Universidad Zaragoza
1542

Trabajo Fin de Grado

Estimación de Escala Absoluta para Structure from Motion en Endoscopio con Fuente de Luz Cercana

Absolute Scale Estimation for Structure from Motion in Endoscope with Nearby Light Source

Autor

Anyiel Fernandes Araujo

Director

José María Martínez Montiel

ESCUELA DE INGENIERÍA Y ARQUITECTURA

2022

AGRADECIMIENTOS

Me gustaría agradecer al profesor José María Martínez Montiel, por todos sus consejos, guías y confianza que tanto he valorado a lo largo de los últimos meses. Al profesor Juan Domingo Tardós por ofertarme el TFG en colaboración con el proyecto de investigación EndoMapper y ayudarme a pulir los últimos detalles del trabajo.

Doy mi más sincero agradecimiento a mi padre y a mi madre, por la educación que me han dado, por su consejo y apoyo incondicional a lo largo de toda mi vida, por el inmenso esfuerzo y sacrificio que han realizado para que sus hijos salgan adelante, y por la confianza y ánimos que me han llevado hasta este punto de mi vida.

Agradezco de corazón a mi hermano, por ser el pilar de mi desarrollo personal y mi compañero de vida.

Agradezco a toda mi familia, por haberme apoyado y aconsejado en todas mis decisiones.

También agradezco a mis compañeros de grado, que han amenizado esta etapa de mi vida.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 863146

Estimación de Escala Absoluta para Structure from Motion en Endoscopio con Fuente de Luz Cercana

RESUMEN

Actualmente es posible reconstruir la geometría de un escenario observado por una cámara monocular a partir de varias imágenes tomadas en distintas posiciones. No obstante, la escala de la escena no es observable y es necesario emplear información adicional a la geometría para recuperar el tamaño real de la escena.

Se considera el caso de una cámara monocular móvil que incorpora varias fuentes de luz, cuya posición relativa a la cámara es fija, conocida y denominada *base* (b). En esta situación, es posible recuperar la escala y el albedo de la escena a partir de una reconstrucción geométrica no escalada. Además, el escenario propuesto se asemeja a un entorno de endoscopia, en el que varias fuentes de luz iluminan una escena cuya iluminación es completamente artificial.

Este trabajo se desarrolla dentro del proyecto europeo EndoMapper, que busca realizar una reconstrucción 3D del interior del cuerpo humano a partir de secuencias de endoscopia, donde el tamaño real de la reconstrucción es ciertamente importante. Se expone en este documento una metodología de estimación de escala y albedo a partir de una reconstrucción geométrica hasta un factor de escala. El método se basa en un modelo fotométrico simple con cámara fotométricamente calibrada y fuentes de luz puntuales uniformes, en el que la escala se estima mediante una optimización no lineal por mínimos cuadrados. La precisión de este método, así como su comportamiento, se analizan mediante una simulación que reproduce un escenario de colonoscopia real.

Se muestra en simulación que el método requiere de 2 imágenes en escala de grises con 100 puntos emparejados y una estimación de la potencia de las fuentes de luz entre 0.01 y 100 veces su valor real. Si la escena está bien condicionada, esto es, su profundidad es menor que $3b$ y la traslación entre las imágenes es similar a la profundidad, es posible estimar la escala absoluta con un error entre el 2 % y el 4 %. Por otro lado, el albedo se puede estimar hasta un factor de escala, con menos de un 4 % de error siempre que traslación y profundidad sean mayores que $2b$. Todo ello incluso en el caso de

desconocer los valores de exposición de la cámara debido a la acción del AGC (control automático de ganancia). En caso de conocerse dichos valores, se evidencia que el error puede reducirse al 1 % tanto en escala como en albedo.

El código de este trabajo puede encontrarse en el repositoriro: `https://github.com/UZ-SLAMLab/absolute-geometrical-scale-from-near-light-monocular-images.git`.

Índice general

Capítulo 1

Introducción

Structure From Motion (SfM) [1] es una técnica de visión por computador capaz de reconstruir la geometría de una escena a partir de una secuencia de imágenes. Este método realiza un seguimiento de una serie de puntos a lo largo de la escena y, utilizando un modelo de cámara, es capaz de estimar la posición de los puntos observados en las imágenes, así como sus normales e incluso el trayecto de la cámara por la escena. No obstante, con esta metodología el tamaño real de la escena no es observable. Se dice entonces que la escena se reconstruye *hasta un factor de escala*.

Sin embargo, bajo ciertas circunstancias, la información fotométrica de la escena puede permitir recuperar su escala, tal y como demostró Iwahori en [2]. Nuestro trabajo considera por primera vez el escenario de una cámara monocular móvil que incorpora varias fuentes de luz en posiciones fijas y conocidas (iluminación co-localizada). Se propone una metodología para recuperar la escala absoluta y el albedo del escenario haciendo uso exclusivo de las imágenes captadas por la cámara al recorrer la escena.

Este trabajo se desarrolla dentro del proyecto europeo EndoMapper, que busca reconstruir mapas 3D del interior del cuerpo humano a partir de secuencias de endoscopia médica. Resulta que el caso considerado se adecúa perfectamente a un escenario de endoscopia, pues un endoscopio es una cámara monocular que incorpora 3 fuentes de luz cuyas posiciones relativas con respecto a la cámara se pueden conocer por las propias especificaciones del endoscopio, o mediante un proceso de calibración. Además, en este área, la escala de la escena es de vital importancia para diagnosticar distintas afecciones en el organismo. Así pues, en este trabajo se desarrolla la aplicación del problema en el entorno de endoscopia, en concreto, un escenario de colonoscopia.

En el Capítulo 2 se introduce un modelo fotométrico basado en [3] que explica, de manera simplificada, la formación de imágenes en cámaras monoculares con iluminación co-localizada. A continuación, en el Capítulo 3 se realiza un análisis teórico del problema y se propone un método de estimación de escala basado en optimización no lineal. Posteriormente, en el Capítulo 4 se presenta un modelo de simulación aplicado al entorno de colonoscopia, que se utiliza finalmente en el Capítulo 5 para analizar la precisión de la solución y caracterizar su comportamiento.

Todo el código de este trabajo puede encontrarse en el siguiente repositorio, que se hará público en cuanto reciba la aprobación del proyecto: `https://github.com/UZ-SLAMLab/absolute-geometrical-scale-from-near-light-monocular-images.git`.

1.1. Revisión bibliográfica

Horn, en su tesis doctoral [4], es el primero en aprovechar la fotometría para reconstruir la geometría de una escena asumiendo una escena lambertiana de albedo uniforme con fuentes de luz en el infinito (*shape from shading*). No obstante, la escala real del escenario no era observable. Posteriormente, Woodham en [5] propone un *estéreo fotométrico* que utiliza múltiples fuentes de luz que se van apagando y encendiendo selectivamente para realizar la reconstrucción de los puntos, las normales y los albedos de la escena, pero sin poder determinar su escala. Iwahori en [2] propone el primer estéreo fotométrico capaz de recuperar la escala de la escena, con una cámara y fuentes de luz a distancia finita. Clark en [6] también es capaz de recuperar la escala de la escena, pero suponiendo cámara fija y una fuente de luz móvil cuyas posiciones eran conocidas.

En este trabajo se parte de una reconstrucción de la escena realizada con SfM clásico (geometría multivista [1] y estéreo multivista [7]), que produce la geometría de un conjunto de puntos de la escena, así como sus normales y las poses de las cámaras que captaron las imágenes, todo ello hasta un factor de escala. En contraste con otros estudios, nosotros proponemos el caso de una cámara monocular móvil con trayectoria desconocida y unas fuentes de luz rígidamente unidas a la cámara, en el que recuperamos la geometría de la escena en escala absoluta y los albedos de sus puntos. Otros trabajos han tratado problemas parecidos, pero nunca de la manera aquí expuesta. Collins y Bartoli en [8] proponen una idea similar, pero con fuentes de luz de distin-

tos colores, y Wu et al. en [3] abordan el problema en un entorno de endoscopia pero utilizando información de un tracker externo a la cámara. Un trabajo similar se ha desarrollado también en el ámbito de EndoMapper [9], pero requiere solamente de una imagen monocular y no estima albedos.

Capítulo 2

Modelo fotométrico de una cámara con iluminación co-localizada

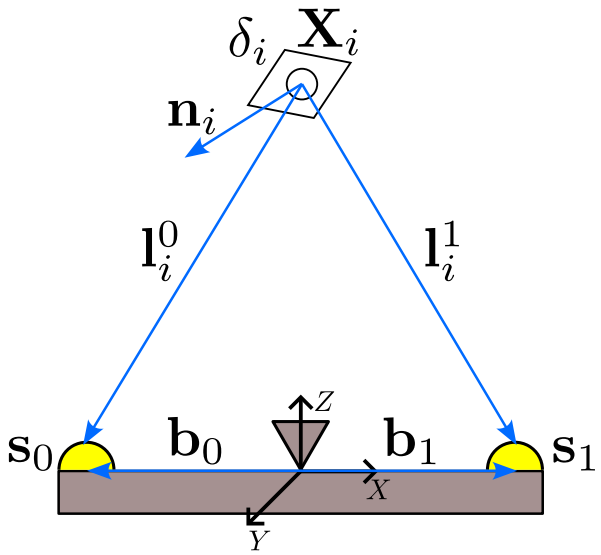


Figura 2.1: Geometría de la formación de imágenes en un entorno con iluminación co-localizada

Atendiendo a la Figura 2.1, sea $\mathbf{X}_i = (x, y, z)^\top$ las coordenadas de un punto sobre una superficie lambertiana, con normal $\mathbf{n}_i$ y albedo δ_i. Sea $\mathbf{s}_j$ la posición de una fuente de luz puntual, y sea I_0 su potencia de iluminación (idéntica en todas las fuentes). Si se ignoran los efectos de iluminación global y se supone $\mathbf{s}_j$ como única fuente de luz,

la radiancia R_i^j que genera esta fuente sobre el punto es [3]:

$$\mathbf{l}_i^j = \mathbf{s}_j - \mathbf{X}_i \tag{2.1}$$

$$R_i^j = I_0 \delta_i \frac{\cos(\mathbf{n}_i, \mathbf{l}_i^j)}{|\mathbf{l}_i^j|^2} = I_0 \delta_i \frac{\mathbf{n}_i \cdot \mathbf{l}_i^j}{|\mathbf{l}_i^j|^3} \tag{2.2}$$

La cámara integra varias fuentes de luz cuya posición se puede conocer si se sabe la ubicación y pose de la cámara. Además, todas las fuentes de luz están a la misma distancia euclídea del centro óptico de la cámara. Esta distancia se denomina *base* o b, y es la única medida real que se conoce de la escena.

Sea $\mathbf{b}_j = (x, y, z)^\top$ las coordenadas de una fuente de luz con respecto a la posición de la cámara ($|\mathbf{b}_j| = b$). Sea $\mathbf{R}$ una matriz 3×3 correspondiente a la rotación de la cámara y sea $\mathbf{t} = (x, y, z)^\top$ la traslación de esta, la posición de la fuente de luz $\mathbf{s}_j$ es:

$$\tilde{\mathbf{s}}_j = \begin{bmatrix} \mathbf{R}^\top & -\mathbf{R}^\top \mathbf{t} \\ 0 & 1 \end{bmatrix} \tilde{\mathbf{b}}_j \tag{2.3}$$

donde $\tilde{\mathbf{s}}_j$ y $\tilde{\mathbf{b}}_j$ son las coordenadas homogéneas de $\mathbf{s}_j$ y $\mathbf{b}_j$. La radiancia total R_i que mediría la cámara al observar un punto es la suma de las radiancias individuales de cada fuente de luz. Suponiendo imágenes en blanco y negro, la radiancia percibida se transformará en un valor de nivel de gris L_i gracias al control de ganancia automático de la cámara (AGC), que se modela como una transformación afín con dos parámetros de regulación α y β:

$$L_i = \alpha \sum_j R_i^j + \beta \tag{2.4}$$

En una escena de escala geométrica desconocida, se puede expresar la posición de la fuente de luz y del punto observado en función de un parámetro λ que representa la escala (real o no) de la escena (y en consecuencia también R_i^j y L_i):

$$\tilde{\mathbf{s}}_j(\lambda) = \begin{bmatrix} \mathbf{R}^\top & -\lambda \mathbf{R}^\top \mathbf{t} \\ 0 & 1 \end{bmatrix} \tilde{\mathbf{b}}_\mathbf{j} \tag{2.5}$$

$$\mathbf{l}_i^j(\lambda) = \mathbf{s}_j(\lambda) - \lambda \mathbf{X}_i \tag{2.6}$$

Finalmente, supóngase que la cámara recorre la escena siguiendo una trayectoria conocida. Se disponen de m imágenes de varias cámaras[1] cuyas traslaciones y rotaciones se conocen en relación con las de la primera cámara (referencia) hasta un factor de escala. Cada cámara k tienen un AGC propio e independiente (α_k y β_k) que se puede expresar también en función de la iluminación percibida en la primera cámara para cada punto (α'_k y β'_k), de modo que toda la fotometría es relativa a la cámara de referencia:

$$0 \leq k < m \tag{2.7}$$

$$\alpha'_0 = \alpha_0 I_0 \tag{2.8}$$

$$\alpha'_k = \frac{\alpha_k}{\alpha_0}; \quad k > 0 \tag{2.9}$$

$$L_i^k(\lambda) = \begin{cases} \alpha'_0 \delta_i \left(\sum_j R_i^{jk}(\lambda) \right) + \beta'_0 & \text{si } k = 0 \\ \\ \alpha'_k \alpha'_0 \delta_i \left(\sum_j R_i^{jk}(\lambda) \right) + \beta'_k & \text{si } k > 0 \end{cases} \tag{2.10}$$

Con este modelo se supone que la iluminación de la escena proviene de varias fuentes de luz puntuales uniformes y que la cámara nos proporciona información de la radiancia. Para llevar esto a la práctica es necesario un calibrado fotométrico de la cámara y de las fuentes de luz. Por otro lado, es importante remarcar que la escena se supone lambertiana y se ignoran los efectos de iluminación global.

[1] No se hace referencia a la cámara física, sino a su pose (rotación y traslación). Por ejemplo, al desplazar una cámara un número de unidades a la derecha, se considerarán dos cámaras distintas, aunque no coexistan en escena y físicamente sean la misma.

Capítulo 3

Estimación de escala absoluta

La escala métrica real (λ_m) de una escena no es observable si solo se considera la información geométrica de una secuencia de imágenes monoculares. Podría observarse una escena gigantesca en la que la cámara presenta una traslación enorme, o por el contrario podría tratase de una escena diminuta observada por una cámara con traslación también minúscula.

No obstante, la escala métrica es observable cuando se valora la información fotométrica de la escena en entornos controlados, tal y como se ha abordado tradicionalmente en problemas de estéreo fotométrico [4]. Se considera ahora un problema similar: una cámara monocular móvil con iluminación co-localizada y *base* conocida. En este caso es posible estimar la escala de la escena siempre que la cámara esté lo suficientemente cerca de la misma y presente una traslación similar a la base. Ocurre que una endoscopia se ajusta a este escenario: se conoce la base, la cámara está muy cerca de la escena y su iluminación es completamente artificial. Además, existe una ventaja adicional, y es que se puede estimar a priori qué tamaño aproximado tendría la escena a partir del conocimiento médico.

En este capítulo se realiza, en primer lugar, un análisis teórico de la viabilidad de la recuperación de escala suponiendo una versión simplificada del problema. Posteriormente se hace una presentación formal del caso de estudio y, finalmente, se propone una solución basada en optimización no lineal.

3.1. Modelo simplificado con un punto y una fuente de luz

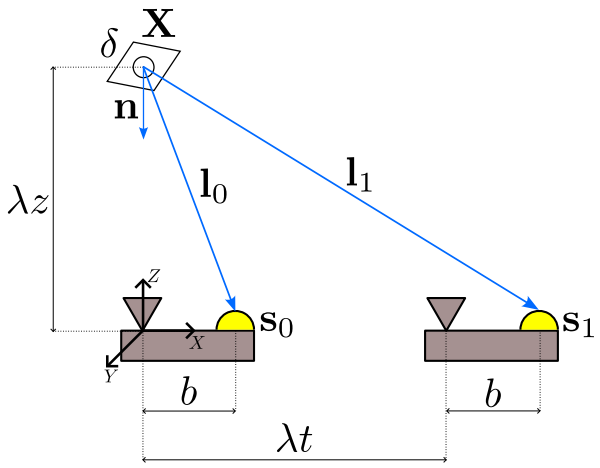


Figura 3.1: Diagrama de una simplificación del problema de recuperación de escala en una cámara monocular con iluminación co-localizada.

En la Figura 3.1 se muestra una simplificación del problema de recuperación de escala. Una cámara observa el punto $\mathbf{X}$ de albedo δ, cuya profundidad se conoce hasta un factor de escala (λz) y cuya normal $\mathbf{n}$ apunta al centro óptico de la cámara. La cámara integra una única fuente de luz que se encuentra b unidades a la derecha de la misma (base). A continuación, la cámara se desplaza λt unidades a la derecha (traslación conocida hasta un factor de escala) y vuelve a observar el punto.

Por la Ecuación 2.2, la radiancia que capturan los sensores de cada cámara $R_0(\lambda)$ y $R_1(\lambda)$ son:

$$R_0(\lambda) = I_0\delta\frac{\mathbf{n}\cdot\mathbf{l}_0}{|\mathbf{l}_0|^3} = I_0\delta\frac{\lambda z}{\left(b^2+\lambda^2z^2\right)^{3/2}} \tag{3.1}$$

$$R_1(\lambda) = I_0\delta\frac{\mathbf{n}\cdot\mathbf{l}_1}{|\mathbf{l}_1|^3} = I_0\delta\frac{\lambda z}{\left((\lambda t+b)^2+\lambda^2z^2\right)^{3/2}} \tag{3.2}$$

El albedo observado para el punto es invariable (superficie lambertiana), y por tanto es posible despejar δ de estas ecuaciones. Tras simplificar $I_0\lambda z$ se tiene que:

$$R_0(\lambda)\left(b^2+\lambda^2z^2\right)^{3/2} = R_1(\lambda)\left((\lambda t+b)^2+\lambda^2z^2\right)^{3/2} \tag{3.3}$$

Sea λ_m la escala real de la escena, para la que se conocen $R_0^m = R_0(\lambda_m)$ y $R_1^m = R_1(\lambda_m)$ (valores captados por los sensores de las cámaras). Se puede simplificar la Ecuación 3.3

elevando ambos términos a 2/3:

$$b^2 + \lambda_m^2 z^2 = \left(\frac{R_1^m}{R_0^m}\right)^{2/3} \left((\lambda_m t + b)^2 + \lambda_m^2 z^2\right) = k\left((\lambda_m t + b)^2 + \lambda_m^2 z^2\right) \tag{3.4}$$

Donde $k = (R_1^m/R_0^m)^{2/3}$ es una constante conocida. Ahora bien, la Ecuación 3.4 no restringe λ si $b = 0$, porque cualquier λ_m cumple la ecuación:

$$\lambda_m^2 z^2 = k\lambda_m^2 t^2 + k\lambda_m^2 z^2 \tag{3.5}$$

Es decir, si la fuente de luz está en el centro óptico de la cámara ya no hay una "referencia" que permita recuperar la escala real. No obstante, si $b \neq 0$, tras simplificar la Ecuación 3.4, se obtiene una ecuación cuadrática de solución:

$$\lambda_m = b\,\frac{-kt \pm \sqrt{kt^2 - z^2(k-1)^2}}{kt^2 + z^2(k-1)} \tag{3.6}$$

Por tanto, existen dos soluciones a la ecuación. Sin embargo, una de ellas ha sido introducida al elevar la ecuación a 2/3, y no justificará los valores de $R_0(\lambda_m)$ y $R_1(\lambda_m)$ observados, por lo que es trivial seleccionar la solución correcta.

Se concluye así que en una escena observada por una cámara monocular móvil e iluminación co-localizada, en ciertos casos es posible determinar la escala del escenario si se suponen superficies lambertianas y siempre que la base sea mayor que 0. Todo ello gracias a la invarianza del albedo observado.

3.2. Formalización del problema

Se supone una cámara que tiene asociadas l fuentes de luz, todas ellas a una distancia b del centro óptico de la cámara, y de las que se conocen sus coordenadas locales reales $\mathbf{b}_j$ $(0 \leq j < l)$ en relación a dicho centro. Se disponen de m imágenes en escala de grises tomadas por la cámara en distintas posiciones, es decir, $L_i^k(\lambda_m)$. Además, se conoce la geometría de n puntos de la escena y las poses de las cámaras hasta un factor de escala, esto es: $\mathbf{R}_k, \mathbf{t}_k, \mathbf{X}_i$ y $\mathbf{n}_i$ $(0 \leq k < m,\ 0 \leq i < n)$.

El AGC es un parámetro interno de la cámara que suele ser desconocido, pero no existe una limitación que impida exponer sus valores si se desea. Por tanto, es posible que en algún caso el AGC de la cámara se pueda conocer con exactitud y se simplificaría

el problema. A continuación se proponen dos variantes del problema: con AGC conocido y con AGC desconocido. Este último es el caso más realista a día de hoy y en el que se centra este trabajo.

En la versión simplificada, se supone que se conoce a la perfección el AGC de las cámaras, esto es: α'_k y β'_k. En consecuencia, se conoce también I_0. Por tanto, quedan como incógnitas: λ_m y δ_i. Cada punto proporciona m ecuaciones, y se tienen $n+1$ incógnitas de modo que, si el número de imágenes es fijo, se necesitan $n \geq 1/(m-1)$ puntos para tener un sistema determinado, por lo que se requieren al menos dos imágenes.

Por el contrario, si no se conoce el AGC de las cámaras, se busca estimar $\lambda_m, \delta_i, \alpha'_k$ y β'_k. En este caso, además, se tiene una estimación de I_0 (cuya justificación es abordada en la Sección 5.5). De nuevo, cada punto proporciona m ecuaciones, pero hay $2m+n+1$ incógnitas. Si el número de imágenes es fijo, se requieren $n \geq (2m+1)/(m-1)$ puntos, es decir, mínimo dos imágenes y se necesitan más puntos que en el caso de AGC conocido.

Si el AGC es desconocido se puede observar que no es posible estimar correctamente los albedos porque van siempre en producto con α'_0 y, en consecuencia, lo que se estima realmente es $\alpha'_0\delta_i$ y no sus valores por separado. En otras palabras, se estima δ'_i, un albedo cuyo valor puede ser superior a 1 y que representa la iluminación del punto en la primera cámara, por lo que α'_0 puede eliminarse de la Ecuación 2.10 y suponer:

$$\delta'_i = \alpha_0 \delta_i I_0 \tag{3.7}$$

$$L_i^k(\lambda) = \begin{cases} \delta'_i \left(\sum_j R_i^{jk}(\lambda) \right) + \beta'_0 & \text{si } k = 0 \\ \alpha'_k \delta'_i \left(\sum_j R_i^{jk}(\lambda) \right) + \beta'_k & \text{si } k > 0 \end{cases} \tag{3.8}$$

No obstante, las relaciones entre los albedos se mantienen. Este albedo "escalado" (δ'_i) supone un problema para analizar la precisión del modelo, pues interesa estudiar cuán bien se captan las relaciones entre los albedos sin evaluar su "escala". Por ello, antes de realizar cualquier comparación, se realizará un escalado de los albedos estimados, buscando k tal que:

$$\arg\min_k \sum_i \left(\delta_i - k\hat{\delta}_i \right)^2 \tag{3.9}$$

Y después se comparará $k\hat{\delta}'_i$ con δ_i de la manera pertinente. El valor de k se obtiene con una optimización no lineal por el algoritmo Levenberg–Marquardt [10], y se inicializa k a $1/(\text{máx}\,\delta_i)$ (lo que mantiene $k\hat{\delta}_i$ entre 0 y 1).

3.3. Estimación de escala absoluta mediante optimización no lineal

Se busca recuperar la escala métrica λ_m utilizando optimización no lineal por mínimos cuadrados. Se dispone de n puntos de los cuales se desconoce el albedo δ_i y m imágenes de cuyas cámaras se desconoce el control de ganancia (α'_k, β'_k). El objetivo es encontrar los valores para $\lambda_m, \delta_i, \alpha'_k$ y β'_k que generen niveles de gris lo más parecidos a los observados para cada punto. En este caso:

$$\underset{\lambda,\alpha'_0,\ldots,\alpha'_{m-1},\beta'_0,\ldots,\beta'_{m-1},\delta_0,\ldots,\delta_{n-1}}{\arg\min} \sum_{i,k} \left(\hat{L}_i^k(\lambda, \alpha'_0, ..., \alpha'_{m-1}, \beta'_0, ..., \beta'_{m-1}, \delta_0, ..., \delta_{n-1}) - L_i^k\right)^2 \tag{3.10}$$

Donde $\hat{L}_i^k$ es el nivel de gris estimado en la cámara k para el punto i, cuyo valor se computa de acuerdo a la Ecuación 2.10.

Se trata de un problema de mínimos cuadrados sin restricciones que se podría resolver por el algoritmo de Levenberg–Marquardt [10]. Sin embargo, se ha utilizado el algoritmo TRF (*trust-region reflective)* [11] porque la implementación disponible del Levenberg–Marquardt no permitía un jacobiano disperso, mientras que la de TRF sí lo hacía (lo que agiliza mucho la optimización) [12]. El jacobiano analítico de la función objetivo (3.10) se adjunta en el Apéndice A.

3.4. Inicialización de la optimización no lineal

La inicialización de los valores de la optimización es crucial para evitar que el optimizador se estanque en un mínimo local. Para ello, interesa que sus valores iniciales estén lo más cercanos a los valores óptimos.

Tras varios experimentos, se ha concluido que la mejor manera de inicializar la optimización es la siguiente:

1. Inicialización de albedos (δ_i): se realiza una estimación de cuál es el albedo medio

de la escena y se inicializan todos los albedos a dicho valor: $\forall i,\ \delta_i = \bar{\delta}$. Si no se pudiese estimar, se supone que el albedo medio es 0.5.

2. Inicialización de control de ganancia:

   - En la inicialización de α'_k, se supone que $\alpha_k = 1$ y por consiguiente:

     - $\alpha'_0 \simeq I_0$: es necesario tener una estimación inicial de I_0, lo que es posible y se analiza en la Sección 5.5.

     - $\alpha_k \simeq 1/I_0,\ k > 0$: por la Ecuación 2.9, esto conserva la proporción entre α'_0 y α'_k.

   - $\forall k,\ \beta'_k = 0$

3. Inicialización de λ: se hace una búsqueda exhaustiva en el dominio de L, buscando el valor de λ que minimice la suma de las diferencias al cuadrado (3.10). La búsqueda exhaustiva se acota en en el dominio $[a_0, a_1]$, donde a_0 y a_1 se determinan manualmente en función del tamaño mínimo y máximo esperado para la escena tratada. Esto es factible en un entrono de endoscopia porque se sabe que ciertos tamaños son demasiado pequeños o demasiado grandes para lo observado. Por ejemplo, en la simulación propuesta en el siguiente capítulo, $a_0 = 0$ y $a_1 = 6$, mientras que $\lambda_m = 0.5$. En este dominio, se realizan 50 evaluaciones de la función objetivo, con distribución lineal, en busca del valor óptimo.

De este modo se dispone de unos valores iniciales más o menos factibles para $\lambda_m, \delta_i, \alpha'_k$ y β'_k, lo que asegurará que el optimizador pueda encontrar la solución correcta si la escena está bien condicionada (analizado en el Capítulo 5).

Capítulo 4

Modelo de simulación

Se desea verificar mediante simulación que el método de estimación de escala propuesto en el Capítulo 3 es útil en un entorno de endoscopia. Para ello, se ha realizado una simulación simplista de un escenario de colonoscopia con el objetivo de refinar la metodología propuesta, así como detectar y caracterizar las variables más relevantes del problema. En este capítulo se describe el modelo de simulación empleado para llevar a cabo los experimentos expuestos en el Capítulo 5.

4.1. Geometría de la escena y características del endoscopio

El objetivo de una colonoscopia es identificar posibles anomalías en el colon que puedan dar origen a afecciones más graves en un futuro. En este contexto, una de las anomalías más comunes es el pólipo: una acumulación de células en la pared del colon que frecuentemente es inofensiva, pero que puede llegar a ocasionar enfermedades más graves como el cáncer de colon. El tamaño de los pólipos es un indicador del riesgo para el desarrollo de cáncer colonorectal. Las guías basan el seguimiento en el tamaño del pólipo a pesar de que la estimación de su tamaño a partir de las imágenes es subjetiva y por ello es muy relevante el tamaño real de la escena.

Así pues, se ha tratado de simular una escena con un pólipo de tamaño interesante. Se ha hecho uso del dataset público de colonoscopias del proyecto EndoMapper [13] para modelar una escena realista (Figura 4.1). En la Figura 4.2 se muestra un diagrama con las medidas y componentes de la escena simulada. Esta es la escena de referencia,

pero para ciertos experimentos se tratará de alejar o acercar los puntos de la cámara para analizar el papel de la profundidad en el problema.

Por otro lado, se ha tratado de simular el endoscopio EVIS EXERA III CF-H190L/I mediante una cámara pinhole con un campo de visión de 120° en horizontal y vertical y una resolución de 512×512 píxeles. Tiene asociadas tres fuentes de luz dispuestas en un triángulo equilátero, todas ellas a una distancia de 3.89 mm de su centro de visión. Esta medida (base) es el factor clave que permite la recuperación de escala.

En la simulación, el albedo de los puntos de la escena sigue una distribución aleatoria uniforme en $[0.3, 0.7]$ (media 0.5).

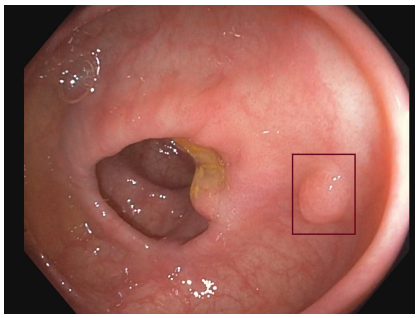


Figura 4.1: Pólipo de 4mm adherido a la pared del colon. Imagen extraída de la secuencia seq_47 minuto 3:38, del dataset de colonoscopias EndoMapper [13].

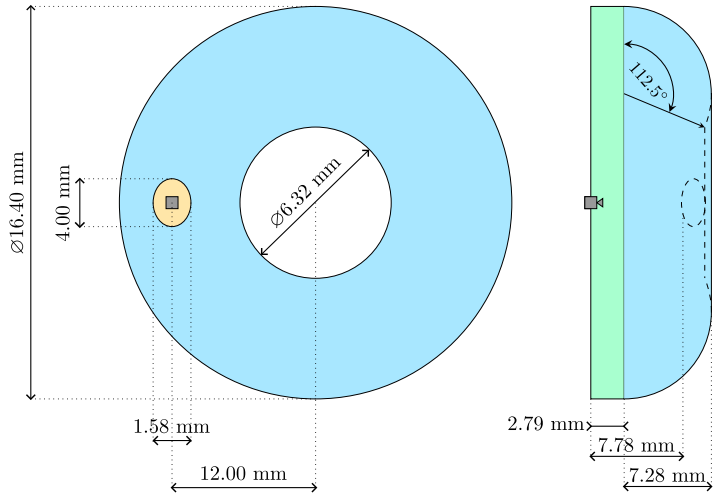


Figura 4.2: Escena simulada. Se compone de un cilindro (verde) que representa las paredes del colon, un elipsoide (naranja) que representa un pólipo, y una sección de toroide (azul) que representa el haustra sobre la que se encuentra el pólipo.

4.2. Formación de imágenes

Dado un conjunto de puntos con geometría conocida, se forman las imágenes siguiendo el modelo fotométrico expuesto en el Capítulo 2 y el modelo de cámara pinhole [14]. Con esto es posible conocer el píxel que corresponde a cada punto, así como la radiancia que percibe la cámara. No obstante, dicha radiancia debe convertirse a niveles de gris simulando el AGC. Ello se hace con una transformación afín sencilla que genera niveles de gris en el rango $[L_{min}, L_{max}]$ a partir de las radiancias R_i^k de cada punto en la cámara k:

$$\alpha_k = \frac{1}{\text{máx}\, R_i^k} \cdot (L_{max} - L_{min}) \tag{4.1}$$

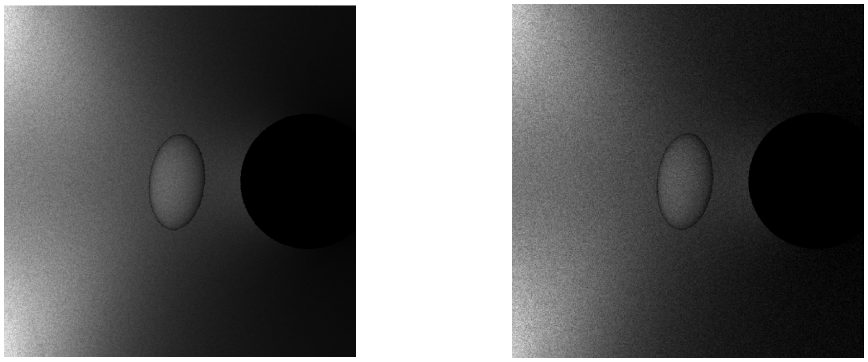
$$\beta_k = L_{min} \tag{4.2}$$

Por otro lado, en un caso real, la radiancia percibida por la cámara es perturbada por el error de su sensor, que se supondrá gaussiano:

$$L_i^k = \alpha_k R_i^k + \beta_j + \mathcal{N}(\mu, \sigma) \tag{4.3}$$

Al añadir el ruido, se busca respetar el rango de valores $[L_{min}, L_{max}]$. En este trabajo, $\mu = 0$, $\sigma = 2.5$, $L_{min} = 12$ y $L_{max} = 255$. En la Figura 4.3 se muestra una comparativa de la imagen obtenida sin ruido y con ruido.

El algoritmo de optimización se basa en rasterización, es decir, las imágenes solamente muestran el conjunto de puntos con el que se trabaja y no la escena completa, porque no es necesaria. Esto es eficiente en computación pero no permite ver qué es lo que realmente observa la cámara, lo que sin duda es necesario para analizar el comportamiento del algoritmo. Para ello se ha implementado un algoritmo de ray tracing [15] con el que se generan las imágenes que ve realmente la cámara simulada.



(a) Sin ruido gaussiano *(b) Ruido gaussiano: $\mu = 0$, $\sigma = 2.5$*

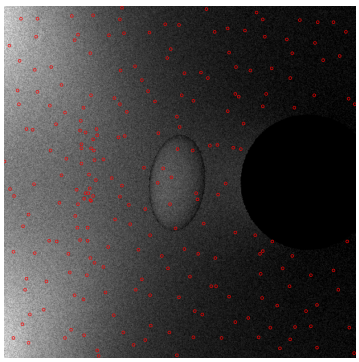
Figura 4.3: Imágenes generadas mediante ray tracing desde la posición inicial del endoscopio frente al pólipo.

4.3. Simulación de reconstrucción geométrica

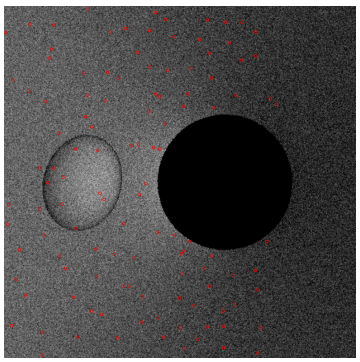
El problema parte de la suposición de que se dispone de una reconstrucción de la geometría de la escena hasta un factor de escala, esto es: un conjunto de puntos y sus normales, y las traslaciones y rotaciones de las cámaras. La pose de la cámara es algo trivial de simular y variará de un experimento a otro, pero los puntos que se reconstruyen de la escena deben tratarse con más cuidado. Esta parte de la reconstrucción geométrica se ha simulado muestreando un conjunto de puntos de la escena observada por la cámara de referencia, utilizando para ello el algoritmo de ray tracing ya mencionado (Figura 4.4).

Se adelanta que los experimentos realizados en el Capítulo 5 simularán movimientos de cámara en horizontal y hacia la derecha, sin ningún tipo de rotación, y además harán uso de solamente dos imágenes. Ello requiere, por tanto, que los puntos utilizados sean visibles desde las dos cámaras (co-visibilidad). Por otro lado, para poder realizar las comparativas necesarias es imprescindible que el conjunto de puntos no varíe de un experimento a otro y, en consecuencia, se busca utilizar un conjunto de puntos que no presente problemas de co-visibilidad para un amplio número de traslaciones.

Dado que los puntos al borde de la imagen tienden a desaparecer incluso con traslaciones muy pequeñas (Figura 4.4), se ha buscado concentrar los puntos muestreados en el centro de la imagen, ignorando un 25 % de los bordes de la misma (Figura 4.5), lo que permite realizar traslaciones grandes manteniendo un gran porcentaje de co-visibilidad. De este modo se separa el problema de la co-visibilidad (que aparece por la naturaleza de las traslaciones simuladas) del problema de recuperación de escala.

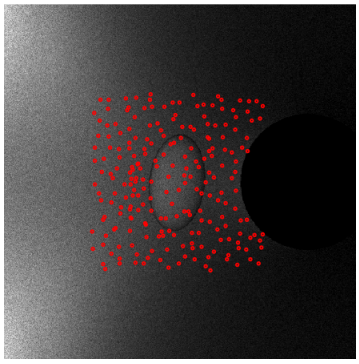


(a) Cámara de referencia

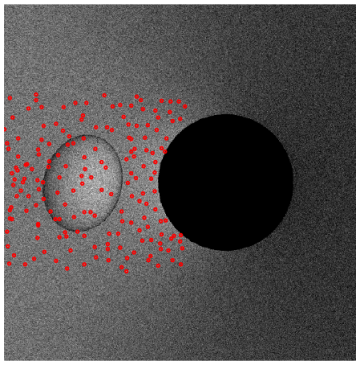


(b) Cámara trasladada 2b mm hacia la derecha.

Figura 4.4: 225 puntos muestreados con ray tracing en todo el dominio de la imagen, vistos desde varias cámaras. Una gran cantidad de puntos se pierde en la segunda imagen.



(a) Cámara de referencia



(b) Cámara trasladada 2b mm hacia la derecha.

Figura 4.5: 225 puntos muestreados con ray tracing ignorando el 25 % de los píxeles en los márgenes, vistos desde varias cámaras.

Capítulo 5

Experimentación mediante simulación

Utilizando el modelo de simulación descrito en el Capítulo 4, se han realizado varios experimentos para analizar la precisión de la metodología propuesta para estimación de escala.

En este capítulo se recopilan los experimentos más interesantes y representativos: se comienza por un caso muy similar al modelo simplificado expuesto en la Sección 3.1, con AGC conocido y solamente una fuente de luz. Posteriormente, se procede con un caso más realista en el que se desconoce el AGC y el endoscopio tiene 3 fuentes de luz. Con el tercer experimento se analiza el número de puntos requeridos para resolver el problema con el menor error posible y, con el cuarto, la importancia del número de fuentes de luz en el endoscopio. Se sigue con un análisis de la variabilidad aceptada en la estimación de I_0 al inicializar la optimización y se finaliza con un análisis del costo computacional de la solución.

Durante la experimentación, se utilizará la base (b) como distancia de referencia, puesto que es la única medida que se conoce con exactitud. Por ello se hablará de traslaciones y profundidades en función de la base. Por otro lado, se utilizará $\hat{x}/x$ como medida de error, donde $\hat{x}$ es la medida estimada y x es la medida real (*ground truth*). En cada caso de experimento se realizan 100 intentos distintos con distintas variaciones (por ruido) en la imagen original para analizar la variabilidad del error. En todos los casos, la búsqueda exhaustiva de λ se realiza en el dominio $[0, 6]$, siendo $\lambda_m = 0.5$ y se supone que se conoce I_0. El rango de búsqueda de λ es bastante amplio,

y en la práctica se podría reducir todavía más.

La simulación con la que se ha experimentado está basada en python, y utiliza la implementación de la librería *SciPy* [16] del método TRF [12].

5.1. AGC conocido, 1 fuente de luz

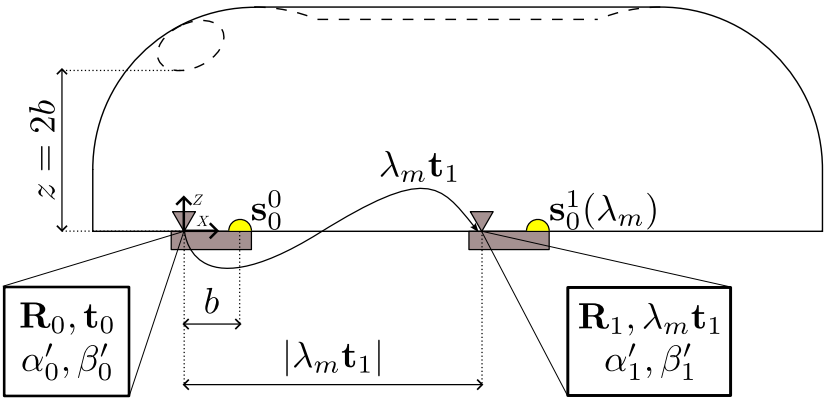


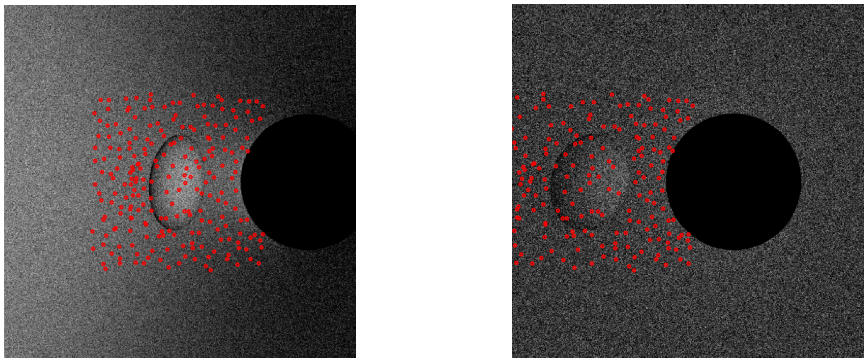
Figura 5.1: Representación del movimiento de la cámara durante los experimentos que requieren traslación.

En una aproximación simplista al problema, bastante similar a la detallada en la Sección 3.1, se supone un escenario de colonoscopia con una única fuente de luz, y en el que se conoce con exactitud el control de ganancia de la cámara (α'_k, β'_k). Se obtendrán dos imágenes: una desde el origen (cámara de referencia), y otra en la que la cámara se ha desplazado $\lambda_m t$ unidades en la dirección de la base (dirección positiva del eje x, tal y como se ilustra en la Figura 5.1). Se busca analizar el impacto de la traslación de la segunda cámara en el error de estimación de λ_m, para lo que se procede de la siguiente manera:

1. Se genera un conjunto de 225 puntos de manera uniforme en la región central de la imagen de la primera cámara.

2. Para cada traslación a analizar, se generan las imágenes que obtiene cada cámara. Se ignorarán los puntos que desaparezcan del campo de visión de la segunda cámara.

3. Se añade un ruido gaussiano de $\sigma = 2.5$ unidades de iluminación a la imagen y se estima λ_m. Se repite este paso 100 veces.



(a) Cámara 0. *(b) Cámara 1 cuando* $|\lambda_m \mathbf{t}_1| = 2b$

Figura 5.2: Imágenes y puntos vistos por un par de cámaras en una iteración del experimento con AGC conocido y una fuente de luz.

El experimento se ha realizado con traslaciones hasta 5 octavas por debajo y 2 octavas por encima de la base. Los resultados (Figura 5.3) muestran que cuando la traslación es mayor o igual que la base, se puede obtener un error que oscila entre el 1 % y el 2 %. En este caso se dice que la escena está bien condicionada. Cuando la traslación es muy pequeña, la escena no está bien condicionada y el error puede aproximarse al 20 %. También se aprecian problemas de co-visibilidad cuando la traslación es demasiado grande (4 veces la base), porque muchos puntos salen del campo de visión de la segunda cámara.

Por otro lado, en contraste con las técnicas tradicionales de *shape from shading* [4], esta metodología permite estimar los albedos con una precisión pareja a la de la escala cuando la escena está bien condicionada (1 % de error), tal y como se observa en la Figura 5.4. Si bien es cierto que la variación del error es igual, se pueden encontrar errores más exagerados en algunos los albedos, pues no se consigue la precisión del 1 % para todos los puntos.

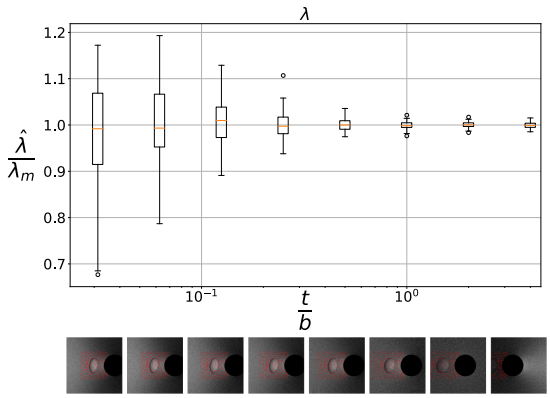


Figura 5.3: Error de estimación en escala en función de la traslación, con AGC conocido y 1 fuente de luz.

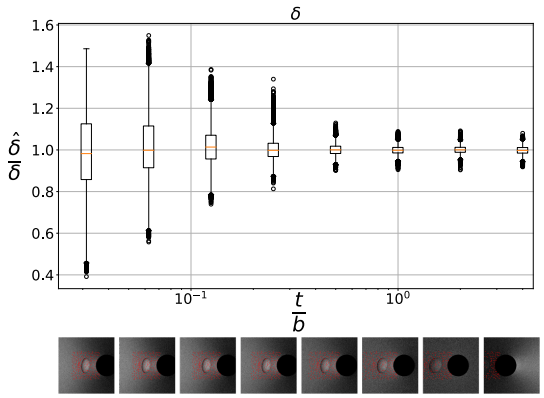


Figura 5.4: Error de estimación en albedos en función de la traslación, con AGC conocido y 1 fuente de luz.

En la Sección 3.1 se justificó que el tamaño de la base es un factor fundamental en el condicionamiento del problema. Dado que todo se referencia en relación a la base, acercar o alejar la escena de la cámara equivale a aumentar o disminuir la base de manera virtual, y por ende es de esperar que la profundidad del escenario influya en el condicionamiento de la escena. En busca de caracterizar esta relación, se ha repetido el experimento anterior pero variando la profundidad de la escena. Los resultados se analizan a continuación.

Por un lado, la mediana del error de estimación en escala y albedos (figuras 5.5b y 5.5d) muestra que la optimización converge al valor deseado, pues la mediana siempre oscila entorno a 1.

La desviación típica del error en escala expone cierta relación entre la profundidad y la traslación para el caso de escena bien condicionada. Conforme aumenta la profundidad, también aumenta la mínima traslación requerida, y cuanto mayor es la traslación, menor es la desviación típica del error. Esto implica, por tanto, que el condicionamiento de la escena depende de la traslación de la cámara en función de la profundidad de la escena. Ignorando los problemas de co-visibilidad, una traslación similar a la profundidad es lo que caracteriza una escena bien condicionada. En dichos casos, el error esperado es cercano al 2 % (desviación típica de 0.1).

Curiosamente, la varianza en el error de estimación de albedos (Figura 5.5c) no muestra la misma relación: el error disminuye al aumentar la profundidad y solo depende de la traslación cuando la escena es muy cercana. En escenas profundas, el error esperado es del 2 % (desviación típica 0.0125), mientras que en los peores casos supera el 6 %.

Ocurre que, cuando la escena está muy cerca de la cámara, el ángulo de incidencia de la fuente de luz sobre un punto es mucho más relevante que el albedo, lo que aumenta la tolerancia en el error del albedo al estimar la escala. Si todos los puntos están muy cerca y la traslación es muy pequeña, la iluminación del punto será prácticamente la misma en las dos imágenes. No obstante, si se traslada la cámara lo suficiente, el albedo y el ángulo de incidencia vuelven a tener una importancia similar, y en consecuencia el error es bajo.

Con este experimento se concluye que, con una sola fuente de luz, si el AGC es conocido y la traslación de la cámara es similar a la profundidad de la escena, es posible estimar la escala y los albedos con un error menor que el 2 %.

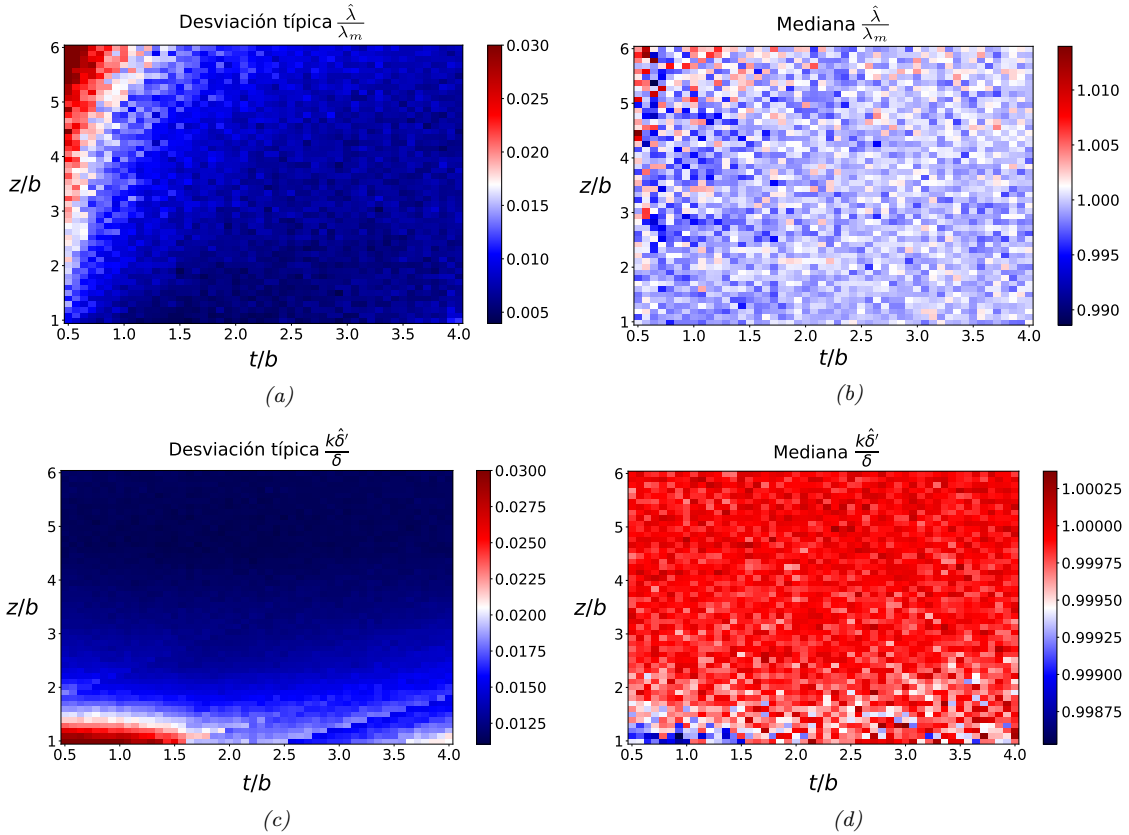


Figura 5.5: Desviación típica y mediana del error de estimación en escala y albedo a varias profundidades y traslaciones, con AGC conocido y 1 fuente de luz. La desviación típica se ha saturado en 0.03 por conveniencia en la representación de colores.

5.2. AGC desconocido, 3 fuentes de luz

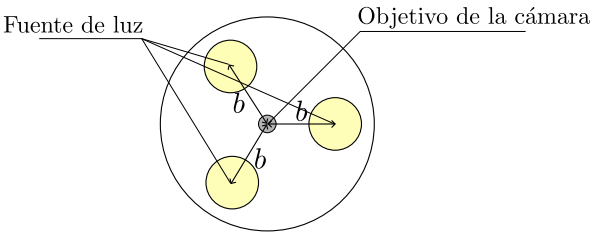


Figura 5.6: Posiciones de las fuentes de luz en relación con el objetivo de la cámara del endoscopio.

Se aborda un escenario que podría encontrarse en un entorno real: el endoscopio cuenta con tres fuentes de luz a una distancia b (Figura 5.6) y se desconoce el control de ganancia de cada cámara. El experimento se desarrolla igual que en el caso anterior: se captura una imagen con la cámara de referencia y otra en la que la cámara se ha desplazado cierta cantidad hacia la derecha. Se debe recordar que en este caso no es posible estimar los albedos reales, si no un albedo "escalado" que se alinea con los originales para poder realizar la comparativa (como se expuso en la Sección 3.2).

El experimento base, con los puntos a una profundidad $2b$, arroja las mismas conclusiones que el experimento anterior: la escena está bien condicionada cuando la traslación es mayor o igual que la base (figuras 5.7 y 5.8). De nuevo, el error en el caso óptimo se acerca al 2 %, pero por el contrario, en los casos desfavorables puede superar el 20 % (a diferencia del experimento anterior). Cuando la traslación es superior a $2b$, el error aumenta fruto de la pérdida de co-visibilidad, que afecta mucho más a la escala que a los albedos.

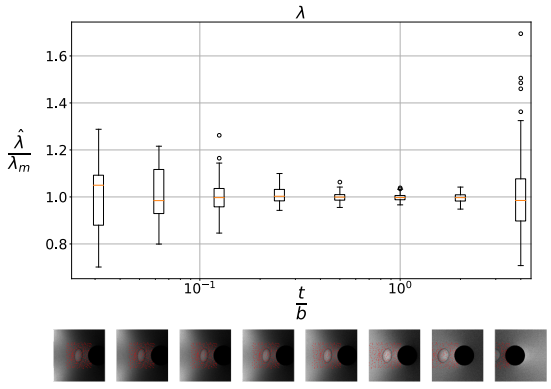


Figura 5.7: Error de estimación en escala en función de la traslación, con AGC desconocido y 3 fuentes de luz.

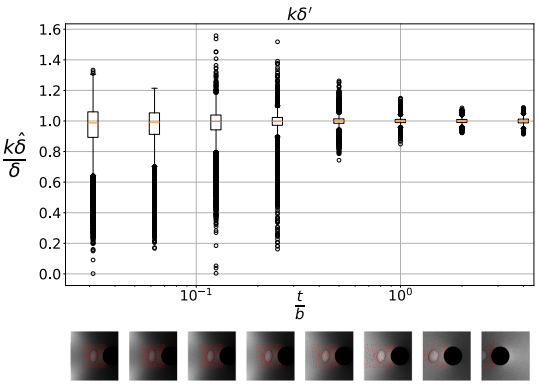


Figura 5.8: Error de estimación en albedos en función de la traslación, con AGC desconocido y 3 fuentes de luz.

En la Figura 5.9 se muestra la mediana y desviación típica del error de estimación en escala y albedos según la profundidad de la escena y la traslación de la cámara. La desviación típica del error en escala muestra una clara relación entre la traslación de la cámara y la profundidad de la escena.

Por un lado, se aprecia que a profundidades mayores que $3b$ el error aumenta y existe la posibilidad de optimización sesgada. Esto se aprecia claramente en la mediana del error de escala y albedo, pues hay un conjunto de casos en los que la mediana es 0.2, (un 80 % de error) y por tanto la optimización no tiende al valor real de la escala. Así pues, una escena bien condicionada debe tener como máximo una profundidad de $3b$. Se cree que el sesgo que aparece cuando la escena tiene de profundidad entre $3b$ y $4b$ pueda deberse al amplio rango de búsqueda en la estimación inicial de λ (entre 0 y 6), lo que provoca una mala estimación inicial que deriva en mala convergencia. Cuando la escena es menos profunda que $3b$, se requiere que la traslación sea similar a la profundidad para que la escena esté bien condicionada (error entre el 2 % y el 4 %). Si la escena es poco profunda (menor que $2b$), se admite casi cualquier traslación entre $0.5b$ y $2b$.

En la estimación de albedos no se aprecia la misma relación entre profundidad y traslación. Si la traslación y la profundidad son mayores que $2b$, el albedo se estima con un error entre el 2 % y el 4 % exceptuando los casos sesgados. Si la traslación es muy grande, aparecen problemas de co-visibilidad y el error aumenta y, si es menor que $2b$, el error mínimo es del 6 % y puede llegar a sobrepasar el 12 % cuando la escena es muy profunda. Debe destacarse que el error en escala no se propaga a la estimación de albedos.

Los problemas de co-visibilidad fruto de no rotar la cámara se aprecian claramente en la desviación típica del error, y son mucho más influyentes que en el experimento con AGC conocido. Cuando la traslación es mayor que $3b$, empiezan a desaparecer muchos puntos del campo de visión de la segunda cámara y el error aumenta notablemente. A razón de esto, en la siguiente sección se analizará el número de puntos que se requieren para obtener el máximo rendimiento del algoritmo.

Este experimento permite concluir, por un lado, que una escena bien condicionada es aquella con una profundidad menor que $3b$ y una traslación similar a su profundidad, en cuyo caso el error de estimación en la escala oscila entre el 2 % y el 4 %. Por suerte, el caso de escena bien condicionada es común en una endoscopia médica (cámara muy

cerca de la escena, y la traslación similar a la profundidad). Por otro lado, la estimación de albedos se realiza con un error menor que el 4 % siempre que profundidad y traslación sean mayores que $2b$. Si la traslación es menor que b, incluso en escenas bien condicionadas el error puede ser del 10 %.

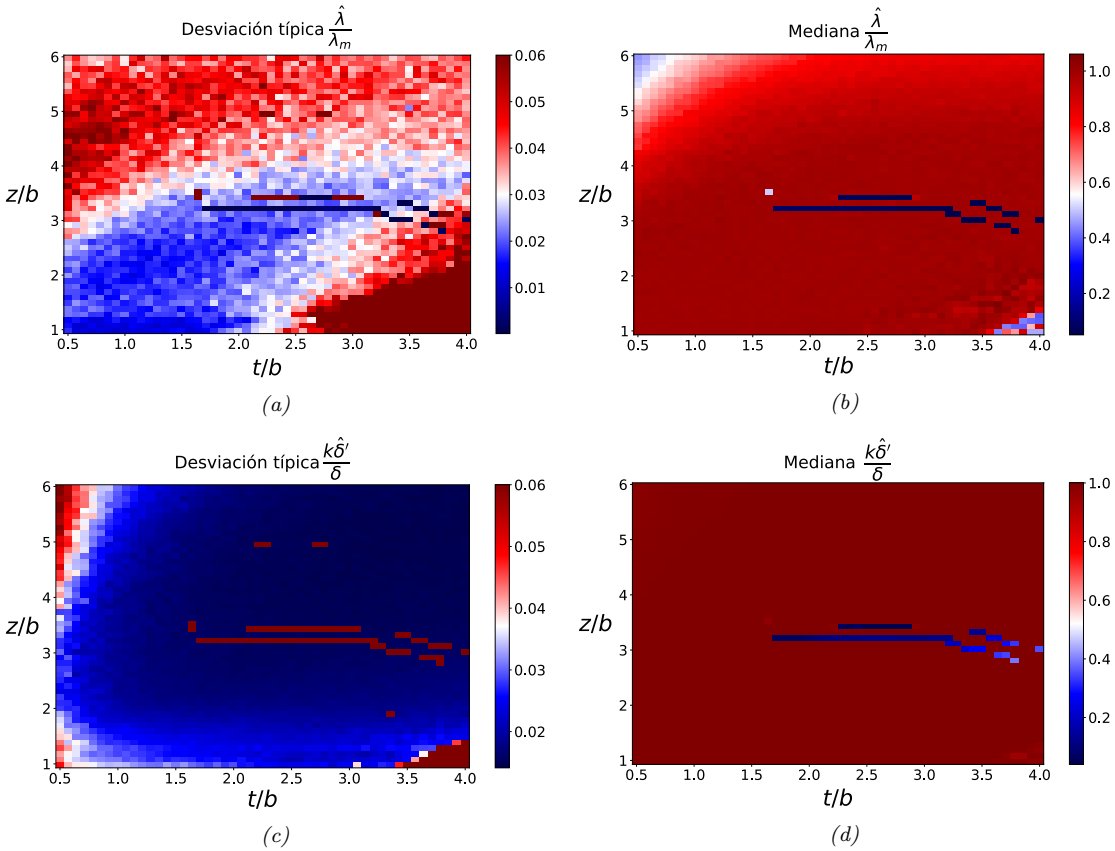


Figura 5.9: Desviación típica y mediana del error de estimación en escala y albedo a varias profundidades y traslaciones, con AGC desconocido y 3 fuentes de luz. La desviación típica se ha saturado en 0.06 para apreciar mejor los resultados en la escala de colores.

5.3. Influencia del número de puntos en la estimación

Si se supone AGC desconocido y 2 cámaras, es necesario disponer de al menos 5 puntos para que el sistema no sea indeterminado (detallado en la Sección 3.2). Este es, de manera teórica, el mínimo número de puntos que requiere el problema para recuperar la escala de la escena. No obstante, en la sección anterior se han detectado algunos problemas de co-visibilidad que permiten sospechar que el algoritmo requiere un número mínimo de puntos, superior al mínimo teórico, para llegar al máximo rendimiento en escenas bien condicionadas.

En esta sección se plantea un experimento en una escena bien condicionada, con profundidad $2b$, traslación b (Figura 5.1) y AGC desconocido, en la que se irán realizando distintas estimaciones variando el número de puntos disponibles (entre 10 y 400).

Los resultados son claros (Figura 5.10): con menos de 100 puntos, el error es más alto de lo que se ha visto en apartados anteriores. Incluso se puede observar un caso desastroso con un 90 % de error con 36 puntos. Por otro lado, con más de 100 puntos el error se estabiliza por debajo del 4 %, como se ha analizado anteriormente. De esto se concluye que se necesita un mínimo de 100 puntos para optimizar el rendimiento de la estimación.

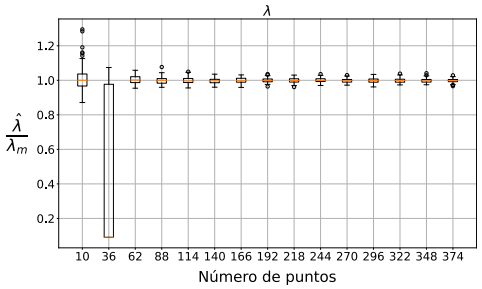


Figura 5.10: Variación del error de estimación en escala en función del número de puntos de la escena.

5.4. Influencia del número de fuentes de luz en la estimación

Intuitivamente, pareciera que con AGC desconocido pudiese haber alguna relación entre el número de fuentes de luz y el condicionamiento de la escena. En busca de esta posible relación, se ha realizado un experimento sencillo en una escena bien condicionada con profundidad y traslación b. Se han acercado los puntos a esta profundidad para aumentar la influencia de la iluminación en la imagen.

La prueba consiste en tratar de estimar la escala y los albedos variando el número de fuentes de luz que iluminan la escena. Para que el experimento tenga sentido se debe respetar b, pues posicionar las luces a una distancia mayor equivale a modificar la profundidad de la escena (que es relativa a b). Por ello, las luces se posicionaran de manera uniforme en un disco de radio b utilizando una distribución de cabeza de girasol [17]. En la Figura 5.11 se muestran las imágenes obtenidas con hasta 128 fuentes de luz. Se puede observar que la iluminación es cada vez más uniforme, pero el AGC

controla que la imagen no se sature completamente y las diferencias son cada vez más sutiles.

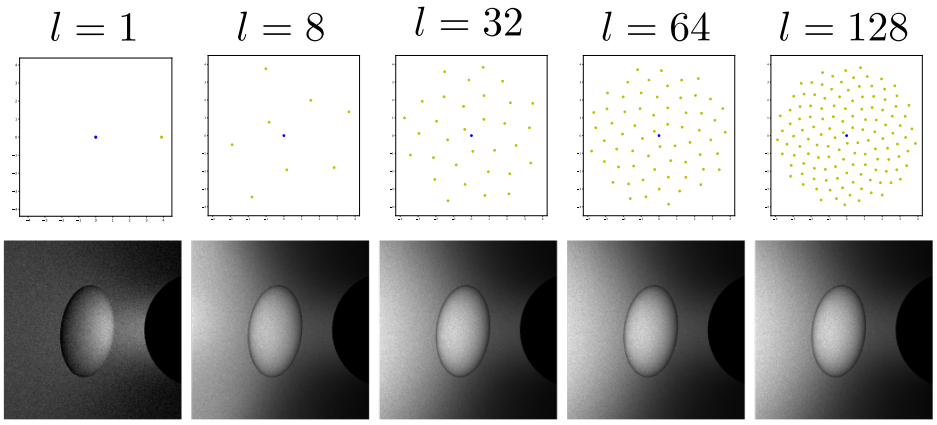


Figura 5.11: Imágenes captadas con la cámara de referencia en frente del pólipo (a profundidad b), variando el número de fuentes de luz (l), junto con la distribución de las fuentes de luz (amarillo) con respecto a la cámara (azul).

Los resultados muestran que la estimación de escala no es afectada por el número de fuentes de luz (Figura 5.12). El algoritmo necesita variación de luz entre las imágenes, de ahí la importancia de la traslación de la cámara. Por ello, no influye el número de luces mientras la traslación sea suficiente. Es decir, el algoritmo necesita que la iluminación varíe entre imágenes y no es afectado por la variación de iluminación dentro de una misma imagen.

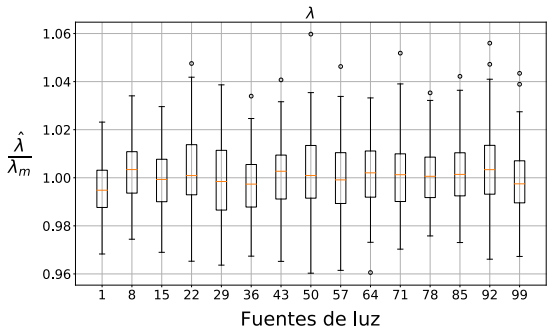


Figura 5.12: Error en la estimación de escala según el número de luces del endoscopio (distribución de cabeza de girasol), con profundidad b y traslación b.

5.5. Estimación inicial de la potencia de la fuente de luz

En el Capítulo 4 se adelantó que para inicializar α'_k (con AGC desconocido) se requería de una estimación de I_0. En esta sección se busca analizar la sensibilidad del optimizador ante la estimación inicial de I_0. Para ello, se ha seleccionado una escena bien condicionada: profundidad de $2b$, traslación b, y en ella se han realizado varias estimaciones variando el valor inicial de I_0 por un factor multiplicativo F, es decir: $\hat{I}_0 = FI_0$. El experimento se ha realizado con 1 y 3 fuentes de luz.

Con una fuente de luz (Figura 5.13), se pueden obtener buenas estimaciones con valores de I_0 iniciales entre 0.8 y 10 veces el valor de I_0. Por otro lado, cuando se trabaja con tres fuentes de luz, el límite inferior, si existe, está por debajo de 0.01, mientras que el superior está en $100I_0$. El comportamiento en el caso de los albedos es idéntico.

En conclusión, si se tiene una estimación de la potencia de las bombillas en base a la documentación del endoscopio no habrá problema porque muy probablemente esté dentro del rango aceptado. Si por el contrario, se desconoce completamente I_0, se pueden realizar un par de pruebas con valores muy pequeños de I_0, ya que no parece haber un límite inferior en el caso de AGC desconocido. Además, el rango de esta estimación inicial crece con el número de luces.

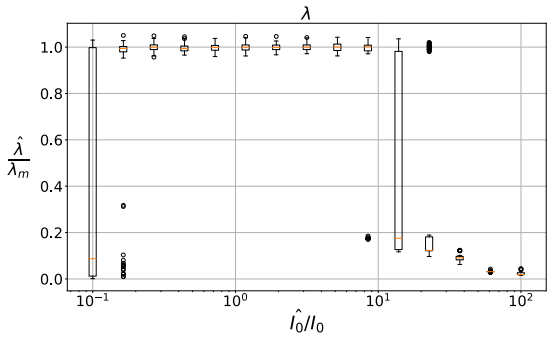
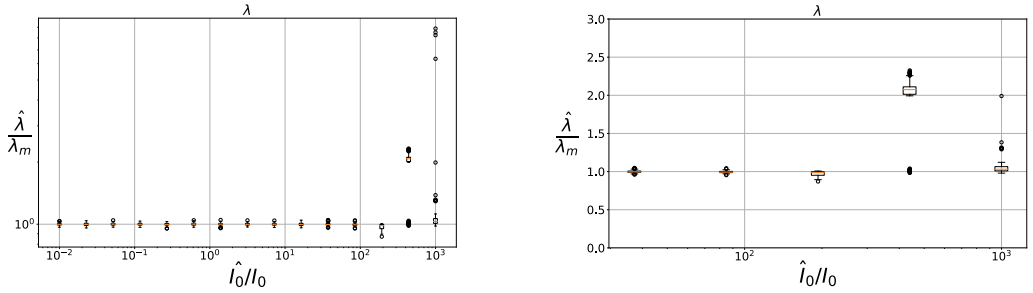


Figura 5.13: Error en la estimación de escala según la estimación inicial de I_0 con una fuente de luz.



(a) *(b) Zoom en la zona crítica*

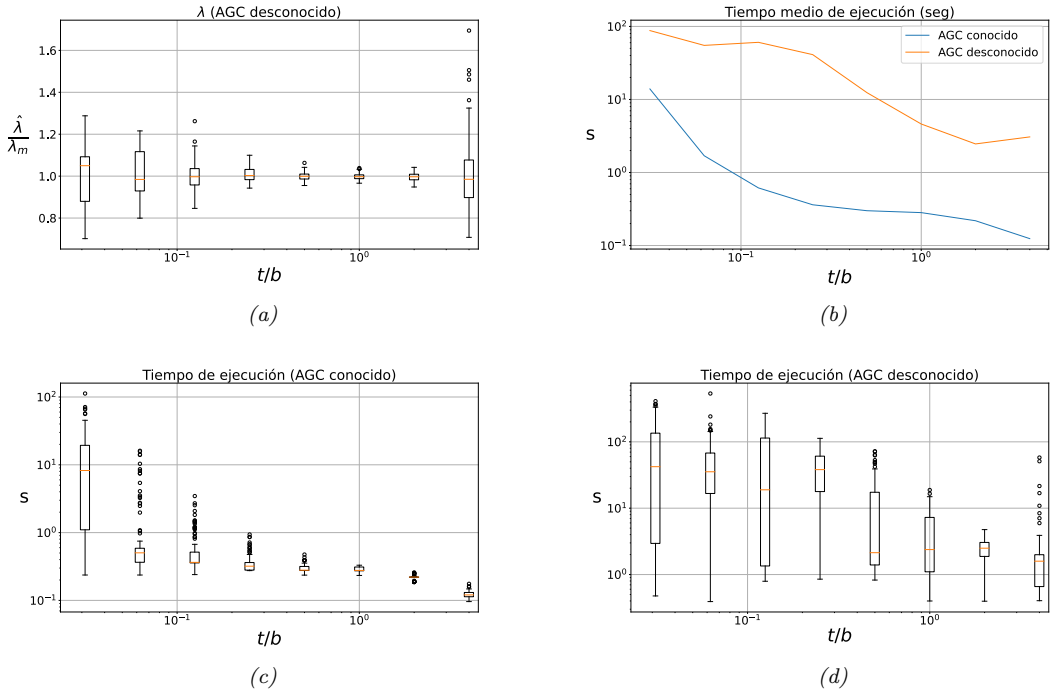
Figura 5.14: Error en la estimación de escala según la estimación inicial de I_0 con tres fuentes de luz.

5.6. Coste computacional

El tiempo de convergencia de la optimización es un aspecto fundamental al analizar la viabilidad de esta metodología. Es importante caracterizar el comportamiento temporal de la optimización en función de sus variables principales, en este caso: el condicionamiento de la escena, el número de puntos, la estimación de I_0 y el conocimiento o desconocimiento del AGC.

El análisis se centra en el caso de AGC desconocido por ser el caso más realista. Todas las pruebas se han ejecutado en un sistema linux con CPU `Intel i7-9700K @ 3.60GHz` con uso exclusivo para estos experimentos, y suponiendo un caso de escena bien condicionada: profundidad de $2b$, 3 fuentes de luz y un desplazamiento en la cámara de b mm (Figura 5.1).

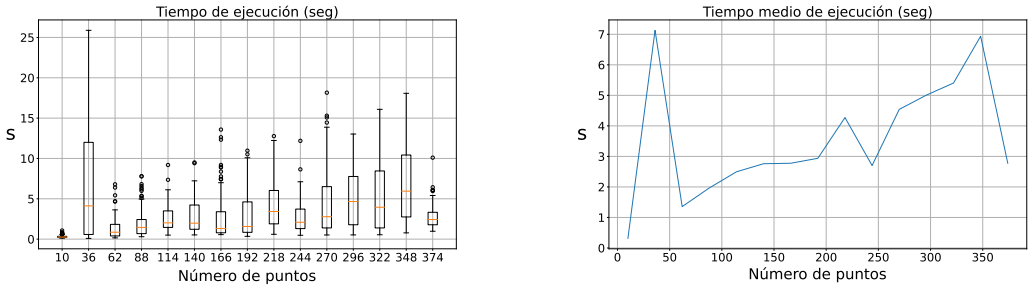
En primer lugar, se ha realizado una comparación entre el tiempo de ejecución cuando se conoce y se desconoce el control de ganancia. Se ha procedido con el experimento de la misma manera que en la Sección 5.1, y en la Figura 5.15 se muestran los resultados más relevantes. Destaca una clara relación entre el tiempo de cómputo y el condicionamiento de la escena, pues en los casos bien condicionados el tiempo de ejecución es de unos 25 ms, mientras que en los casos mal condicionados se alcanza hasta los 6 minutos de ejecución. Por otro lado, conocer el control de ganancia supone una reducción de un orden de magnitud en el tiempo de ejecución (Figura 5.15b).



(a) *(b)* *(c)* *(d)*

Figura 5.15: Comparativa de tiempos de ejecución según la traslación de la cámara, conociendo y desconociendo control de ganancia

En lo referente al número de puntos de la escena, para el experimento realizado en la Sección 5.3, el tiempo de ejecución es el mostrado en la Figura 5.16. Se aprecia cierta tendencia ascendente en el tiempo de ejecución con el incremento del número de puntos de la escena. Sin embargo, el condicionamiento de la escena sigue siendo más relevante. Destaca el caso de 36 puntos, en el que el tiempo de ejecución es el más alto porque la escena no está bien condicionada.

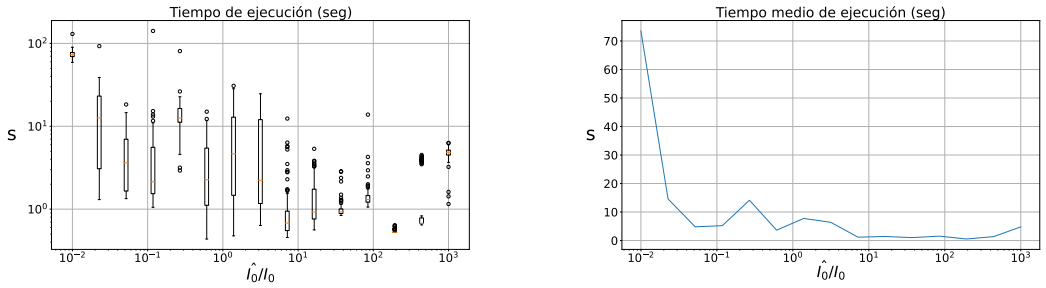


(a) *(b)*

Figura 5.16: Tiempo de convergencia de la optimización en función del número de puntos de la escena.

Por último, es interesante analizar el impacto de la estimación inicial de I_0 en el tiempo de ejecución. Para el experimento ejecutado en la Sección 5.5, los tiempos de

ejecución son los adjuntados en la Figura 5.17. Se aprecia que cuanto menor es el valor inicial de I_0, más tiempo de ejecución requiere el optimizador. En este caso, el tiempo de ejecución no se relaciona con la viabilidad de la estimación de I_0, pues cuando $\hat{I}_0 \simeq 100 I_0$ el error de estimación aumentaba notablemente y sin embargo el tiempo de ejecución no lo hace.



(a) *(b)*

Figura 5.17: Tiempo de convergencia de la optimización en función de la estimación inicial de I_0.

En conclusión, si la escena está bien condicionada y se dispone de una estimación de I_0 en $[0.1I_0, 100I_0]$, es posible obtener un resultado en un máximo de 25 ms. En caso contrario, la ejecución puede alargarse hasta los 6 minutos o más.

Capítulo 6

Conclusión

En este trabajo se ha propuesto una solución para la estimación de escala métrica de una escena observada por una cámara monocular móvil con iluminación co-localizada. A diferencia de las técnicas tradicionales de estéreo fotométrico, este método se sustenta en el movimiento de la cámara y en la base (distancia entre la(s) fuente(s) de luz y el centro óptico de la cámara), y es capaz de estimar el albedo de la escena.

Esta metodología se basa en un modelo fotométrico simplificado que supone una escena de materiales lambertianos, ignora efectos de iluminación global y modela el AGC como una función afín. La escala se estima con una optimización por mínimos cuadrados por el método TRF. Finalmente, la metodología se ha analizado experimentalmente en una simulación de colonoscopia basada en primitivas geométricas sencillas (elipsoides, cilindros y toroides), en la que el movimiento de la cámara se limita a un solo eje.

El algoritmo requiere de al menos dos imágenes en escala de grises, 100 puntos emparejados y una estimación de la potencia de las fuentes de luz (I_0) entre 0.01 y 100 veces su valor real. El rendimiento del algoritmo depende de que el condicionamiento de la escena le sea favorable o no. Una escena bien condicionada es aquella con con una profundidad menor que $3b$ y una traslación de la cámara similar a dicha profundidad, aunque si la escena es poco profunda, se admiten traslaciones algo menores. En estos casos, el error en la estimación de la escala está entre el 2 % y el 4 % (unas 2 décimas de mm) y la solución se obtiene en unos 25 ms. En caso contrario, el error puede sobrepasar el 20 % y la ejecución puede durar varios minutos. La estimación de albedos, por su parte, tiene un error menor que el 4 % siempre que profundidad y traslación sean

mayores que $2b$.

También se ha observado que el número de luces de la escena no influye en el error de estimación, y que, en caso de conocer el AGC de la cámara, el error se puede reducir al 1 % y el tiempo de ejecución disminuirse en un orden de magnitud.

En un escenario de endoscopia, donde la cámara está muy cerca de la escena y se mueve bastante, es factible encontrar una escena bien condicionada, y por tanto la recuperación de la escala es posible.

6.1. Trabajo futuro

Este trabajo supone una toma de contacto con un problema complejo. Por ello se realizan varias simplificaciones y se aborda el análisis en un entorno simulado. Hay una clara línea de trabajo para poder llegar a aplicar esta solución a un entorno real:

- **Modelado fotométrico completo**: se ha supuesto una escena lambertiana de albedo aleatorio pero uniforme, y se han ignorado efectos de iluminación global. En una escena de endoscopia, no solo es común encontrar materiales reflectantes si no que la iluminación global juega un papel importante en el alumbrado. Para aplicar esta solución a un entorno real es posible que se requiera de un modelado fotométrico más complejo.

- **Aplicación en entornos más realistas**: las pruebas se realizan en una simulación con geometría sencilla. Antes de aplicar esta solución a un entorno real, es de interés observar y pulir esta metodología en entornos simulados con más realismo.

- **Modelo de cámara**: el modelo de cámara pinhole no es el más adecuado para tratar la formación de imágenes en endoscopia. Por ejemplo, entre otros aspectos, no se ha modelado el efecto de *ojo de pez* que presentan este tipo de cámaras.

- **Ruido en la geometría escalada**: en este trabajo se ha supuesto que no hay ruido en la geometría cuya escala se quiere recuperar. En la práctica, el proceso de SfM no recuperará la geometría exacta y ello influirá en la estimación de escala.

- **Calibración fotométrica y geométrica de la cámara**: En la práctica, la formación de imágenes en una cámara es compleja y está sujeta a varios fenómenos

físicos que perturban la imagen generada: viñeteo, distorsión de ojo de pez, enfoque de la cámara, etc. Mediante un proceso de calibración de la cámara es posible detectar y corregir estos fenómenos, lo que es necesario para aplicar esta solución a la realidad. Del mismo modo, se pueden calibrar el perfil de iluminación de cada una de las fuentes de luz de la cámara.

- **Estimación automática del dominio factible para la escala**: En la metodología propuesta se realiza una estimación inicial de la escala basada en una búsqueda exhaustiva en el dominio de la función objetivo. El dominio de búsqueda se ha establecido manualmente, pero, en la práctica, podría automatizarse y ajustarse al máximo. Solamente es necesario una estimación humana del tamaño que tendría la escena, por ejemplo, saber que el pólipo tiene un tamaño entre 2 mm y 4 mm. Esto podría mejorar tanto los resultados como el tiempo de ejecución de la estimación.

Apéndice A

Jacobiano analítico de L_i^k

Para simplificar notación, se supone que $L_i^k = L_i^k(\lambda, \alpha'_0, ..., \alpha'_{m-1}, \beta'_0, ..., \beta'_{m-1}, \delta_0, ..., \delta_{n-1})$, definido en la Ecuación 2.10. Se suponen n puntos, m cámaras y l fuentes de luz para cada cámara. El jacobiano correspondiente es:

$$J_n^m = \begin{bmatrix} \frac{\partial L_0^0}{\partial \lambda} & \frac{\partial L_0^0}{\partial \alpha'_0} & \cdots & \frac{\partial L_0^0}{\partial \alpha'_{m-1}} & \frac{\partial L_0^0}{\partial \beta'_0} & \cdots & \frac{\partial L_0^0}{\partial \beta'_{m-1}} & \frac{\partial L_0^0}{\partial \delta_0} & \cdots & \frac{\partial L_0^0}{\partial \delta_{n-1}} \\ \frac{\partial L_0^1}{\partial \lambda} & \frac{\partial L_0^1}{\partial \alpha'_0} & \cdots & \frac{\partial L_0^1}{\partial \alpha'_{m-1}} & \frac{\partial L_0^1}{\partial \beta'_0} & \cdots & \frac{\partial L_0^1}{\partial \beta'_{m-1}} & \frac{\partial L_0^1}{\partial \delta_0} & \cdots & \frac{\partial L_0^1}{\partial \delta_{n-1}} \\ \vdots & & & & & \vdots & & & & \vdots \\ \frac{\partial L_{n-1}^{m-2}}{\partial \lambda} & \frac{\partial L_{n-1}^{m-2}}{\partial \alpha'_{m-2}} & \cdots & \frac{\partial L_{n-1}^{m-2}}{\partial \alpha'_{m-1}} & \frac{\partial L_{n-1}^{m-2}}{\partial \beta'_0} & \cdots & \frac{\partial L_{n-1}^{m-2}}{\partial \beta'_{m-1}} & \frac{\partial L_{n-1}^{m-2}}{\partial \delta'_0} & \cdots & \frac{\partial L_{n-1}^{m-2}}{\partial \delta_{n-1}} \\ \frac{\partial L_{n-1}^{m-1}}{\partial \lambda} & \frac{\partial L_{n-1}^{m-1}}{\partial \alpha'_0} & \cdots & \frac{\partial L_{n-1}^{m-1}}{\partial \alpha'_{m-1}} & \frac{\partial L_{n-1}^{m-1}}{\partial \beta'_0} & \cdots & \frac{\partial L_{n-1}^{m-1}}{\partial \beta'_{m-1}} & \frac{\partial L_{n-1}^{m-1}}{\partial \delta_0} & \cdots & \frac{\partial L_{n-1}^{m-1}}{\partial \delta_{n-1}} \end{bmatrix} \tag{A.1}$$

La mayoría de las componentes de J_n^m valen 0, por ejemplo, con $n = 2$ y $m = 2$:

$$J_2^2 = \begin{bmatrix} \frac{\partial L_0^0}{\partial \lambda} & \frac{\partial L_0^0}{\partial \alpha'_0} & 0 & \frac{\partial L_0^0}{\partial \beta'_0} & 0 & \frac{\partial L_0^0}{\partial \delta_0} & 0 \\ \frac{\partial L_0^1}{\partial \lambda} & \frac{\partial L_0^1}{\partial \alpha'_0} & \frac{\partial L_0^1}{\partial \alpha'_1} & 0 & \frac{\partial L_0^1}{\partial \beta'_1} & \frac{\partial L_0^1}{\partial \delta_0} & 0 \\ \frac{\partial L_1^0}{\partial \lambda} & \frac{\partial L_1^0}{\partial \alpha'_0} & 0 & \frac{\partial L_1^0}{\partial \beta'_0} & 0 & 0 & \frac{\partial L_1^0}{\partial \delta_1} \\ \frac{\partial L_1^1}{\partial \lambda} & \frac{\partial L_1^1}{\partial \alpha'_0} & \frac{\partial L_1^1}{\partial \alpha'_1} & 0 & \frac{\partial L_1^1}{\partial \beta'_1} & 0 & \frac{\partial L_1^1}{\partial \delta_1} \end{bmatrix} \tag{A.2}$$

Las derivadas parciales son tal que:

$$\frac{\partial L_i^k}{\partial \delta_i} = \begin{cases} \alpha_0' \sum_{j=1}^{l} \frac{\mathbf{n}_i \cdot \mathbf{l}_i^{jk}(\lambda)}{|\mathbf{l}_i^{jk}(\lambda)|^3} & \text{si } k = 0 \\ \alpha_k' \alpha_0' \sum_{j=1}^{l} \frac{\mathbf{n}_i \cdot \mathbf{l}_i^{jk}(\lambda)}{|\mathbf{l}_i^{jk}(\lambda)|^3} & \text{si } k > 0 \end{cases} \tag{A.3}$$

$$\frac{\partial L_i^k}{\partial \beta_j} = 1 \tag{A.4}$$

$$\frac{\partial L_i^k}{\partial \alpha_0'} = \begin{cases} \delta_i \sum_{j=1}^{l} \frac{\mathbf{n}_i \cdot \mathbf{l}_i^{jk}(\lambda)}{|\mathbf{l}_i^{jk}(\lambda)|^3} & \text{si } k = 0 \\ \alpha_k' \delta_i \sum_{j=1}^{l} \frac{\mathbf{n}_i \cdot \mathbf{l}_i^{jk}(\lambda)}{|\mathbf{l}_i^{jk}(\lambda)|^3} & \text{si } k > 0 \end{cases} \tag{A.5}$$

$$\frac{\partial L_i^k}{\partial \alpha_k'} = \begin{cases} \delta_i \sum_{j=1}^{l} \frac{\mathbf{n}_i \cdot \mathbf{l}_i^{jk}(\lambda)}{|\mathbf{l}_i^{jk}(\lambda)|^3} & \text{si } k = 0 \\ \alpha_0' \delta_i \sum_{j=1}^{l} \frac{\mathbf{n}_i \cdot \mathbf{l}_i^{jk}(\lambda)}{|\mathbf{l}_i^{jk}(\lambda)|^3} & \text{si } k > 0 \end{cases} \tag{A.6}$$

$$\frac{\partial L_i^k}{\partial \lambda} = \sum_{j=1}^{l} \frac{a_{ik}' b_{ijk} - a_{ijk} b_{ijk}'}{b_{ijk}^2} \tag{A.7}$$

$$a_{ijk} = \begin{cases} \alpha_k' \delta_i \left(\mathbf{n}_i \cdot \mathbf{l}_i^{jk}(\lambda) \right) & \text{si } j = 0 \\ \alpha_k' \alpha_0' \delta_i \left(\mathbf{n}_i \cdot \mathbf{l}_i^{jk}(\lambda) \right) & \text{si } j > 0 \end{cases} \tag{A.8}$$

$$a_{ik}' = \begin{cases} \alpha_k' \delta_i \left(\mathbf{n}_i \cdot \left(-\mathbf{R}_k^\top \mathbf{t}_k - \mathbf{X}_i \right) \right) & \text{si } k = 0 \\ \alpha_k' \alpha_0' \delta_i \left(-\mathbf{R}_k^\top \mathbf{t}_k - \mathbf{X}_i \right) & \text{si } k > 0 \end{cases} \tag{A.9}$$

$$b_{ijk} = \left| \mathbf{l}_i^{jk}(\lambda) \right|^3 \tag{A.10}$$

$$b_{ijk}' = 3 \left| \mathbf{l}_i^{jk}(\lambda) \right| \left(\mathbf{l}_i^{jk}(\lambda) \cdot \left(-\mathbf{R}_k^\top \mathbf{t}_k - \mathbf{X}_i \right) \right) \tag{A.11}$$

Apéndice B

Gestión de proyecto

Este proyecto se ha realizado gracias a una Beca de Colaboración del MEFP (Ministerio de Educación y Formación Profesional). La colaboración ha durado cerca de 7 meses y medio y en la Tabla B.1 se muestra el tiempo dedicado a cada tarea del proyecto. En la Figura B.1 se adjunta un diagrama con la planificación temporal del mismo.

Tarea	Horas
Revisión de teoría y bibliografía	15
Desarrollo de las bases de la simulación	30
Análisis con una cámara estática	60
Análisis con cámara móvil	149
- Modelo fotométrico relativo a cámara de referencia	12
- Optimización no lineal (AGC conocido y desconocido)	80
- Prueba con puntos especulares de albedo conocido	20
- Integración de múltiples fuentes de luz	6
- Análisis de variabilidad de $\hat{I}_0$	8
- Análisis de estimación en función de la profundidad	15
- Análisis de número de puntos	4
- Análisis del coste computacional	4
Ray tracing	8
Simulación de escenarios simples	30
Simulación de escenario de colonoscopia real (escena con pólipo)	6
Refactoring	15
Reuniones de coordinación	40
Memoria y presentación	75
Total	**428**

Cuadro B.1: Tiempo dedicado a cada tarea del proyecto.

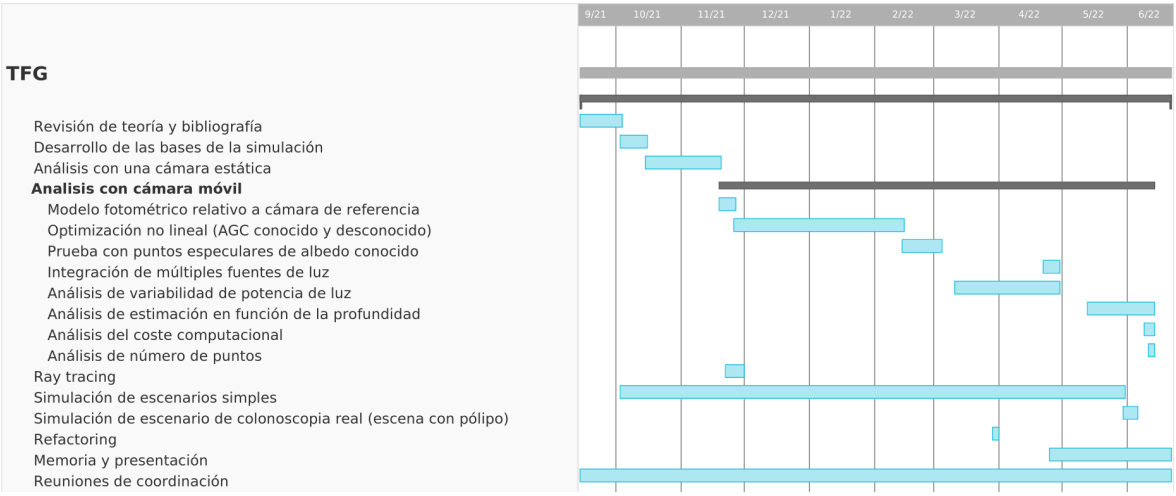


Figura B.1: Diagrama de Gantt del proyecto.

Bibliografía

[1] J. L. Schonberger y J.-M. Frahm, «Structure-from-motion revisited,» en *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, págs. 4104-4113.

[2] Y. Iwahori, H. Sugie y N. Ishii, «Reconstructing shape from shading images under point light source illumination,» en *[1990] Proceedings. 10th International Conference on Pattern Recognition*, IEEE, vol. 1, 1990, págs. 83-87.

[3] C. Wu, S. G. Narasimhan y B. Jaramaz, «A multi-image shape-from-shading framework for near-lighting perspective endoscopes,» *International Journal of Computer Vision*, vol. 86, n.º 2, págs. 211-228, 2010.

[4] B. K. Horn, «Shape from shading; a method for obtaining the shape of a smooth opaque object from one view.,» Tesis doct., Massachusetts Institute of Technology, 1970.

[5] R. J. Woodham, «Photometric method for determining surface orientation from multiple images,» *Optical engineering*, vol. 19, n.º 1, págs. 139-144, 1980.

[6] J. J. Clark, «Active photometric stereo.,» en *CVPR*, vol. 92, 1992, págs. 29-34.

[7] Y. Furukawa y C. Hernández, «Multi-view stereo: A tutorial,» *Foundations and Trends® in Computer Graphics and Vision*, vol. 9, n.º 1-2, págs. 1-148, 2015.

[8] T. Collins y A. Bartoli, «3d reconstruction in laparoscopy with close-range photometric stereo,» en *International conference on medical image computing and computer-assisted intervention*, Springer, 2012, págs. 634-642.

[9] V. Martínez Batlle y J. D. Tardós Solano, «Reconstrucción 3D a escala real a partir de imágenes monoculares de endoscopio.,» 2021.

[10] K. Madsen, H. B. Nielsen y O. Tingleff, «Methods for non-linear least squares problems,» 2004.

[11] Y.-x. Yuan, «A review of trust region algorithms for optimization,» en *Iciam*, vol. 99, 2000, págs. 271-282.

[12] SciPy. «scipy.optimize.least_squares.» (), dirección: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html (visitado 19-06-2022).

[13] P. Azagra, C. Sostres, Á. Ferrandez y col., «EndoMapper dataset of complete calibrated endoscopy procedures,» *arXiv preprint arXiv:2204.14240*, 2022.

[14] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[15] A. S. Glassner, *An introduction to ray tracing*. Morgan Kaufmann, 1989.

[16] SciPy. «SciPy Documentation.» (), dirección: https://docs.scipy.org/doc/scipy/ (visitado 19-06-2022).

[17] A. Mathai y T. Davis, «Constructing the sunflower head,» *Mathematical Biosciences*, vol. 20, n.º 1, págs. 117-133, 1974, ISSN: 0025-5564.