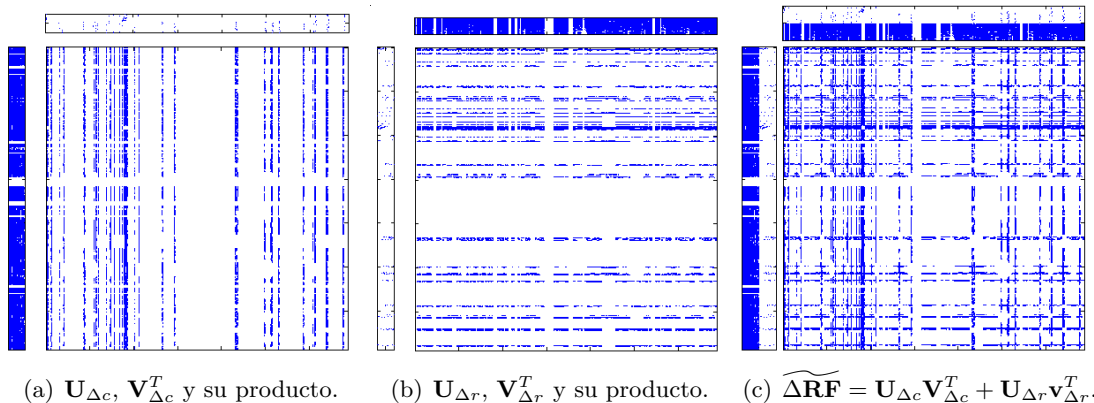


Figura 4.11: Generación de la matriz aproximada a $\Delta\mathbf{RF}$ a partir del producto de matrices de rango bajo.

4.5.3. Cálculo de \hat{B}_F luego de conocer $\widetilde{\Delta(\mathbf{RF})}$

Para calcular la radiosidad B_F , la matriz de radiosidad es sustituida por alguna aproximación que facilite el cómputo y ahorre memoria. Si se dispone de \mathbf{U}_F y \mathbf{V}_F , en lugar de calcular B_F se calcula una aproximación \tilde{B}_F . Esta aproximación se halla mediante la aplicación directa de la Ecuación (3.21). Si no se dispone de \mathbf{U}_F ni de \mathbf{V}_F , pero sí se dispone de \mathbf{U}_I , \mathbf{V}_I , \mathbf{U}_Δ y de \mathbf{V}_Δ , se calcularía una aproximación diferente a \tilde{B}_F , que llamaremos \hat{B}_F . El cálculo de \hat{B}_F puede ser realizado a través de la ecuación de Sherman-Morrison-Woodbury, según su formulación en las Ecuaciones (3.42) y (3.43).

4.5.4. Algunos resultados del cálculo de \hat{B}_F

En la Figura 4.12 se ejemplifica el resultado final del cálculo de B_F exacto y los cálculos aproximados de \tilde{B}_F y de \hat{B}_F , así como el error en la radiosidad de cada parche. La Figura 4.12 muestra diferentes resultados de la radiosidad de la escena, cuando el objeto se encuentra en la posición final. En (a) se muestra el resultado exacto, generado con el uso de la ecuación matricial de radiosidad y la matriz \mathbf{F} . En (b) y (d), se muestran los resultados de radiosidad, cuando la matriz \mathbf{F} es reemplazada por las matrices $\mathbf{U}_F\mathbf{V}_F^T$ y $\mathbf{U}_I\mathbf{V}_I^T + \mathbf{U}_\Delta\mathbf{V}_\Delta^T$ respectivamente. Por último, en (c) y (e) se observan las diferencias entre la radiosidad real y la aproximada para cada parche.

La Figura 4.12 muestra que al calcular \hat{B}_F , a los errores de \tilde{B}_F se agregan los errores de las sombras en las posiciones inicial y final del objeto, provocados por los elementos de $\Delta(\mathbf{RF})$ que no fueron considerados para el cálculo de $\mathbf{U}_\Delta\mathbf{V}_\Delta^T$. Cuando la iluminación genera sombras alargadas en los objetos en movimiento, el error de \hat{B}_f es más evidente, dado que las sombras pueden extenderse sobre parches cuyas filas y columnas no han

sido consideradas en la aproximación a $\Delta(\mathbf{RF})$. En (d) se observa que la sombra del cubo está truncada. En (e) se observan zonas blancas en la parte inferior izquierda de la imagen, indicando una diferencia importante en la radiosidad de algunos parches alejados del cubo, sobre los que se debe proyectar su sombra. La magnitud de estas diferencias dependerá del valor k_Δ elegido.

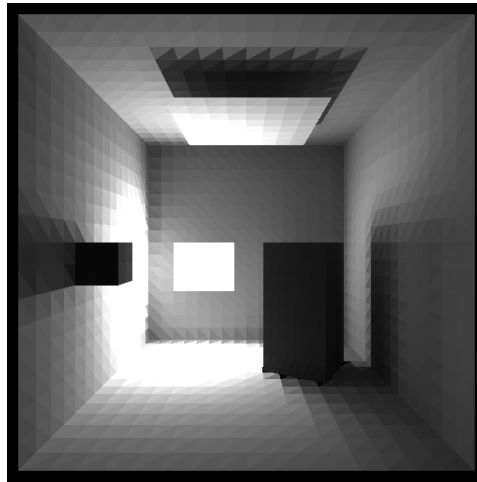
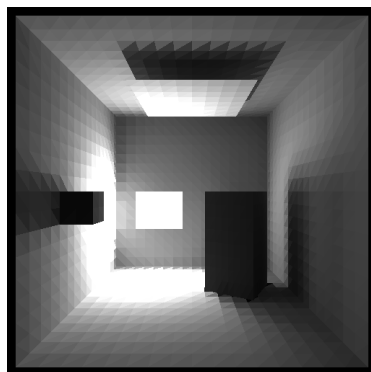
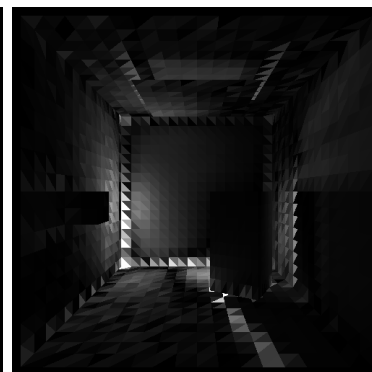
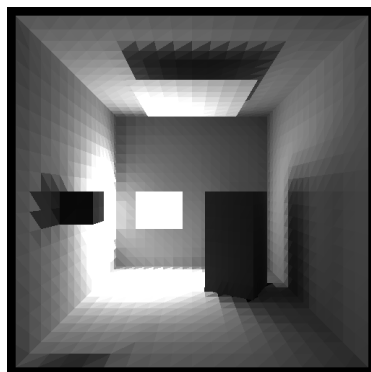
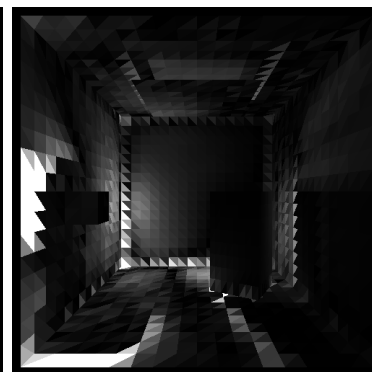
(a) B_f .(b) \tilde{B}_f .(c) $20 * |\tilde{B}_f - B_f|$.(d) \hat{B}_f .(e) $20 * |\hat{B}_f - B_f|$.

Figura 4.12: Cálculo de B_f , \tilde{B}_f y \hat{B}_f para una iluminación dada.

Capítulo 5

Conclusiones y trabajo futuro

Esta sección presenta las conclusiones del trabajo de investigación relativo a la resolución del problema de radiosidad utilizando matrices de rango bajo. También se formulan las principales líneas de trabajo futuro que surgieron durante el desarrollo de esta tesis.

5.1. Conclusiones

Esta tesis investigó la resolución del problema de radiosidad utilizando matrices de rango bajo. En las últimas tres décadas ha habido un avance importante en los algoritmos de iluminación global, durante el cual las técnicas desarrolladas para la resolución de la ecuación de radiosidad han tenido una importancia sustantiva. Entre los objetivos actuales en el área de computación gráfica se encuentra el desarrollo de aplicaciones interactivas en tiempo real que resuelvan problemas de iluminación global. Las soluciones propuestas al problema de radiosidad requieren volúmenes de memoria importantes y los algoritmos desarrollados están basados principalmente en métodos iterativos computacionalmente costosos.

En este contexto, se abordó el problema de aproximar las matrices \mathbf{RF} en la ecuación matricial de radiosidad utilizando matrices de rango bajo, con los objetivos de lograr ahorros importantes en el uso de memoria y procurar resolver el sistema lineal en tiempo real.

Con respecto a las posibilidades de generar buenas aproximaciones a la matriz \mathbf{RF} , se estudiaron sus propiedades utilizando el concepto de coherencia espacial en las escenas, llegándose a la conclusión de que para la mayoría de las escenas las matrices asociadas poseen rango numérico bajo.

Para aproximar la matriz \mathbf{RF} mediante una matriz de rango bajo se comenzó realizando un relevamiento de las técnicas existentes para la creación de aproximaciones de rango bajo. En particular, se relevaron las descomposiciones SVD y CUR, el uso de las transformadas discretas de Fourier y aquellas basadas en wavelets. El relevamiento permitió concluir que todas las técnicas estudiadas (salvo CUR) poseen deficiencias serias en cuanto al volumen de memoria necesaria y a los órdenes de los tiempos de cálculo de las aproximaciones.

A continuación, se propuso la ecuación de radiosidad de rango bajo como alternativa a la ecuación matricial de radiosidad y se desarrollaron métodos para su resolución. Se propusieron métodos iterativos y directos para escenas en blanco y negro, para escenas en colores y escenas con objetos en movimiento. Todos estos métodos poseen complejidad

$O(nk)$ en al menos una de sus variantes (iterativo o directo) y utilizan un volumen de memoria $O(nk)$, siendo n el número de elementos de la escena y k el rango de la aproximación a **RF**. Además, se llevó a cabo un estudio del error en el cálculo de la radiosidad cuando se utiliza la ecuación de radiosidad de rango bajo en lugar de la ecuación matricial de radiosidad. Los resultados indicaron que el error relativo en el cálculo de la radiosidad es linealmente proporcional a la distancia entre **RF** y su aproximación de rango bajo.

Por último, se abordó el problema de construcción de aproximaciones de rango bajo específicas para la matriz **RF**, encontrándose que en el problema de radiosidad la reflexión de los parches es muy poco sensible a los cambios en la emisión. Este problema se debe a la coherencia espacial y queda explicitado al desarrollar la ecuación matricial de la radiosidad reflejada. Esta observación se puede aprovechar para la construcción de aproximaciones de rango bajo a la matriz **RF**. En consecuencia, se desarrolló una metodología basada en la generación de dos mallas: una malla gruesa encargada de simplificar las emisiones de las escenas y una malla fina utilizada para modelar la reflexión de la escena. Se definieron los conceptos de mallas compatibles, parche con coherencia espacial y mallas con coherencia espacial, a partir de los cuales se desarrollaron los algoritmos 2MF y 2MCE para la construcción de aproximaciones de rango bajo de **RF**. El algoritmo 2MF presupone que las mallas fina y gruesa empleadas poseen coherencia espacial, y permite construir la matriz \mathbf{U}_k con complejidades $O(nk)$ u $O(n^2)$ según la variante utilizada para su implementación. El volumen de memoria necesario para el algoritmo 2MF es $O(nk)$. Por otra parte, el algoritmo 2MCE construye las mallas finas y gruesas para que posean coherencia espacial. Este algoritmo posee una complejidad aproximada $O(n^2)$ y utiliza una memoria $O(n^2)$.

El análisis experimental realizado en este trabajo de investigación se enfocó en cuatro aspectos principales: (i) el estudio de la coherencia espacial de las escenas y su influencia en la emisión y reflexión de sus superficies, (ii) en la comparación del error en la radiosidad calculada dependiendo de la factorización utilizada, (iii) en la comparación de la eficiencia computacional de implementaciones sobre CPU y GPU del cálculo de radiosidad basado en la ecuación RRB con geometría fija, y (iv) el estudio de las matrices resultantes de modelar escenas que incluyen objetos en movimiento.

Respecto al análisis de la coherencia espacial de las escenas, los resultados experimentales confirman que la reflexión de la escena es muy poco sensible (relación de 20 a 1) a las perturbaciones en la emisión. El problema inverso, hallar la emisión a partir de la reflexión, es extremadamente sensible a las perturbaciones en los datos, debido al rango bajo de la matriz **RF**. Estas observaciones confirman que la utilización de dos mallas, una gruesa para modelar las emisiones y una fina para modelar las reflexiones, es una solución válida para resolver el problema de radiosidad.

Respecto al análisis del error de los valores calculados con la ecuación de radiosidad de rango bajo respecto a la solución exacta de la ecuación matricial de radiosidad, los resultados experimentales se detallan a continuación. Si se comparan ecuaciones de rango bajo con matrices de igual rango, las mejores aproximaciones a B se obtienen con las ecuaciones que utilizan aproximaciones basadas en SVD, seguidas de aquellas que utilizan 2MF, 2MCE y CUR. Más destacables resultan los resultados si se comparan ecuaciones de rango bajo que contienen matrices con aproximadamente el mismo volumen de información (la importancia de esta comparación radica en que 2MF y 2MCE generan siempre matrices \mathbf{V}_k dispersas mientras que SVD y CUR no lo hacen). En este

caso, las mejores aproximaciones a B se obtienen con las aproximaciones 2MF en 3 de las 4 instancias experimentales, seguido de SVD, 2MCE y CUR. Por último, se realizaron comparaciones de la distancia entre las aproximaciones de rango bajo y las matrices \mathbf{RF} y los mejores resultados (sin considerar SVD, dado que se puede demostrar que genera los resultados óptimos) se obtienen por la aproximación basada en la descomposición CUR. Esta última conclusión es reveladora de que no alcanza con disponer de métodos que generen buenas aproximaciones a la matriz \mathbf{RF} para hallar buenas soluciones al problema de radiosidad, dado que CUR produce buenas aproximaciones a \mathbf{RF} que generan errores grandes en el cálculo de la ecuación de radiosidad de rango bajo.

Respecto al análisis de la eficiencia computacional de los métodos de radiosidad de rango bajo implementados en CPU y GPU para resolver escenas con geometría fija, los resultados indican que es posible lograr tiempos menores a los 15 ms en GPU (correspondientes a procesar 67 imágenes por segundo) para geometrías de 55296 elementos y 864 parches. En escenas con hasta 221184 elementos y 216 parches los tiempos llegaron a los 45 ms en GPU (correspondientes a procesar 22 imágenes por segundo). El speedup de la versión implementada en GPU respecto a la versión en CPU tuvo valores entre 2 y 6 para las matrices utilizadas. En ambas arquitecturas se realizaron cálculos de simple precisión lograndose resultados de calidad indistinguible de los que se pueden obtener con doble precisión. Un resultado no esperado generado durante la etapa de análisis fue la comprobación de que la implementación de la radiosidad de rango bajo con geometría fija es extremadamente sencilla en CPU y GPU si se utiliza la biblioteca BLAS.

Respecto al análisis de las matrices resultantes para una escena que incluye objetos en movimiento, los resultados indican que al mover un objeto los cambios en la matriz \mathbf{RF} se concentran en unas pocas filas y columnas, y que esos cambios se pueden aproximar por matrices de rango bajo. Se desarrolló una metodología para generar la matriz que contiene los cambios y se evaluaron las ecuaciones desarrolladas para el cálculo de \mathbf{B} en escenarios que contienen objetos en movimiento, obteniendose resultados satisfactorios.

En resumen, las principales contribuciones de la investigación que se reporta en esta tesis son:

- El avance en el estudio del rango de las matrices \mathbf{F} y \mathbf{RF} y de las consecuencias que un rango numérico bajo tiene en la resolución del problema de radiosidad.
- La jerarquización de la importancia de la coherencia espacial como causa principal del rango numérico bajo de la matriz \mathbf{RF} y de la insensibilidad de la reflexión de la escena a las perturbaciones en la emisión que la posicionan como una propiedad deseable para las mallas de una escena.
- La propuesta e implementación de varios algoritmos (iterativos y directos) basados en el uso de aproximaciones de rango bajo a las matrices \mathbf{RF} , para el cálculo de radiosidad en diferentes contextos (escenas en blanco y negro, escenas con color, escenas con objetos en movimiento). Varios de estos algoritmos poseen complejidad $O(nk)$, con $k \ll n$, posibilitando la implementación de código eficiente para el cálculo de la radiosidad, intentando lograr el objetivo de cálculo en tiempo real.
- La verificación de que la ecuación RRB con geometría fija posibilita el desarrollo de algoritmos muy sencillos para el cálculo de la radiosidad, que utilizando arquitecturas basadas en GPU son capaces de hallar la radiosidad de una escena a velocidades compatibles con el tiempo real.

- El desarrollo de algoritmos de complejidad $O(nk)$ para el cálculo de aproximaciones de rango bajo de la matriz \mathbf{RF} que sean competitivas con las calculadas a partir de la descomposición SVD en términos de la memoria utilizada para su almacenamiento.
- El desarrollo de los conceptos de mallas compatibles, parches con coherencia espacial y mallas con coherencia espacial, utilizados para el cálculo de nuevas aproximaciones de \mathbf{RF} , y el desarrollo de un algoritmo para el cálculo de mallas con coherencia espacial.
- La verificación experimental de que los cambios en las matrices \mathbf{RF} debidos al movimiento de objetos pueden aproximarse a matrices de rango bajo.

La investigación desarrollada generó dos artículos publicados en congresos internacionales:

- **Low-Rank Radiosity** [22], Eduardo Fernández, publicado en los Proceedings of IV Iberoamerican Symposium in Computer Graphics, Isla Margarita, Venezuela, julio de 2009 (texto en inglés).

En este trabajo se expone el concepto de coherencia aplicado al problema de radiosidad y su correspondencia con el rango numérico bajo de la matriz \mathbf{RF} . Como principales aportes se deduce el método directo basado en la ecuación RRB con geometría fija y un método iterativo basado en la serie de Neumann. Además, se realiza una estimación del error relativo de radiosidad y se expone una variante del algoritmo 2MF para el cálculo de aproximaciones de rango bajo a \mathbf{RF} .

- **Implementación en GPU del algoritmo de radiosidad de rango bajo** [23], Eduardo Fernández, Pablo Ezzatti, Sergio Nesmachnow, publicado en los Anales del XVIII Congreso sobre Métodos Numéricos y sus Aplicaciones, Tandil, Argentina, noviembre de 2009.

En este trabajo se introducen los conceptos principales de la ecuación de radiosidad de rango bajo y la ecuación RRB con geometría fija. Luego se presentan implementaciones de la ecuación RRB con geometría fija en CPU y GPU, utilizando la biblioteca BLAS. Se realizan análisis experimentales comparativos de ambas implementaciones en escenas que varían de 3500 a 220000 elementos, comprobándose que la eficiencia computacional en GPU es hasta 6 veces superior a la de la implementación en CPU. El método en GPU registró tiempos de 15 ms para escenas con más de 50000 elementos, lo que posibilita el desarrollo de aplicaciones gráficas que realizan cálculos de iluminación global en tiempo real.

5.2. Trabajo futuro

A lo largo de la investigación, varios aspectos han presentado cuestiones interesantes que merecen una profundización. En este contexto, las principales líneas de trabajos futuros se comentan a continuación.

Para comparar los resultados de este trabajo con otras propuestas de radiosidad e iluminación global que también tienen como objetivo el procesamiento en tiempo real, como primer aspecto, se debe extender el análisis experimental de los algoritmos desarrollados, considerando escenas estándares desarrolladas por la comunidad académica.

Con respecto a la conveniencia de las técnicas RRB, se deben realizar estudios que muestren la relación existente entre k , n y el error relativo de \tilde{B} . Al aumentar n o k es de esperar que el error de \tilde{B} baje debido principalmente al aumento de tamaño y de la información contenida en las matrices. Debe investigarse si ésta y otras hipótesis similares son ciertas y deben cuantificarse las relaciones, intentando definir fórmulas que contemplen los resultados del análisis experimental.

Respecto al uso de mallas compatibles, para contemplar el uso de mallas irregulares se debe estudiar la sensibilidad de los resultados visuales a la variación en el tamaño de los elementos dentro de los parches.

Respecto a los algoritmos estudiados, el análisis experimental mostró que el algoritmo 2MF obtiene mejores resultados que el algoritmo 2MCE. Éste es un resultado un tanto extraño, debido a que se espera que cuando hay mallas adaptadas a la escena y con coherencia espacial se generen mejores resultados que cuando las mallas son uniformes y no contienen ninguna consideración sobre las características de la escena. Por tanto, se debe extender el análisis experimental a otras escenas para comprobar si la tendencia se mantiene y se deben estudiar formas alternativas de generación de mallas con coherencia espacial que permitan obtener mejores resultados.

Asimismo, deben realizarse análisis detallados de las características de las aproximaciones CUR que calculan los diversos algoritmos existentes, para comprender los altos valores de error obtenidos, aún cuando el algoritmo CUR genera muy buenas aproximaciones a la matriz \mathbf{F} .

Hasta el momento de publicación de esta tesis, sólo se ha utilizado a la ecuación RRB con geometría fija y matrices densas para el estudio de eficiencia computacional y speedup entre códigos implementados para GPU y CPU. Por lo tanto, se deben analizar nuevas implementaciones de la ecuación RRB con geometría fija. También se debe extender el estudio de eficiencia y speedup entre códigos para GPU y CPU a los demás resultados de la tesis, que incluyen los algoritmos de generación de aproximaciones de rango bajo y las implementaciones basadas en las demás ecuaciones desarrolladas para el cálculo de radiosidad en diferentes contextos.

En el estudio de escenas con objetos en movimiento, debe profundizarse el análisis teórico de las principales características de las matrices \mathbf{RF}_Δ con el fin de proponer algoritmos para su generación e implementaciones que puedan ejecutarse en tiempo real, contemplando el procesamiento de escenas en aplicaciones interactivas.

Por último, los métodos propuestos en este trabajo pueden extenderse a otras áreas de conocimiento distintas a la computación gráfica. Las ecuaciones de Fredholm se encuentran en múltiples áreas de la física y los modelos numéricos surgidos de ellas contienen sistemas lineales de gran dimensión cuya generación y resolución contienen problemas similares a los de la radiosidad. Por otra parte, la necesidad de analizar y manipular matrices y tensores con billones de elementos es una realidad en muchas áreas de conocimiento. Por tanto, resulta conveniente explorar si las aproximaciones calculadas con los algoritmos 2MF y 2MCE son útiles para otras comunidades científicas.

Bibliografía

- [1] E. ANDERSON, Z. BAI, C. BISCHOF, L. BLACKFORD, J. DEMMEL, J. DONGARRA, J. DU CROZ, S. HAMMARLING, A. GREENBAUM, A. MCKENNEY Y D. SORESENSEN. “LAPACK Users’guide”. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (1999).
- [2] I. ASHDOWN. Radiative transfer networks revisited. En “Proceedings of the 2001 Illuminating Engineering Society of North America Annual Conference”, páginas 431–450, New York, NY (August 2001). Illuminating Engineering Society of North America.
- [3] G. BARANOSKI, R. BRAMLEY Y P. SHIRLEY. Fast radiosity solutions for environments with high average reflectance. En “Proceedings of the Eurographics Workshop”, páginas 345–356. Springer (1995).
- [4] J. BARNES Y P. HUT. A hierarchical $O(n \log n)$ force calculation algorithm. *Nature* (324), 446–449 (1986).
- [5] S. BARRACHINA, M. CASTILLO, F. IGUAL, R. MAYO Y E. QUINTANA-ORTÍ. Evaluation and tuning of level 3 CUBLAS for graphics processors. Informe Técnico ICC 01-01-2008, Universidad Jaume I (Enero 2008).
- [6] R. BARRETT, M. BERRY, T. F. CHAN, J. DEMMEL, J. DONATO, J. DONGARRA, V. EIJKHOUT, R. POZO, C. ROMINE Y H. VAN DER VORST. “Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods”. SIAM, Philadelphia, PA (1994).
- [7] P. BEKAERT. “Hierarchical and Stochastic Algorithms for Radiosity”. Tesis Doctoral, University of Leuven (1999).
- [8] A. BERMAN Y R. PLEMMONS. “Nonnegative matrices in the mathematical sciences”. Computer science and applied mathematics. Academic Press, New York (1979).
- [9] Å. BJÖRCK. “Numerical Methods for Least Squares Problems”. SIAM (1996).
- [10] J. BLINN Y M. NEWELL. Texture and reflection in computer generated images. *Communications of the ACM* **19**(10), 542–547 (1976).
- [11] C. CAIAFA Y A. CICHOCKI. Reconstructing matrices and tensors from few vectors. En “Proceedings of the International Symposium on Nonlinear Theory and its Applications (DVD-ROM)” (2009).

- [12] E. CATMULL. “A subdivision algorithm for computer display of curved surfaces.” Tesis Doctoral, The University of Utah (1974).
- [13] M. COHEN, S. CHEN, J. WALLACE Y D. GREENBERG. A progressive refinement approach to fast radiosity image generation. En “Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques”, páginas 75–84 (1988).
- [14] M. COHEN Y D. GREENBERG. The hemi-cube: a radiosity solution for complex environments. En “Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques”, páginas 31–40 (1985).
- [15] M. COHEN, J. WALLACE Y P. HANRAHAN. “Radiosity and realistic image synthesis”. Academic Press Professional, Inc., San Diego, CA, USA (1993).
- [16] G. COOMBE, M. HARRIS Y A. LASTRA. Radiosity on graphics hardware. En “Proceedings of Graphics Interface 2004”, páginas 161–168. Canadian Human-Computer Communications Society (2004).
- [17] C. DACHSBACHER Y M. STAMMINGER. Reflective shadow maps. En “Proceedings of the 2005 Symposium on Interactive 3D Graphics and Games”, páginas 203–231, New York, NY, USA (2005). ACM.
- [18] G. DAHLQUIST Y Å. BJÖRCK. “Numerical Methods in Scientific Computing: Volume 1”. SIAM, Philadelphia, PA, USA (2008).
- [19] I. DAUBECHIES. “Ten Lectures on Wavelets”. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM (December 1992).
- [20] P. DRINEAS, M. MAHONEY Y S. MUTHUKRISHNAN. Subspace sampling and relative-error matrix approximation: column-row-based methods. En “Proceedings of the 14th conference on Annual European Symposium”, páginas 304–314, London, UK (2006). Springer-Verlag.
- [21] E. ESSELINK. About the order of appel’s algorithm. Computing science note KE5-1, University of Gottingen (1989).
- [22] E. FERNÁNDEZ. Low-rank radiosity. En “Proceedings IV Iberoamerican Symposium in Computer Graphics”, páginas 55–62 (2009).
- [23] E. FERNÁNDEZ, P. EZZATTI Y S. NESMACHNOW. Implementación en GPU del algoritmo de radiosidad de rango bajo. En “Proceedings XVIII Congreso sobre Métodos Numéricos y sus Aplicaciones”, páginas 241–251 (2009).
- [24] J. FOLEY, R. PHILLIPS, J. HUGHES, A. VAN DAM Y S. FEINER. “Introduction to Computer Graphics”. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1994).
- [25] I. FOSTER. “Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering”. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1995).

- [26] A. FRIEZE, R. KANNAN Y S. VEMPALA. Fast Monte-Carlo algorithms for finding low-rank approximations. *Journal of the ACM* **51**(6), 1025–1041 (2004).
- [27] P. GAUTRON, J. KŘIVÁNEK, K. BOUATOUCH Y S. PATTANAİK. Radiance cache splatting: a gpu-friendly global illumination algorithm. En “ACM SIGGRAPH 2008 classes”, páginas 1–10, New York, NY, USA (2008). ACM.
- [28] A. GLASSNER. “Principles of Digital Image Synthesis”. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1994).
- [29] G. GOLUB Y C. VAN LOAN. “Matrix Computations”. The Johns Hopkins University Press (1996).
- [30] C. GORAL, K. TORRANCE, D. GREENBERG Y B. BATTAILE. Modeling the interaction of light between diffuse surfaces. *ACM SIGGRAPH Computer Graphics Quarterly* **18**(3), 213–222 (1984).
- [31] S. GOREINOV, I. OSELEDETS, D. SAVOSTYANOV, E. TYRTYSHNIKOV Y N. ZAMARASHKIN. How to find a good submatrix. Informe técnico 08-10, Institute for Computational Mathematics, Hong Kong Baptist University (2008).
- [32] S. GOREINOV Y E. TYRTYSHNIKOV. The maximal-volume concept in approximation by low-rank matrices. *Contemporary Mathematics* **280**(1), 47–51 (2001).
- [33] S. GOREINOV, E. TYRTYSHNIKOV Y N. ZAMARASHKIN. A theory of pseudoskeleton approximations. *Linear Algebra and its Applications* **261**, 1–21 (1997).
- [34] S. GORTLER, P. SCHRÖDER, M. COHEN Y P. HANRAHAN. Wavelet radiosity. En “Proceedings of the 20th annual conference on Computer graphics and interactive techniques”, páginas 221–230. ACM (1993).
- [35] H. GOURAUD. Continuous shading of curved surfaces. *IEEE Transactions on Computers* **20**(6), 623–629 (1971).
- [36] L. GREENGARD. “The Rapid Evaluation of Potential Fields in Particle Systems”. MIT Press (1988).
- [37] M. GRÖLLER. “Coherence in Computer Graphics”. Tesis Doctoral, Institute of Computer Graphics and Algorithms, Vienna University of Technology (1992).
- [38] P. HANRAHAN, D. SALZMAN Y L. AUPPERLE. A rapid hierarchical radiosity algorithm. En “Proceedings of the 18th annual conference on Computer graphics and interactive techniques”, páginas 197–206. ACM (1991).
- [39] M. HAŠAN. “Matrix Sampling For Global Illumination”. Tesis Doctoral, Cornell University (2009).
- [40] M. HAŠAN, F. PELLACINI Y K. BALA. Matrix row-column sampling for the many-light problem. *ACM Transactions on Graphics* **26**(3), 26 (2007).
- [41] W. HEIDRICH Y H. SEIDEL. View-independent environment maps. En “Proceedings of the ACM SIGGRAPH/EUROGRAPHICS workshop on Graphics hardware”, páginas 39–45, New York, NY, USA (1998). ACM.

- [42] J. HOWELL. “A Catalog of Radiation Heat Transfer Configuration Factors”. University of Texas, Austin, Texas, USA (2001).
- [43] F. IGUAL, G. QUINTANA-ORTÍ Y R. VAN DE GEIJN. Level-3 BLAS on a GPU: Picking the low hanging fruit. En “Minisymposium on High-Performance Computing and Numerical Linear Algebra. 7th International Conference of Numerical Analysis and Applied Mathematics (DVD-ROM)”, Rethymno, Crete, Greece (2009).
- [44] H. JENSEN. “Realistic image synthesis using photon mapping”. A. K. Peters, Ltd., Natick, MA, USA (2001).
- [45] H. JENSEN, J. ARVO, P. DUTRÉ, A. KELLER, A. OVEN, M. PHARR Y P. SHIRLEY. Course 44, Monte Carlo Ray Tracing. En “ACM SIGGRAPH 2003 Full Conference (DVD-ROM)” (2003).
- [46] J. KAJIYA. The rendering equation. En “Proceedings of the 13th annual conference on Computer graphics and interactive techniques”, páginas 143–150, New York, NY, USA (1986). ACM.
- [47] M. KALOS Y P. WHITLOCK. “Monte Carlo methods. Vol. 1: basics”. Wiley-Interscience, New York, NY, USA (1986).
- [48] A. KELLER. Instant radiosity. En “Proceedings of the 24th annual conference on Computer graphics and interactive techniques”, páginas 49–56, New York, NY, USA (1997). ACM Press/Addison-Wesley Publishing Co.
- [49] D. KIRK Y J. ARVO. Unbiased variance reduction for global illumination. En “Proceedings of the Second Eurographics Workshop on Rendering”, páginas 153–156, Barcelona, Spain (May 1991).
- [50] J. KŘIVÁNEK. “Radiance Caching for Global Illumination Computation on Glossy Surfaces”. Ph.D. thesis, Université de Rennes 1 and Czech Technical University in Prague (2005).
- [51] C. L. LAWSON, R. J. HANSON, D. R. KINCAID Y F. T. KROGH. Basic Linear Algebra Subprograms for Fortran Usage. *ACM Transactions on Mathematical Software* **5**(3), 308–323 (1979).
- [52] J. LIENHARD IV Y J. LIENHARD V. “The heat transfer textbook”. Phlogiston Press, Cambridge, Massachusetts, USA (2004).
- [53] M. MAHONEY Y P. DRINEAS. Cur matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences of the United States of America* **106**(3), 697–702 (2009).
- [54] T. MALLEY. A shading method for computer generated images. Proyecto Fin de Carrera, Department of Computer Science, University of Utah (1988).
- [55] N. METROPOLIS Y S. ULAM. The Monte Carlo Method. *Journal of the American Statistical Association* **44**(247), 335–341 (1949).
- [56] J. MORICE, K. MER-NKONGA Y A. BACHELOT. Fast multipole method for solving the radiosity equation. En “Proceedings of the 6th European conference on numerical mathematics and advanced applications”, páginas 609–617. Springer (2006).

- [57] M. NIJASURE, S. PATTANAIK Y V. GOEL. Realtime global illumination on GPU. *Journal of Graphics Tools* **10**(2), 55–71 (2005).
- [58] NVIDIA. “CUDA CUBLAS Library”. NVIDIA Corporation, Santa Clara, California (2008).
- [59] I. OSELEDETS, D. SAVOSTIANOV Y E. TYRTYSHNIKOV. Tucker dimensionality reduction of three-dimensional arrays in linear time. *SIAM Journal on Matrix Analysis and Applications* **30**(3), 939–956 (2008).
- [60] W. PRESS, S. TEUKOLSKY, W. VETTERLING Y B. FLANNERY. “Numerical Recipes: The Art of Scientific Computing”. Cambridge University Press, New York, NY, USA (2007).
- [61] T. RITSCHER, T. GROSCH, M. H. KIM, H.-P. SEIDEL, C. DACHSBACHER Y J. KAUTZ. Imperfect shadow maps for efficient computation of indirect illumination. *ACM Transactions on Graphics* **27**(5), 1–8 (2008).
- [62] Y. SAAD. “Iterative Methods for Sparse Linear Systems”. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2003).
- [63] M. SBERT. “The Use of Global Random Directions to Compute Radiosity: Global Monte Carlo Techniques”. Tesis Doctoral, Universitat Politècnica de Catalunya, Barcelona, Spain (1996).
- [64] R. SIEGEL Y J. HOWELL. “Thermal radiation heat transfer”. Series in thermal and fluids engineering. Hemisphere Pub. Corp, New York, USA (1981).
- [65] P. SLOAN, J. KAUTZ Y J. SNYDER. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. En “Proceedings of the 29th annual conference on Computer graphics and interactive techniques”, páginas 527–536, New York, NY, USA (2002). ACM.
- [66] G. STEWART. “Matrix Algorithms – Volume I: Basic Decompositions”. Society for Industrial and Applied Mathematics (1998).
- [67] I. SUTHERLAND, R. SPROULL Y R. SCHUMACKER. A characterization of ten hidden-surface algorithms. *ACM Computing Surveys* **6**, 1–55 (1974).
- [68] L. SZIRMAY-KALOS, L. SZÉCSI Y M. SBERT. “GPU-Based Techniques for Global Illumination Effects”. Synthesis Lectures on Computer Graphics and Animation. Morgan & Claypool Publishers (2008).
- [69] R. WANG, R. WANG, K. ZHOU, M. PAN Y H. BAO. An efficient GPU-based approach for interactive global illumination. En “ACM SIGGRAPH 2009 papers”, páginas 1–8, New York, NY, USA (2009). ACM.
- [70] G. WARD, F. RUBINSTEIN Y R. CLEAR. A ray tracing solution for diffuse inter-reflection. En “Proceedings of the 15th annual conference on Computer graphics and interactive techniques”, páginas 85–92, New York, NY, USA (1988). ACM.
- [71] C. WARE. “Information Visualization: Perception for Design”. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2004).

-
- [72] R. WHALEY, A. PETITET Y J. DONGARRA. Automated empirical optimizations of software and the atlas project. *Parallel Computing* **27**(1-2), 3 – 35 (2001).
- [73] T. WHITTED. An improved illumination model for shaded display. *Communications of the ACM* **23**(6), 343–349 (1980).
- [74] O. ZIENKIEWICZ Y R. TAYLOR. “The finite element method”. Butterworth-Heinemann, Oxford (2000).