

Trabajo Fin de Grado

Evaluación de métodos densos de reconstrucción 3D a partir de imágenes

Autor

Rosa Castellón Lacasa

Director

Alejo Concha Belenguer

Ponente

Javier Civera Sancho

Escuela de Ingeniería y Arquitectura
2015



DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD

(Este documento debe acompañar al Trabajo Fin de Grado (TFG)/Trabajo Fin de Máster (TFM) cuando sea depositado para su evaluación).

TRABAJOS DE FIN DE GRADO / FIN DE MÁSTER

D./D^a. Rosa Castellón Lacasa

con nº de DNI 73205230 W en aplicación de lo dispuesto en el art.

14 (Derechos de autor) del Acuerdo de 11 de septiembre de 2014, del Consejo de Gobierno, por el que se aprueba el Reglamento de los TFG y TFM de la Universidad de Zaragoza,

Declaro que el presente Trabajo de Fin de (Grado/Máster)
Grado en Ingeniería Electrónica y Automática, (Título del Trabajo)
Evaluación de métodos densos de reconstrucción 3D a partir de imágenes.

es de mi autoría y es original, no habiéndose utilizado fuente sin ser citada debidamente.

Zaragoza, 26 de Junio de 2015

Fdo: Rosa Castellón Lacasa

A mi abuelo Joaquín.

RESUMEN

EVALUACIÓN DE MÉTODOS DENSOS DE RECONSTRUCCIÓN 3D A PARTIR DE IMÁGENES

La reconstrucción 3D es uno de los campos más relevantes de la visión por computador. En la actualidad son muchas las aplicaciones que utilizan o podrían utilizar este tipo de reconstrucciones; por ejemplo navegación de robots, topografía y realidad virtual.

Existen varias técnicas de reconstrucción 3D. En este trabajo se centra la atención en la visión estéreo multivista; y en concreto en las reconstrucciones densas. A diferencia de las técnicas tradicionales no densas, cuya limitación es que sólo pueden reconstruir unos cuantos cientos de puntos salientes, en este tipo de reconstrucciones densas se estiman todos los puntos de la escena. Para ello se utiliza una secuencia de imágenes de la misma escena, tomadas desde distintos puntos de vista cercanos.

Este tipo de técnicas son muy recientes y están todavía en fase de desarrollo. En este proyecto se ha utilizado la técnica DTAM (Dense Tracking and Mapping) de 2011. Este algoritmo reconstruye el mapa de la escena y estima la posición de todas las cámaras. En este caso se ha estudiado únicamente, la fase del mapeo de la escena. El objetivo es mejorar la precisión del algoritmo, y para ello se han evaluado diferentes parámetros.

Se utiliza el color de los píxeles como identificador para calcular la profundidad, en este proyecto se ha aumentado el tamaño de este identificador y se han evaluado diferentes métodos para la comparación de identificadores. También se ha estudiado como afecta el regularizador en las reconstrucciones, el cual impone que píxeles cercanos tengan profundidad similar en caso de que tengan color parecido. Todos estos parámetros se han ajustado mediante experimentación. Para calcular la mejora de los resultados se han comparado las profundidades estimadas con el algoritmo con un ground truth creado a partir de datos extraídos de un sensor Kinect.

Los resultados muestran que es posible mejorar la precisión en las reconstrucciones mediante el ajuste de los parámetros estudiados. En concreto la mejora obtenida en este TFG es del 40 %.

Índice general

1. Introducción	1
1.1. Objetivos	4
2. Dense Tracking and Mapping	7
2.1. Reconstrucción densa y no densa	7
2.2. Dense Mapping	8
2.2.1. Término Fotométrico	9
2.2.2. Término Regularizador	10
2.2.3. Minimización de la energía	11
2.2.4. Comparación de identificadores	13
2.3. Dense tracking	14
3. Resultados experimentales	15
3.1. Tamaño del identificador	15
3.2. Comparación de identificadores	21
3.3. Influencia del regularizador	23
4. Conclusiones	29
Bibliografía	31

Capítulo 1

Introducción

La inteligencia artificial es una rama de la ciencia computacional cuyo objetivo es la creación de máquinas inteligentes. Es decir, máquinas que sean capaces de aprender nuevos conceptos y tareas a partir de datos y, además puedan razonar de manera abstracta sobre dichos conceptos. La inteligencia artificial tiene su origen en los años cincuenta, cuando se planteó la pregunta “¿puede pensar una máquina?”

Las ramas más importantes de la inteligencia artificial son los sistemas basados en el conocimiento, aprendizaje y razonamiento automático, la robótica y la visión por computador. Todas estas disciplinas están en constante evolución y todavía tienen numerosos problemas abiertos.

El objetivo de la visión por computador es la extracción de información a partir de imágenes. Los humanos somos capaces de reconocer objetos, personas o acciones que se realizan en una escena en imágenes o vídeos. La visión por computador intenta que las máquinas imiten estas habilidades.

Uno de los campos más activos de la visión artificial es la reconstrucción tridimensional. En los últimos años, los algoritmos para la reconstrucción 3D han recibido atención significativa ya que, son utilizados en una gran variedad de aplicaciones. Sirven como herramienta para la medicina, la robótica, la arqueología, o los procesos de fabricación.

La reconstrucción 3D tiene como objetivo principal crear un modelo en 3 dimensiones a partir de una secuencia de imágenes, recuperando los valores de profundidad de cada pixel [2]. Así, por ejemplo, podemos utilizar la reconstrucción de una escena para que un robot navegue por ella sin colisionar con los objetos. Existen diversas técnicas de reconstrucción que podemos clasificar en dos grupos: métodos activos y métodos pasivos.

Métodos activos

Los métodos activos son aquéllos que precisan alguna acción sobre la escena a reconstruir. Entre ellos se encuentran la iluminación láser, los métodos basados en la proyección de luz estructurada o los de tiempo de vuelo [7].

La telemetría láser, por ejemplo, consiste en determinar el mapa de profundidades en base al tiempo transcurrido entre la emisión y la detección de un pulso láser. Lo mismo ocurre con las técnicas de tiempos de vuelo, siendo en este caso la naturaleza de la señal emitida acústica o electromagnética. Podemos ver un ejemplo de una de estas técnicas en la Figura 1.1.

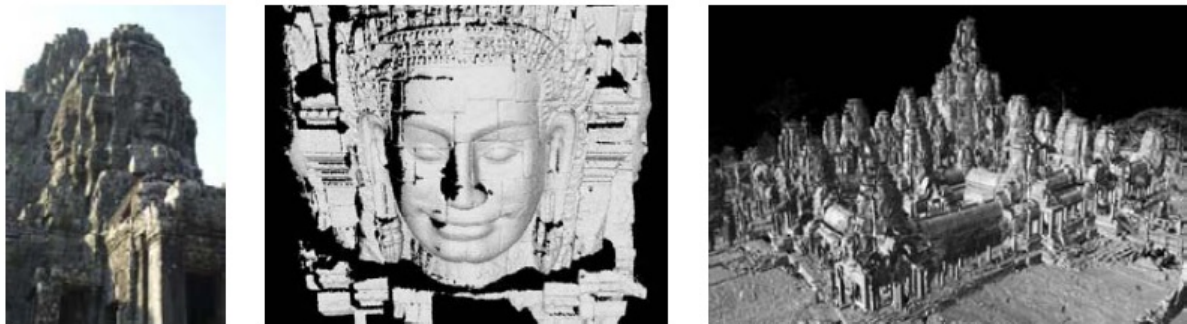


Figura 1.1: Reconstrucción utilizando láser y visión. Imagen extraída de [6]

Métodos pasivos

Por otro lado los métodos pasivos son los que no precisan de ninguna de estas interferencias con la escena y se basan únicamente en las imágenes para realizar la reconstrucción. Para ello, se necesitan, en general, varias imágenes de la misma escena. Se puede estimar la profundidad de imágenes 2D a partir de las sombras, la intersección de volúmenes, o la visión estereoscópica sin calibrar.

Este trabajo se centra en la visión estéreo multivista. Conforme a este método, se toman varias imágenes de una misma escena desde distintos puntos de vista, y a partir de ellas se puede realizar una reconstrucción de la misma. Existen, en este caso, dos tipos de reconstrucción: densa y no densa.

Las reconstrucciones no densas reconstruyen puntos salientes de la escena (bordes, esquinas...) que son fáciles de emparejar entre imágenes. Su principal limitación es la falta de una reconstrucción completa.

Por el contrario la reconstrucción densa reconstruye todos los puntos de la escena. Esta variante tiene un tiempo de cómputo mucho más elevado y se podría utilizar en todas

aquellas aplicaciones en las que se necesita una reconstrucción realista, por ejemplo en navegación robótica.

Aplicaciones

Por citar algunos ejemplos de aplicaciones de la visión 3D, en la medicina se ha extendido el uso de la reconstrucción tridimensional. Así, la visión 3D es de mucha utilidad en cirugía, análisis de tejidos, reconstrucción de vasos sanguíneos o visualización neuronal. Puede servir para visualizar, en el interior del cuerpo humano, modelos artificiales generados a partir de imágenes (Figura 1.2). Incluso permite situar cuerpos extraños en el interior del paciente.

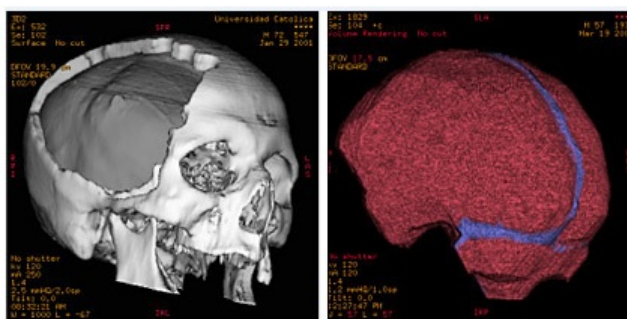


Figura 1.2: Ejemplo de aplicación de las técnicas de reconstrucción en el campo de la medicina.

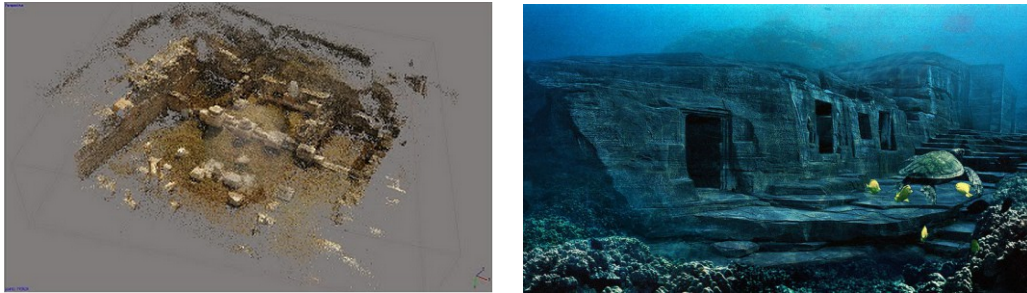
También se utilizan estas técnicas para reconstruir una escena y facilitar la navegación de un robot como el de la Figura 1.3 a través de ella, permitiendo localizar y reconocer objetos en la misma.



Figura 1.3: Ejemplo de robot móvil.

En el terreno de la cartografía y la topografía posibilita la creación de mapas tridimensionales e imágenes en 3D de un terreno, como se ve en la Figura 1.4. Incluso se han

desarrollado técnicas que permiten representar el relieve submarino.



(a) Reconstrucción 3D de un yacimiento arqueológico a partir de imágenes. (b) Reconstrucción de un relieve submarino.

Figura 1.4: Aplicación de las técnicas de reconstrucción 3D en topografía.

Una reconstrucción tridimensional también permite verificar, los procesos de un control de calidad, así como los objetos que se están fabricando.



Figura 1.5: Reconstrucción densa de una escena real en la que se ha introducido un coche virtual.

Por último, la realidad virtual emplea imágenes de una escena para introducir al usuario dentro de la misma. Éste es un campo muy amplio que abarca disciplinas como la arquitectura o la industria aero espacial. En la Figura 1.5 se puede ver cómo, a partir de la reconstrucción densa de una escena, se ha podido introducir en ella un coche virtual .

1.1. Objetivos

El objetivo principal de este proyecto consiste en el estudio de un método de reconstrucción 3D denso de escenas a partir de imágenes y su precisión. Para ello evaluaremos la influencia de tres parámetros en la precisión de dicho método.

- Uno de los parámetros que se considera es el tamaño del identificador de cada píxel, valorando como la variación de su tamaño influye en la exactitud con la que se

reconstruye la nube de puntos que da lugar a la reconstrucción 3D de la escena. El identificador es un descriptor del color de los píxeles que permitirá comparar unos con otros.

- Los identificadores son vectores, y se deberá encontrar la forma de compararlos. Por eso se evalúan diferentes métodos para la métrica del error, con el objeto de seleccionar, entre ellos, el más eficiente. Se estudiarán en concreto, la norma, la media y la mediana
- Otro de los objetivos del proyecto es la evaluación del efecto que produce el regularizador en las reconstrucciones. El regularizador es el encargado de suavizar la imagen, y eliminar el ruido que aparece en las reconstrucciones. Para ello se evaluará como afecta la variación de el parámetro más relevante que interviene en la regularización.

Capítulo 2

Dense Tracking and Mapping

DTAM (Dense Tracking and Mapping)¹ es una técnica de reconstrucción 3D densa [3]. Permite realizar en tiempo real la reconstrucción de una escena en 3D y a su vez, conocer las posiciones de la cámara.

2.1. Reconstrucción densa y no densa

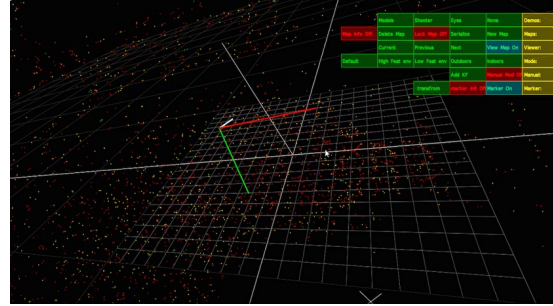
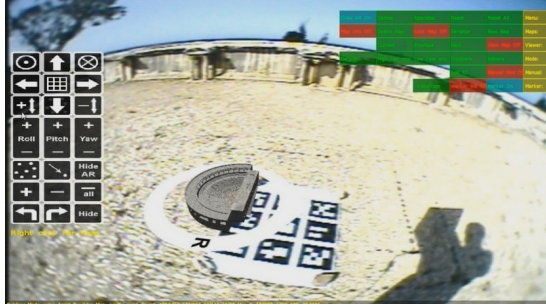
Este proyecto centra su atención en la reconstrucción densa. Como ya se ha dicho, la diferencia principal entre la reconstrucción densa y la no densa es que, mientras la primera se realiza reconstruyendo todos los puntos de la imagen, la segunda se limita a utilizar únicamente puntos característicos. Estos puntos característicos se diferencian claramente de los que tienen a su alrededor, y, más específicamente, son los puntos que tienen un alto gradiente de color en dos direcciones. Por ejemplo, la esquina de un objeto.

La técnica PTAM (Parallel tracking and mapping), se basa en la reconstrucción de dichos puntos característicos, realizando de forma paralela el cálculo de la localización de la cámara y el mapeado de la escena.

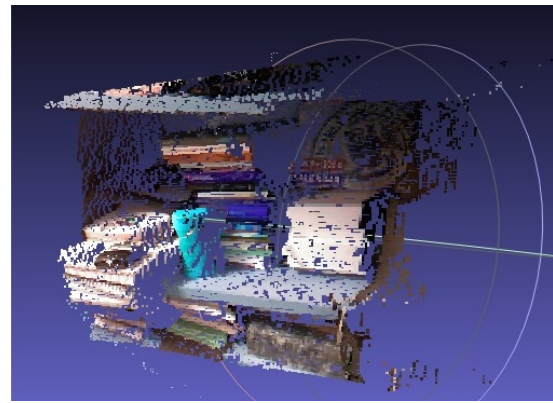
En la Figura 2.1 se observa la diferencia entre una escena reconstruida con una técnica tradicional, a partir de puntos característicos, y otra reconstruida con un método denso. En la Figura 2.1(b) aparecen señalados los puntos característicos reconstruidos, y a su lado, en la 2.1(a) se ve la escena real. Se observa que solo han sido reconstruidos unos cientos de puntos que para determinadas aplicaciones son insuficientes. Por otro lado, en la Figura 2.1(d) aparecen los puntos reconstruidos con una reconstrucción densa. En la Figura 2.1(c), correspondiente a la escena real, se observa que, efectivamente, se han

¹Seguimiento de una cámara y mapeado denso de la escena. Se utilizará este término en inglés ya que es el que aparece en la literatura técnica. De manera análoga se utilizará la notación en inglés a lo largo del TFG cuando su uso esté extendido.

reconstruido todos los puntos de la imagen.



(a) Escena real reconstruida utilizando una técnica basada en puntos característicos en la Figura 2.1(b). Imagen extraída de [5].
 (b) Puntos característicos reconstruidos con la técnica PTAM. Imagen extraída de [5].



(c) Escena real de la reconstrucción que aparece en la Figura 2.1(d).
 (d) Puntos de una escena reconstruidos con la técnica DTAM.

Figura 2.1: Comparación entre la cantidad de puntos utilizados en las técnicas PTAM y DTAM.

2.2. Dense Mapping

Para estimar el mapa de una escena es preciso disponer de una secuencia de imágenes de la misma, tomadas desde distintos puntos de vista. Partiendo de unas condiciones de brillo y luz constantes en todas ellas, se puede estimar la profundidad inversa de todos los píxeles de la imagen de referencia de la secuencia. Para ello se minimiza la función de energía global:

$$E_{\xi} = \int (g(\mathbf{u}) \|\nabla \xi(\mathbf{u})\|_{\epsilon} + \lambda \mathbf{C}(\mathbf{u}, \xi(\mathbf{u}))) \partial \mathbf{u} \quad (2.1)$$

Está compuesta por dos términos que se suman. El primer término, $g(\mathbf{u}) \|\nabla \xi(\mathbf{u})\|_{\epsilon}$, corresponde al término regularizador y el segundo, $\mathbf{C}(\mathbf{u}, \xi(\mathbf{u}))$, es el término fotométrico.

El parámetro λ es el peso relativo entre los dos términos. Ambos se analizan y explican a continuación.

2.2.1. Término Fotométrico

El error fotométrico es la diferencia de color entre un píxel de la imagen de referencia \mathbf{I}_r y su reproyección en el resto de las imágenes de la secuencia. Cada píxel \mathbf{u} de la imagen de referencia se proyecta en cada una de las otras imágenes \mathbf{I}_m . El sumatorio de la diferencia de color de todas ellas corresponde al término $\mathbf{C}(\mathbf{u}, \xi)$:

$$\mathbf{C}(\mathbf{u}, \xi) = \frac{1}{|\mathbf{I}_s|} \sum_{m \in \mathbf{I}_s} \|\rho(\mathbf{I}_m, \mathbf{u}, \xi)\|_1 \quad (2.2)$$

$$\rho(\mathbf{I}_m, \mathbf{u}, \xi) = \mathbf{I}_r(\mathbf{u}) - \mathbf{I}_m(\mathbf{T}_{mr}(\xi, \mathbf{u})) \quad (2.3)$$

La reproyección de los píxeles de la imagen de referencia en el resto de imágenes se consigue de la siguiente manera.

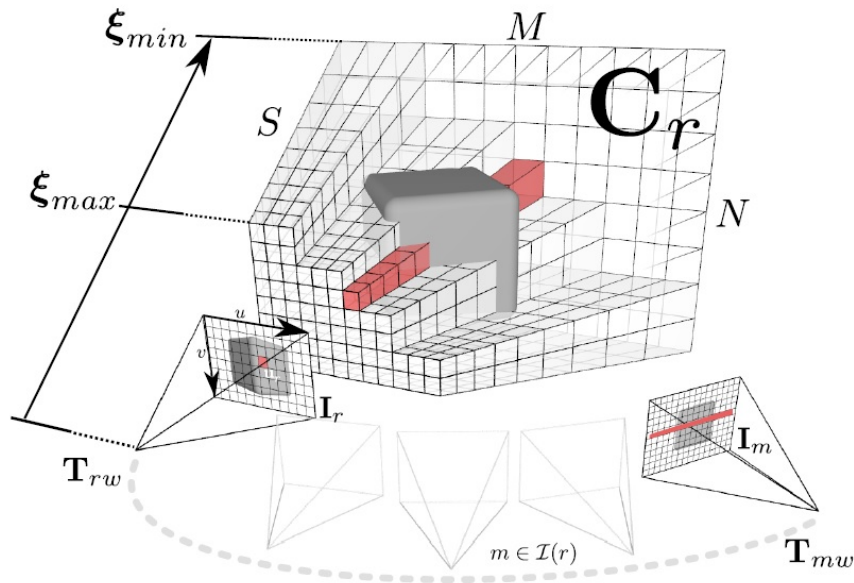


Figura 2.2: Cada píxel de la imagen de referencia I_r con posición T_{rw} , tiene punto asociado en el modelo 3D. Y desde el modelo 3D, cada píxel, se reproyecta en el resto de imágenes I_m . Imagen extraída de [3].

En primer lugar se obtiene un mapa 3D a partir de una serie de puntos característicos estimados previamente con los cuales se determina un valor máximo y uno mínimo para la profundidad (ξ_{min}, ξ_{max}). A continuación, se realiza un muestreo desde ξ_{min} hasta ξ_{max} y para cada píxel \mathbf{u} a cada profundidad ξ se calcula su posición 3D, \mathbf{X}_{cr} , en la referencia de la

cámara a partir de la posición 2D de cada píxel y los parámetros intrínsecos de la cámara. (Figura 2.2), en este caso 65 valores de profundidad comprendidos en ese intervalo, y se proyecta la imagen de referencia en el mapa 3D.

$$\mathbf{X}_{cr} = \frac{\mathbf{u} - (x_0 \ y_0 \ 1)^T}{f_r} \frac{1}{\xi} \quad (2.4)$$

Siendo f_r la distancia focal de la cámara de referencia y $(x_0 \ y_0 \ 1)^T$ el punto principal en coordenadas homogéneas. Los píxeles \mathbf{u} también se representan en coordenadas homogéneas.

Una vez proyectados los puntos en el mapa se realiza una transformación inversa en la que los puntos que están referenciados con respecto a la cámara \mathbf{c} pasan a estarlo con respecto al mundo \mathbf{w} :

$$\mathbf{X}_w = \mathbf{R}_r^{-1}(\mathbf{X}_{cr} - \mathbf{t}_r) \quad (2.5)$$

\mathbf{R}_r y \mathbf{t}_r corresponden a la matriz de rotación y traslación respectivamente, son los parámetros de movimiento que determinan la posición de la cámara de referencia. En relación con todas las cámaras \mathbf{m} , se establece a qué punto del mapa 3D corresponde cada píxel de las imágenes.

$$\mathbf{X}_{cm} = \mathbf{R}_m \mathbf{X}_w + \mathbf{t}_m \quad (2.6)$$

Por fin, se realiza la reproyección de los píxeles del mapa 3D al resto de las imágenes, y ello para todas las profundidades, tal como se puede observar en la Figura 2.3. La línea de puntos que se reproyecta es la epipolar, que se corresponde con el píxel de la imagen de referencia para todas las profundidades. Se calcula el error fotométrico entre cada uno de ellos. La profundidad óptima es la que obtiene menor error fotométrico.

2.2.2. Término Regularizador

El término regularizador de la Ecuación 2.1 está formado por la norma de Huber del gradiente del mapa de profundidades inversas $\|\nabla \mathbf{e}(\mathbf{u})\|_\epsilon$ y el peso de cada píxel $g(\mathbf{u})$:

$$g(\mathbf{u}) = e^{-\alpha \|\nabla \mathbf{I}_r(\mathbf{u})\|_2} \quad (2.7)$$

El peso de cada píxel $g(\mathbf{u})$ depende del gradiente de color del mismo $\|\nabla \mathbf{I}_r(\mathbf{u})\|$ en la imagen de referencia. En las zonas sin textura el gradiente de color es pequeño y por lo tanto el peso del píxel es más grande. Si el peso del píxel aumenta, el término regularizador de la Ecuación 2.1 aumenta también. Cuanto más aumenta el término regularizador más se suaviza la imagen. Lo contrario ocurre en las zonas con textura.



(a) Píxel de la imagen de referencia señalado con un punto rojo. (b) Reproyección de las profundidades estimadas para el píxel señalado en 2.3(a) en una de las imágenes de la secuencia.

Figura 2.3: En esta figura se muestra un ejemplo de como se realiza la reproyección de los píxeles de la imagen de referencia en una de las imágenes del resto de la secuencia. Para calcular el error fotométrico mínimo se compara el color del píxel 2.3(a) con todos los contenidos en la epipolar de 2.3(b).

En la Figura 2.4 se puede observar el efecto que produce el regularizador en los mapas de profundidad. Puede observarse que las discontinuidades de profundidad coinciden a menudo con bordes y contornos. Por esta razón se reduce el efecto del regularizador en las zonas de con grandes texturas y se incrementa en las zonas con poca textura.

La norma de Huber es una composición de dos funciones convexas[4]:

$$\|x\|_{\epsilon} = \begin{cases} \frac{\|x\|_2^2}{2\epsilon} & \text{if } \|x\|_2 \leq \epsilon \\ \|x\|_1 - \frac{\epsilon}{2} & \text{otherwise} \end{cases}$$

Como se puede observar en la Figura 2.4, el regularizador suaviza la imagen y elimina el ruido. Por lo general ésto provoca una mejora en las reconstrucciones, pero en algunos casos la regularización en exceso de una imagen provoca la pérdida de información necesaria para la reconstrucción.

2.2.3. Minimización de la energía

La ecuación de la energía está compuesta por el término convexo $g(\mathbf{u})\|\nabla\xi(\mathbf{u})\|_{\epsilon}$ y el término no convexo $\lambda\mathbf{C}(\mathbf{u}, \xi(\mathbf{u}))$, lo que hace en conjunto una función no convexa. Los dos términos están acoplados, lo que hace difícil su minimización [1][3][4]. Por ese motivo utilizamos una variable auxiliar \mathbf{a} , que permite hacerlos independientes el uno del otro:



(a) Mapa de profundidades antes de aplicar el regularizador. (b) Mapa de profundidades tras aplicar el efecto del regularizador. (c) Imagen real de los mapas de profundidad de las Figuras 2.4(a) y 2.4(b).

Figura 2.4: En estas imágenes se muestra el mapa de profundidad de una escena antes y después de aplicar un regularizador que suaviza la imagen.

$$E_{\xi, \mathbf{a}} = \int \left(g(\mathbf{u}) \|\nabla \xi(\mathbf{u})\|_{\epsilon} + \frac{1}{2\theta} (\xi(\mathbf{u}) - \mathbf{a}(\mathbf{u}))^2 + \lambda \mathbf{C}(\mathbf{u}, \mathbf{a}(\mathbf{u})) \right) \partial \mathbf{u} \quad (2.8)$$

El término $\frac{1}{2\theta} (\xi(\mathbf{u}) - \mathbf{a}(\mathbf{u}))^2 + \lambda \mathbf{C}(\mathbf{u}, \mathbf{a}(\mathbf{u}))$ no convexo se optimiza por muestreo y el término convexo $g(\mathbf{u}) \|\nabla \xi(\mathbf{u})\|_{\epsilon} + \frac{1}{2\theta} (\xi(\mathbf{u}) - \mathbf{a}(\mathbf{u}))^2$ se optimiza, de forma más eficiente, en el espacio primal-dual.

El término $\frac{1}{2\theta} (\mathbf{e}(\mathbf{u}) - \mathbf{a}(\mathbf{u}))^2$ hace que \mathbf{e} y \mathbf{a} sean iguales y θ tienda a 0 de forma iterativa.

Se transforma el término regularizador al espacio primal-dual, donde \mathbf{e} corresponde a la variable ξ , \mathbf{a} corresponde a α , \mathbf{q} es la variable dual y \mathbf{G} es el peso de los píxeles.

$$\|\mathbf{AGe}\| =_{\mathbf{q}, \|\mathbf{q}\|_2 \leq 1} \left\{ \langle \mathbf{AGe}, \mathbf{q} \rangle - \delta_{\mathbf{q}}(\mathbf{q}) - \frac{\epsilon}{2} \|\mathbf{q}\|_2^2 \right\} \quad (2.9)$$

Con esta transformación el problema de minimización se convierte en la maximización de este término:

$$\operatorname{argmax}_{\mathbf{q}, \|\mathbf{q}\|_2 \leq 1} \left\{ \operatorname{argmin}_{\mathbf{e}, \mathbf{a}} \mathbf{E}(\mathbf{e}, \mathbf{a}, \mathbf{q}) \right\} \quad (2.10)$$

Se reemplaza el término en la ecuación 3.3 obteniendo:

$$\mathbf{E}(\mathbf{e}, \mathbf{a}, \mathbf{q}) = \left\{ \langle \mathbf{AGe}, \mathbf{q} \rangle - \delta_{\mathbf{q}}(\mathbf{q}) - \frac{\epsilon}{2} \|\mathbf{q}\|_2^2 + \frac{1}{2\theta} (\mathbf{e}(\mathbf{u}) - \mathbf{a}(\mathbf{u}))^2 + \lambda \mathbf{C}(\mathbf{a}) \right\} \quad (2.11)$$

Para la variable \mathbf{q} , la energía se maximiza, por tanto utilizamos la técnica del gradiente ascendente:

$$\frac{\partial \mathbf{E}(\mathbf{e}, \mathbf{a}, \mathbf{q})}{\partial \mathbf{q}} = \mathbf{AGe} - \epsilon \mathbf{q} \quad (2.12)$$

En el caso de la variable \mathbf{e} , la energía se minimiza utilizando el teorema de la divergencia $\frac{\partial \langle A\mathbf{e}, \mathbf{q} \rangle}{\partial \mathbf{e}} = -\text{div}(\mathbf{q}) = \mathbf{A}^T \mathbf{q}$:

$$\frac{\partial \mathbf{E}(\mathbf{e}, \mathbf{a}, \mathbf{q})}{\partial \mathbf{e}} = \mathbf{G} \mathbf{A}^T \mathbf{q} + \frac{1}{\theta}(\mathbf{e} - \mathbf{a}) \quad (2.13)$$

Discretizando y reagrupando los términos se obtiene:

$$\frac{\mathbf{q}^{n+1} - \mathbf{q}^n}{\sigma_q} = \mathbf{A} \mathbf{G} \mathbf{e}^n - \epsilon \mathbf{q}^n \quad (2.14)$$

$$\frac{\mathbf{e}^{n+1} - \mathbf{e}^n}{\sigma_d} = -\mathbf{G} \mathbf{A}^T \mathbf{q}^{n+1} - \frac{1}{\theta^n}(\mathbf{e}^{n+1} - \mathbf{a}^n) \quad (2.15)$$

$$\mathbf{q}^{n+1} = \mathbf{max}(\mathbf{1}, (\mathbf{q}^n + \sigma_q \mathbf{A} \mathbf{G} \mathbf{e}^n) / (\mathbf{1} + \sigma_q \epsilon)) \quad (2.16)$$

$$\mathbf{e}^{n+1} = (\mathbf{e}^n + \sigma_d (\mathbf{G} \mathbf{A}^T \mathbf{q}^{n+1} + \frac{\mathbf{a}^n}{\theta^n})) / (\mathbf{1} + \frac{\sigma_d}{\theta^n}) \quad (2.17)$$

El término no convexo, $\frac{1}{2\theta}(\xi(\mathbf{u}) - \mathbf{a}(\mathbf{u}))^2 + \lambda \mathbf{C}(\mathbf{u}, \mathbf{a}(\mathbf{u}))$, se minimiza, haciendo una búsqueda punto por punto del valor más pequeño para cada \mathbf{a} del rango $\mathbf{a} = [\mathbf{e}_{\min}, \mathbf{e}_{\max}]$:

$${}_a \mathbf{E}^{\text{aux}}(\mathbf{e}, \mathbf{a}) \quad (2.18)$$

$$\mathbf{E}^{\text{aux}}(\mathbf{e}, \mathbf{a}) = \frac{1}{2\theta}(\mathbf{e}(\mathbf{u}) - \mathbf{a}(\mathbf{u}))^2 + \lambda \mathbf{C}(\mathbf{a}) \quad (2.19)$$

2.2.4. Comparación de identificadores

Se persigue el objetivo de minimizar al máximo el término del error fotométrico y así obtener la profundidad óptima; para ello, como se ha explicado en la sección 2.2.2, en lugar de comparar el color de un solo píxel de la imagen de referencia con su reproyección en el resto de imágenes, se introduce una modificación en la técnica DTAM y se utiliza también como identificador el color de los píxeles de alrededor.

A la hora de calcular el error fotométrico puede plantearse un problema cuando algunas regiones pertenecientes al mismo objeto tienen muy poca variación de color. En estos casos, la profundidad estimada no siempre es la óptima, y para evitarlo, en lugar de comparar píxel a píxel comparamos también los píxeles vecinos.

Se comparan los identificadores formados por varios píxeles. Lo natural para comparar dos vectores de elementos, como son en este caso los identificadores, es calcular la norma. En este caso se han utilizado tres maneras de comparar los identificadores, la norma del vector y además, el cálculo de la media y la obtención de la mediana.

Norma de un vector

La norma, longitud o módulo de un vector se representa a través de la ecuación:

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2 + \dots + v_n^2} \quad (2.20)$$

Media aritmética

La media es un valor que se obtiene al sumar todos los valores y dividirlos por el número de sumandos.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n a_i \quad (2.21)$$

Mediana

Para obtener el valor de la mediana, se ordenan todos los valores de forma ascendente y se selecciona el que se encuentra en la posición central.

Los resultados experimentales obtenidos al aplicar estas modificaciones a la técnica DTAM se recogen en el siguiente capítulo.

2.3. Dense tracking

La técnica del DTAM también realiza el tracking de las cámaras, aunque en este proyecto no se realizan estudios sobre esta parte.

El tracking consiste en la estimación de la posición de las cámaras a partir de un modelo 3D dado y la posición de las cámaras previas [3][4].

La estimación de la posición de una cámara se lleva a cabo a través de los parámetros de movimiento (rotación y traslación). Con estos parámetros se genera una imagen sintética a partir del modelo 3D.

La posición se estima en dos fases. En primer lugar se estima la rotación de la cámara a partir de posiciones anteriores. A continuación se estiman los 6GDL a partir del modelo 3D.

Ambas fases se consideran problemas de mínimos cuadrados no lineales, donde el error fotométrico se minimiza iterativamente.

Capítulo 3

Resultados experimentales

En este capítulo se recogen los resultados obtenidos en la evaluación del método utilizado para realizar las reconstrucciones 3D. Para evaluar la precisión de los algoritmos utilizados se comparan los valores de profundidad estimados con los reales, obtenidos con el sensor RGB-D Kinect. Este tipo de dispositivos que utilizan la técnica de luz estructurada, son capaces de capturar el mundo que los rodea en 3D, mediante la información obtenida a través de una cámara RGB y un sensor de profundidad. El resultado es una imagen RGB-D en la que a cada píxel se le asigna una información de color y de profundidad.

Para realizar estos experimentos se han utilizado tres secuencias de imágenes. En la Figura 3.1 se observa una muestra de las imágenes que componen estas secuencias.

3.1. Tamaño del identificador

El objeto de este experimento es evaluar cómo afecta la variación del tamaño del identificador en la minimización del error fotométrico, y por consiguiente en la estimación de la profundidad de cada píxel.

En el artículo [3], para estimar la profundidad óptima de cada píxel se calcula el error fotométrico mínimo; para ello se toma como identificador el color de un píxel y se compara con el color del píxel reproyectado en el resto de imágenes. La profundidad óptima es aquella para la que se tiene el error fotométrico mínimo. Sin embargo en ocasiones, esto puede resultar un problema, por ejemplo en secciones pertenecientes al mismo objeto en las que no hay variación de color. En estas ocasiones, el gradiente de color es similar para todas las profundidades y puede ocurrir que la profundidad estimada no sea la óptima. Por esta razón, además de tomar como identificador el color de un píxel, en este experimento se va a utilizar también el color de los píxeles vecinos. En la Figura 3.2 aparece un ejemplo



(a) Librería.

(b) Mochila.



(c) Estanteria.

Figura 3.1: Muestra de las imágenes que forman las secuencias utilizadas en los experimentos

de como se forman estos identificadores.

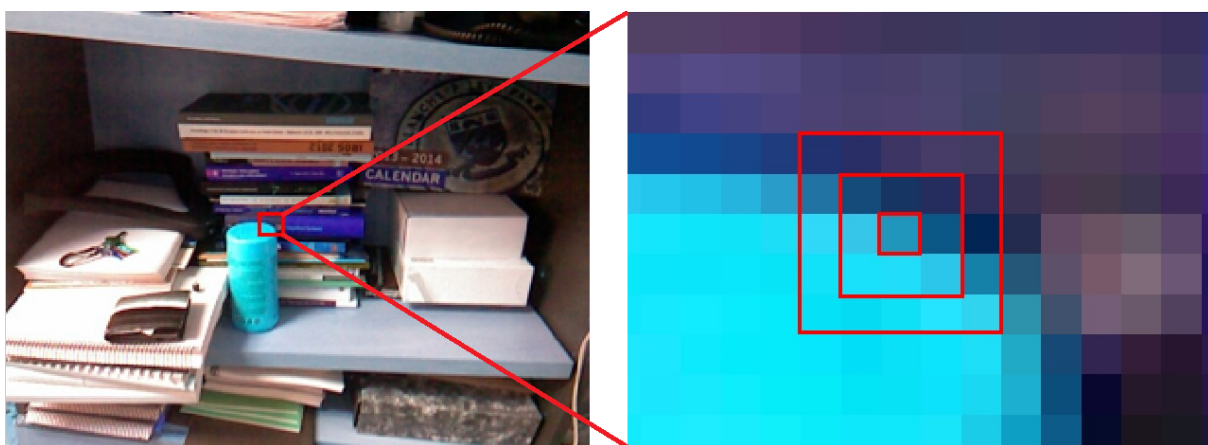


Figura 3.2: Sección ampliada de una imagen en la que se han seleccionado diferentes tamaños de identificador, 1x1, 3x3 y 5x5.

El hecho de comparar el color de varios pixeles presenta mejoras significativas en los resultados, aunque provoca también un aumento del coste computacional. En lugar de comparar el color de un solo píxel se compara el color de 9 (identificador 3x3) o 25 (identificador 5x5), cuestión a tener en cuenta al elegir el tamaño óptimo del mismo.

En primer lugar hay que redefinir la imagen obtenida con la Kinect para que tenga las mismas dimensiones, ya que se trabaja con imágenes de 240x320 y el sensor Kinect aporta imágenes de 480x640.

Las profundidades estimadas y las reales no están en la misma unidad, por lo que se debe calcular la escala para poder compararlas (Los resultados obtenidos se mostrarán en milímetros):

$$S = \frac{1}{i * j} \sum_{i=1}^n \sum_{j=1}^m \frac{e_{estimada}}{e_{kinect}} \quad (3.1)$$

Calculamos el error de la siguiente manera:

$$\mathbf{Error} = \frac{e_{estimada}}{S} - e_{kinect} \quad (3.2)$$

Todo este proceso se ha realizado para diferentes tamaños de identificador. Al aumentar el tamaño se observan mejoras en los mapas de profundidad. En la Figura 3.3 vemos dos mapas de profundidad calculados con distintos tamaños de identificador. Las zonas más claras se corresponden con las zonas más cercanas al observador y las más oscuras con las más lejanas. En los mapas de profundidad es difícil apreciar visualmente las diferencias, por ello también se estudian los resultados cuantitativos.

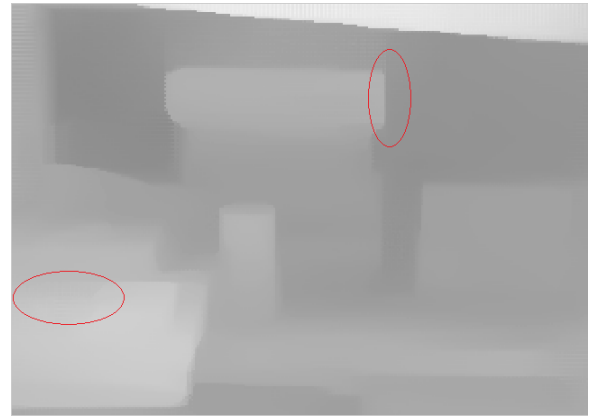
En la Figura 3.4 se observa el histograma de los errores obtenidos al comparar las profundidades estimadas y las obtenidas con el sensor Kinect de la secuencia *Librería*. En el primer histograma de la Figura 3.4(a) se observa que la mayor parte de los errores obtenidos tienen un valor entre 0 y 20 milímetros. El número de errores por encima de este valor va disminuyendo progresivamente. Nótese que en la siguiente Figura 3.4(b), los valores que se muestran son mucho más altos, pero también que el número de errores es muy pequeño.

Ambos histogramas han sido calculados a partir de los errores que se obtienen con un identificador de 11x11 y utilizando el método de la mediana, que como se muestra a continuación en el Cuadro 3.1 es el más eficiente.

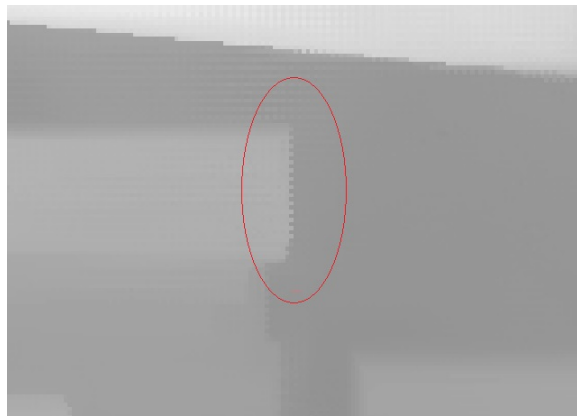
Los gráficos de cajas y bigotes permiten visualizar la distribución de un conjunto de datos de manera general. Estos diagramas están formados por una caja y dos brazos llamados “bigotes”. El lado inferior y el superior de la caja corresponden a los cuartiles Q1



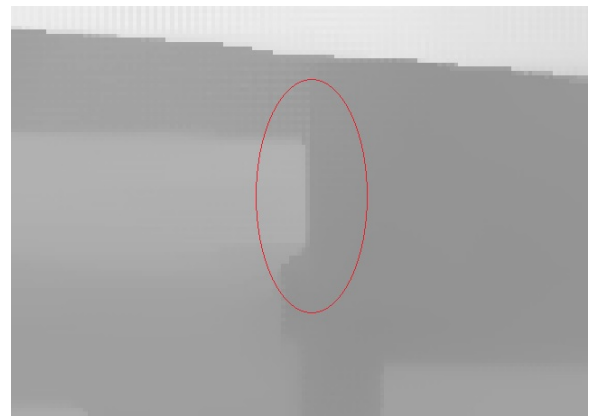
(a) Mapa de profundidad estimado tomando como tamaño de identificador 3x3.



(b) Mapa de profundidad estimado tomando como tamaño de identificador 11x11.



(c) Detalle perteneciente a la Figura 3.3(a).

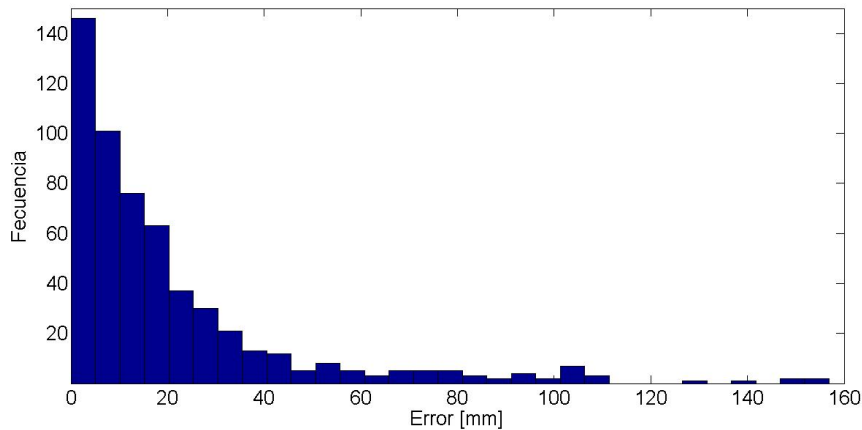


(d) Detalle perteneciente a la Figura 3.3(b).

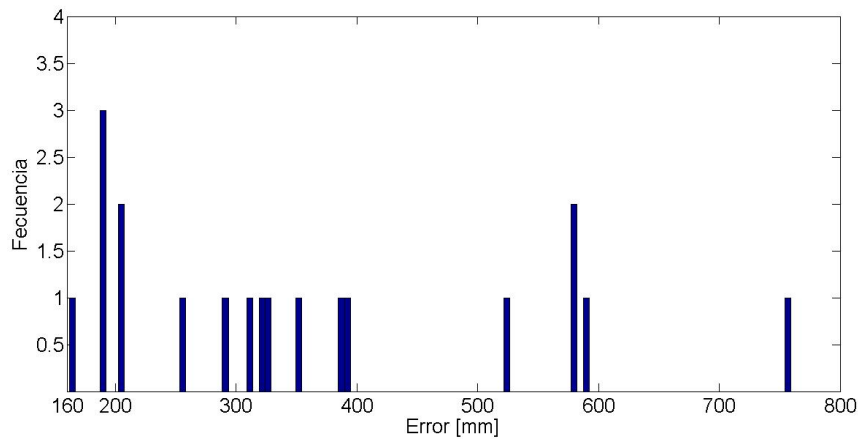
Figura 3.3: En ambas imágenes se muestran los mapas de profundidad obtenidos con distintos tamaños de identificador. En las zonas señaladas se observan algunas de las mejoras obtenidas al aplicar esta modificación.

y Q3, la línea que aparece dibujada en rojo en medio de la caja es la mediana y los brazos marcan el valor máximo y el mínimo del conjunto de datos. Por último el punto verde corresponde al valor de la media. En las Figuras 3.5, 3.6 y 3.7 aparecen los diagramas de cajas de las tres secuencias de imágenes utilizadas en este experimento.

Al analizar los gráficos de las tres secuencias aparecen varios elementos comunes. En primer lugar se observa de nuevo la mejora que se produce al introducir la modificación del tamaño del identificador con respecto a los valores del DTAM original. Los errores son más bajos. Nótese también que los valores máximos disminuyen conforme aumenta el tamaño del identificador, lo que demuestra que las profundidades estimadas son cada vez más próximas a las reales. Por último, destacar que unos valores máximos tan altos con respecto a la mayoría hacen que la media sea un valor notablemente mayor que la



(a) El histograma muestra el número de errores obtenidos con valores de entre 0 y 160 milímetros. Se observa que la máxima concentración de errores se produce con un valor de entre 0 y 20 milímetros.



(b) En este histograma se muestra el número de errores con valores altos. Nótese que el número de errores altos es mucho menor que el de errores bajos.

Figura 3.4: En esta Figura aparecen los valores cuantitativos de los errores calculados entre las profundidades reales y las estimadas.

mediana.

En el Cuadro 3.1 aparece un resumen de los resultados. En las filas del mismo se varía el tamaño tomado como identificador y en las columnas se indica el método utilizado para comparar los identificadores. Nos adelantamos, de esta forma, a lo que aparece en la siguiente sección 3.2. Los resultados de este cuadro corresponden a la primera de las tres secuencias de imágenes, *Librería*, que se están evaluando en este proyecto.

Como se puede observar, conforme se aumenta el tamaño del identificador, el error disminuye, y esto ocurre para todos los métodos. También se observa que, en general, cuando el identificador tiene un tamaño de 11x11 los valores se estabilizan. Nótese que la

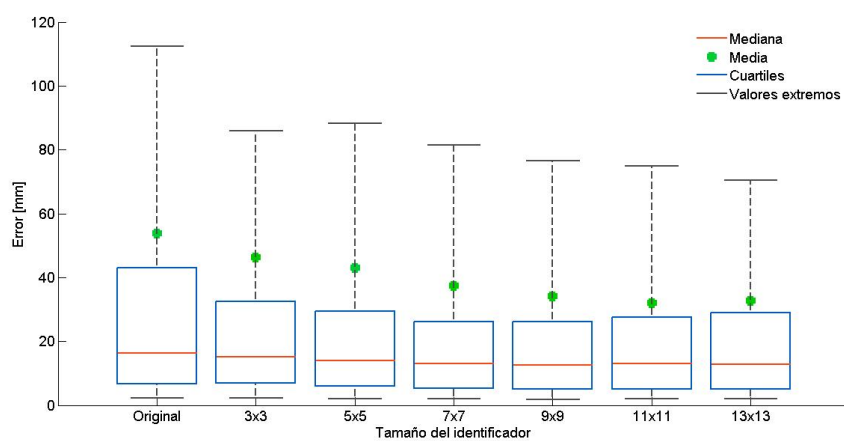


Figura 3.5: Gráfico de cajas de la secuencia *Librería*.

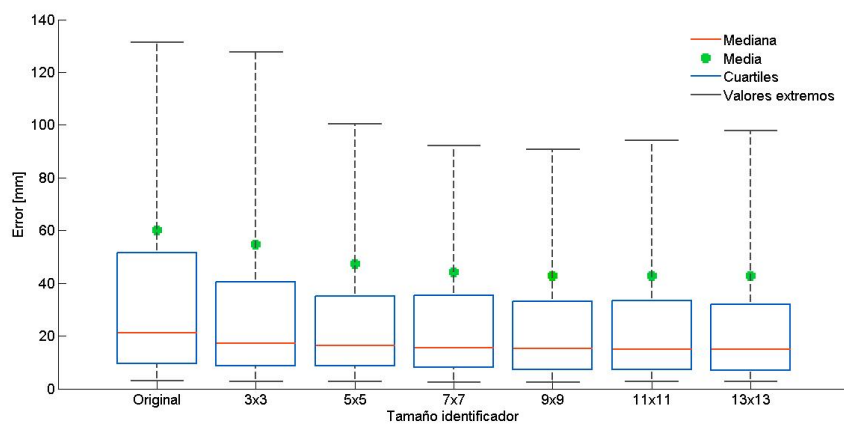


Figura 3.6: Gráfico de cajas de la secuencia *Mochila*.

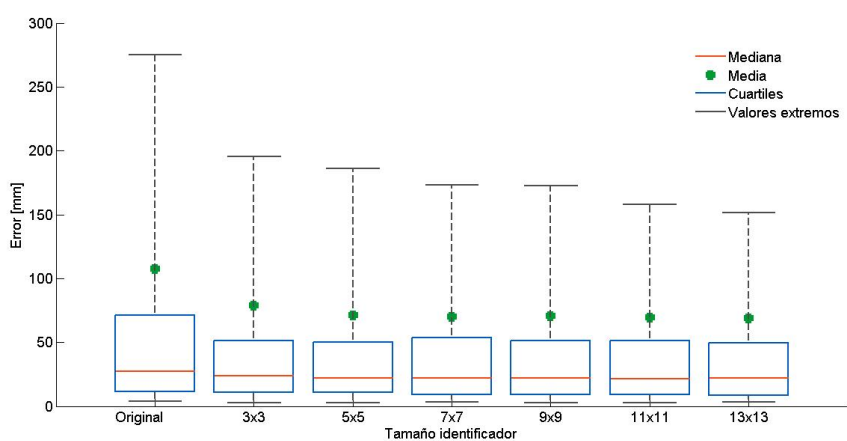


Figura 3.7: Gráfico de cajas de la secuencia *Estantería*.

Error [mm]			
	MEDIA	MEDIANA	NORMA
1x1	53,7	53,7	53,7
3x3	44,5	46,2	46,3
5x5	41,3	42,9	44,1
7x7	37,8	37,4	42,7
9x9	33,6	34,0	42,9
11x11	33,9	31.9	41,9
13x13	33,0	32,6	41,2

Cuadro 3.1: Cuadro resumen de los valores numéricos obtenidos al calcular el error entre las profundidades estimadas con el algoritmo y las obtenidas con el sensor Kinect. En el cuadro se muestran los valores medios.

diferencia obtenida entre este valor y el siguiente es mínima, y aunque en algunos casos el valor sea ligeramente más bajo, se toma como tamaño óptimo el de 11x11. Se ha escogido éste y no el siguiente, porque también hay que tener en cuenta el coste computacional. Si se escoge un tamaño más grande para el identificador también aumenta el coste y teniendo en cuenta que la diferencia es mínima se elige el de menor coste.

Se parte del DTAM original que utiliza el color de un solo píxel para calcular el error fotométrico. La primera fila del Cuadro 3.1 se corresponde con este valor. Con ésto queda descrita la mejora que existe al incluir más píxeles en el identificador. El experimento se ha realizado para tres secuencias de imágenes distintas y en todas ellas se ha obtenido una mejora significativa al modificar la técnica DTAM original aumentando el tamaño del identificador. En la Figura 3.8 se observan los resultados obtenidos.

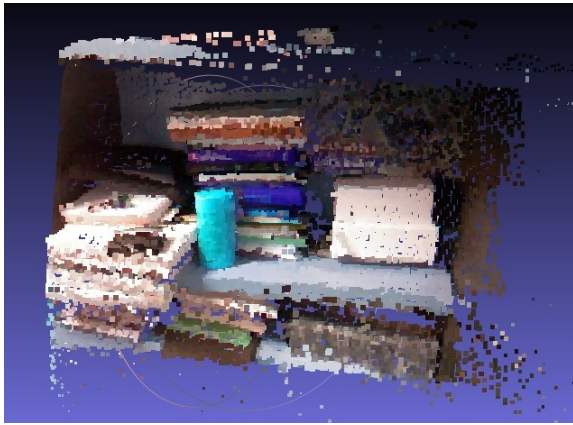
3.2. Comparación de identificadores

La finalidad de este experimento es evaluar la influencia de la métrica para evaluar el error fotométrico en la reconstrucción.

En la sección 2.2 se detalla la técnica DTAM de reconstrucción 3D y cómo se realiza el mapeado de una escena para obtener una reconstrucción, así como las modificaciones que se han realizado para intentar obtener resultados optimizados.

Una de ellas consiste en utilizar varios píxeles para calcular el error fotométrico y así estimar la profundidad óptima. Por ejemplo en el caso de un identificador 3x3, se obtienen 9 valores que forman un vector y los cuales hay que comparar con el vector de sus reproyecciones en el resto de las imágenes.

Se han aplicado distintas distribuciones estadísticas (véase sección 2.2.4) para comparar los identificadores. En el Cuadro 3.1 aparecen todos los valores obtenidos al aplicar



(a) Identificador de tamaño 1x1



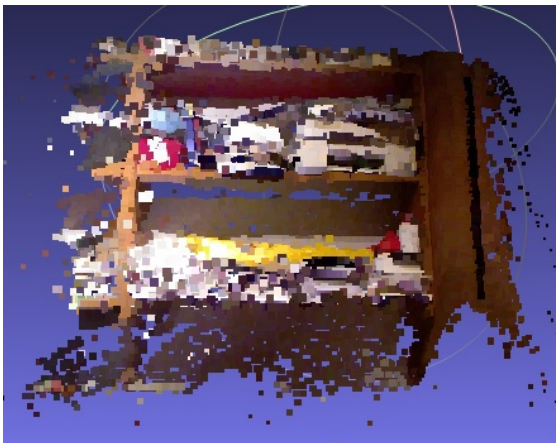
(b) Identificador de tamaño 11x11, utilizando el método de la mediana



(c) Identificador de tamaño 1x1



(d) Identificador de tamaño 11x11, utilizando el método de la mediana



(e) Identificador de tamaño 1x1



(f) Identificador de tamaño 11x11, utilizando el método de la mediana

Figura 3.8: En esta figura se observa una reconstrucción con el método original del DTAM y otra con las mejoras propuestas en este TFG. Nótese visualmente la mejor calidad de la reconstrucción. Cuantitativamente, la mejora del error medio es del 40%.

las diferentes distribuciones, así como los distintos tamaños del identificador.

Lo habitual para hacer la comparación entre dos vectores es calcular la norma euclídea, que coincide con el módulo del vector (Ecuación 2.20).

Lo que ocurre al calcular la norma es que si hay valores muy altos, el resultado será muy alto, y por tanto también el error. Por esta razón se han utilizado además, otras distribuciones para comparar los identificadores. Nótese en el Cuadro 3.1 que se obtienen valores más bajos calculando el error aplicando la media o la mediana.

Aplicando la media se obtienen, en conjunto, los valores más pequeños. Una de las limitaciones de la media aritmética es que sigue siendo algo sensible a los valores extremos. Los valores muy grandes tienden a aumentarla, mientras que los pequeños la reducen.

Por otro lado, la mediana es un método que resta influencia a los valores extremos, por lo que el valor obtenido representa mejor el conjunto de datos. Volviendo al Cuadro 3.1 se observa que el error mínimo se obtiene al aplicar esta métrica.

3.3. Influencia del regularizador

En esta sección se evalúa el efecto del regularizador en las reconstrucciones realizadas a partir de diferentes secuencias de imágenes.

En el artículo [3], así como en la Sección 2.2 de este texto, se detalla el término regularizador de la ecuación de la energía global (Ec.2.1). El regularizador suaviza la imagen, eliminando el ruido que aparece en las reconstrucciones. Véase la ecuación 2.1, cuando $\mathbf{g}(\mathbf{u})$ aumenta, la imagen se regulariza, esto ocurre en las zonas con poca textura (Ec. 2.7). Lo contrario ocurre en las zonas con textura, como son bordes o esquinas, $\mathbf{g}(\mathbf{u})$ tiende a cero y la imagen no se regulariza. En las zonas pertenecientes al mismo objeto, es decir, zonas sin textura, el regularizador consigue la profundidad estimada se ajuste más a la real.

Por lo general el efecto del regularizador supone una mejora en las reconstrucciones. Para demostrarlo se ha calculado el error entre las profundidades estimadas y las obtenidas con el sensor Kinect antes y después de aplicar el regularizador.

En primer lugar se muestra en la Figura 3.9 el efecto del regularizador en un mapa de profundidad. En la imagen de la izquierda aparece el mapa de profundidad antes de aplicar el regularizador, y se ve que la reconstrucción es ruidosa. En la siguiente imagen se ve el mapa de profundidades después de aplicar el regularizador, y se comprueba que ha desaparecido el ruido y se distinguen a la perfección los contornos de los objetos.

En el Cuadro 3.2 aparecen los valores del error de profundidad antes y después de aplicar el regularizador, para las tres secuencias de imágenes utilizadas.



Figura 3.9: Efecto del regularizador en un mapa de profundidades. En la imagen de la izquierda aparece el mapa de profundidades antes de que se haya aplicado el regularizador. En la imagen del centro se observa el efecto del regularizador. Por último la imagen de la derecha se corresponde con la imagen real de la escena.

Error[mm]						
	Librería		Mochila		Estantería	
	Sin Reg	Con Reg	Sin Reg	Con Reg	Sin Reg	Con Reg
3x3	46,2	28,15	54,7	41,1	78,6	70,4
5x5	42,9	27,14	47,3	40,8	70,9	81,4
7x7	37,4	27,4	44,1	41,3	70,1	74,8
9x9	34,0	28,07	42,8	40,7	70,5	74,8
11x11	31,9	28,92	42,6	41,1	69,7	70,2
13x13	32,6	29,54	42,6	40,7	68,8	70,7

Cuadro 3.2: Valores del error de profundidad obtenido antes y después de aplicar el regularizador.

Como ya se adelantó en la sección 2.2, regularizar en exceso el mapa de profundidades puede provocar resultados negativos en las reconstrucciones. De las tres secuencias utilizadas en este experimento, hay en una en la que el error aumenta tras aplicar el regularizador (Cuadro 3.2, secuencia *Estantería*). Esta secuencia se corresponde con las imágenes de la Figura 3.10. En ella se puede ver que el regularizador ha actuado en exceso, ya que prácticamente no se distinguen los contornos de ningún objeto.

Volviendo de nuevo a la ecuación de la energía global, se centrará la atención ahora en el parámetro λ :

$$E_{\mathbf{e}} = \int (g(\mathbf{u}) \|\nabla \mathbf{e}(\mathbf{u})\|_{\epsilon} + \lambda \mathbf{C}(\mathbf{u}, \mathbf{e}(\mathbf{u}))) \partial \mathbf{u} \quad (3.3)$$

El parámetro λ es una constante independiente que multiplica al término fotométrico de manera que, si se aumenta su valor, se disminuye la influencia del término regularizador, consiguiendo menos suavidad en la reconstrucción. Lo contrario ocurre cuando el valor de λ es pequeño; en esta ocasión el término regularizador aumenta su influencia.

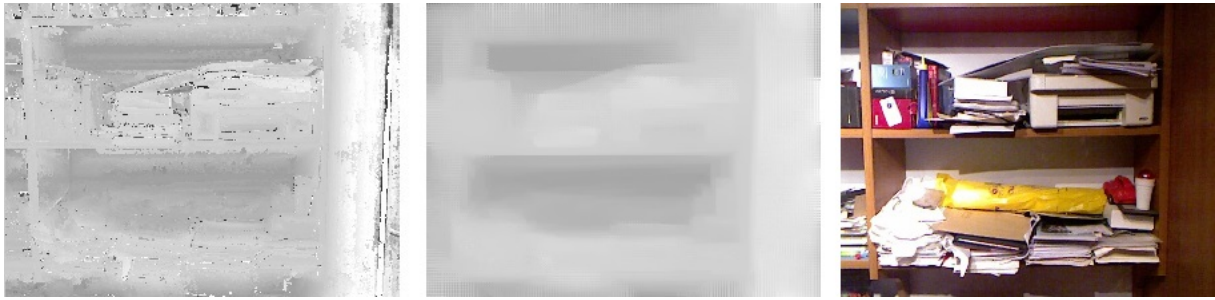


Figura 3.10: En estas imágenes se aprecia el efecto del regularizador, así como la imagen real de la escena. En este caso al aplicar el regularizador la reconstrucción es peor.

Para observar cómo afecta el regularizador a las reconstrucciones, se ha realizado un experimento variando el valor del parámetro λ . Si el valor de λ es muy pequeño la imagen se regulariza demasiado. Sucede que, aunque el peso del píxel sea muy pequeño, en el caso de bordes y esquinas, al reducirse tanto el término del error fotométrico, el regularizador toma relevancia en la ecuación. Un ejemplo del resultado que se obtiene en estas circunstancias se muestra en la Figura 3.11. En ella se puede observar la evolución del mapa de profundidades. Por el contrario, si el valor de λ es muy grande, la imagen no se regulariza. El término fotométrico se hace muy grande en la ecuación eliminando la influencia del término regularizador. En este caso, el mapa de profundidades no varía apenas antes y después de ser regularizado. En la Figura 3.12 se puede ver un ejemplo de ello.

En la sección anterior se ha mostrado que una de las secuencias utilizadas para el experimento empeoraba su reconstrucción tras aplicarle el regularizador. El objetivo es mostrar que variando este parámetro y ajustándolo para cada secuencia de imágenes, es posible hacer que el regularizador mejore la reconstrucción. En un primer momento se han realizado las reconstrucciones con valor estándar del parámetro λ .

En el Cuadro 3.2 de la sección anterior se observa que el error obtenido con el regularizador para la secuencia *Estantería* es mayor que el obtenido sin él. Para esta secuencia de imágenes se ha variado el parámetro λ en la Ecuación 3.3.

En el Cuadro 3.3 aparecen los valores del parámetro λ y el error obtenido con cada uno de ellos. Se puede observar que, conforme aumenta el valor de λ , el error disminuye hasta obtener un mínimo a partir del cual el error vuelve a aumentar aunque en menor medida. En este caso aumentando λ un orden de magnitud ya se obtiene mejora en el resultado. Ésto demuestra que suavizar demasiado empeora la reconstrucción.

En las Figuras 3.13, 3.14 y 3.15 aparece representado el valor del error con respecto al parámetro λ de las tres secuencias *Estantería*, *Mochila* y *Librería* respectivamente. El valor original de λ con el que se han realizado el resto de experimentos es $2,4 * 10^{-4}$.

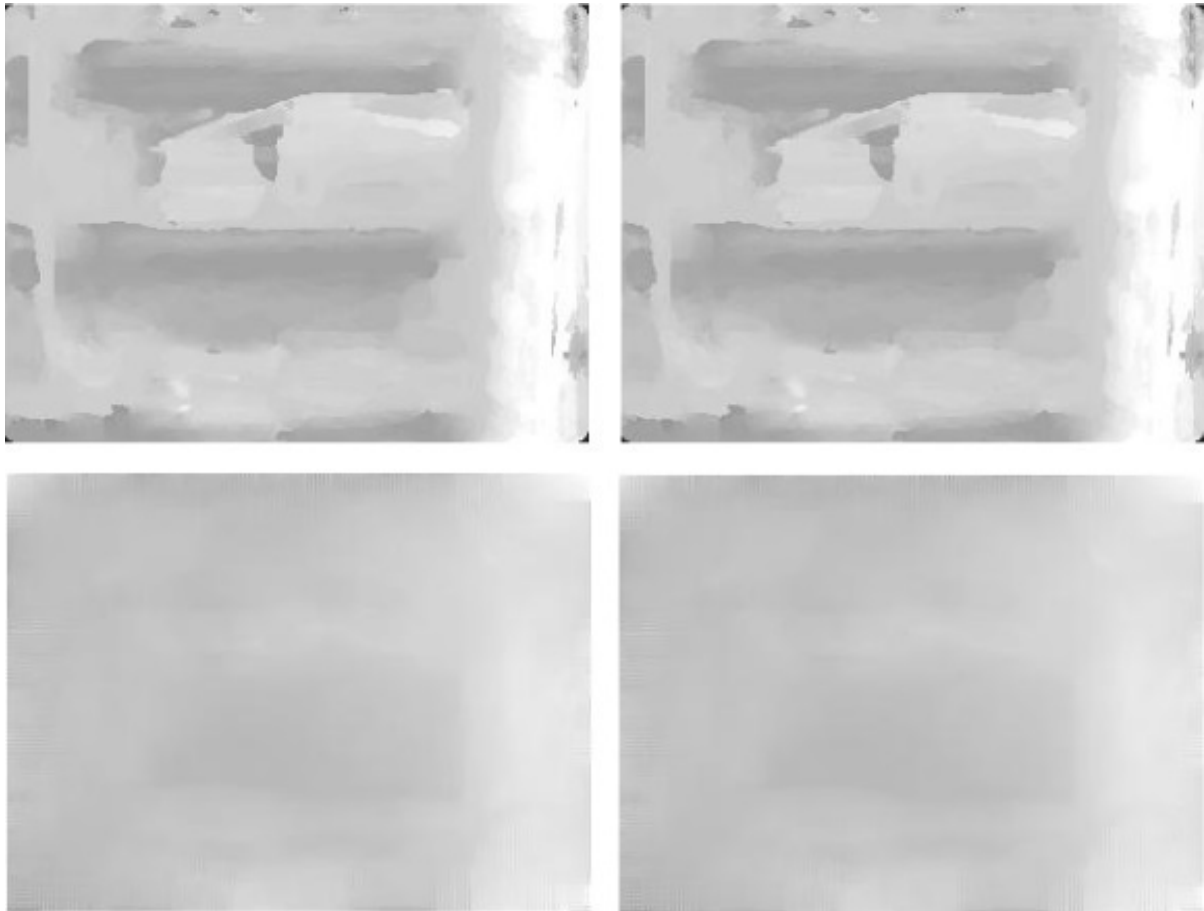


Figura 3.11: Esta figura se corresponde con el mapa de profundidades de la secuencia *Estantería*. En ella se muestra la evolución del mapa de profundidad conforme a la aplicación del regularizador. En la primera de las imágenes se ve el mapa inicial calculado a partir del término fotométrico, en la cual se observa ruido. La última de las imágenes es el mapa de profundidades una vez aplicado el regularizador, como se puede observar no hay ruido en la imagen y los bordes y contornos también han sido regularizados.

En las tres secuencias ocurre que para valores muy bajos de λ , el regularizador actúa en exceso y los errores son muy altos. Conforme el valor de λ aumenta el error disminuye. La línea roja que aparece en los gráficos corresponde al valor del error que se obtiene si no se aplica el regularizador en la reconstrucción. Se ve que, mientras los valores de λ son muy pequeños el regularizador no mejora la reconstrucción. El error alcanza un mínimo para un valor de λ de 0,0024.

Por el contrario si el valor de λ es muy alto llega un momento que tampoco aporta mejoras en la reconstrucción.

En todas secuencias evaluadas se ha obtenido el mismo valor óptimo de λ , ésto se debe a que las tres son muy similares. Puede ocurrir que para otras secuencias con distintas características el valor óptimo de λ sea otro.

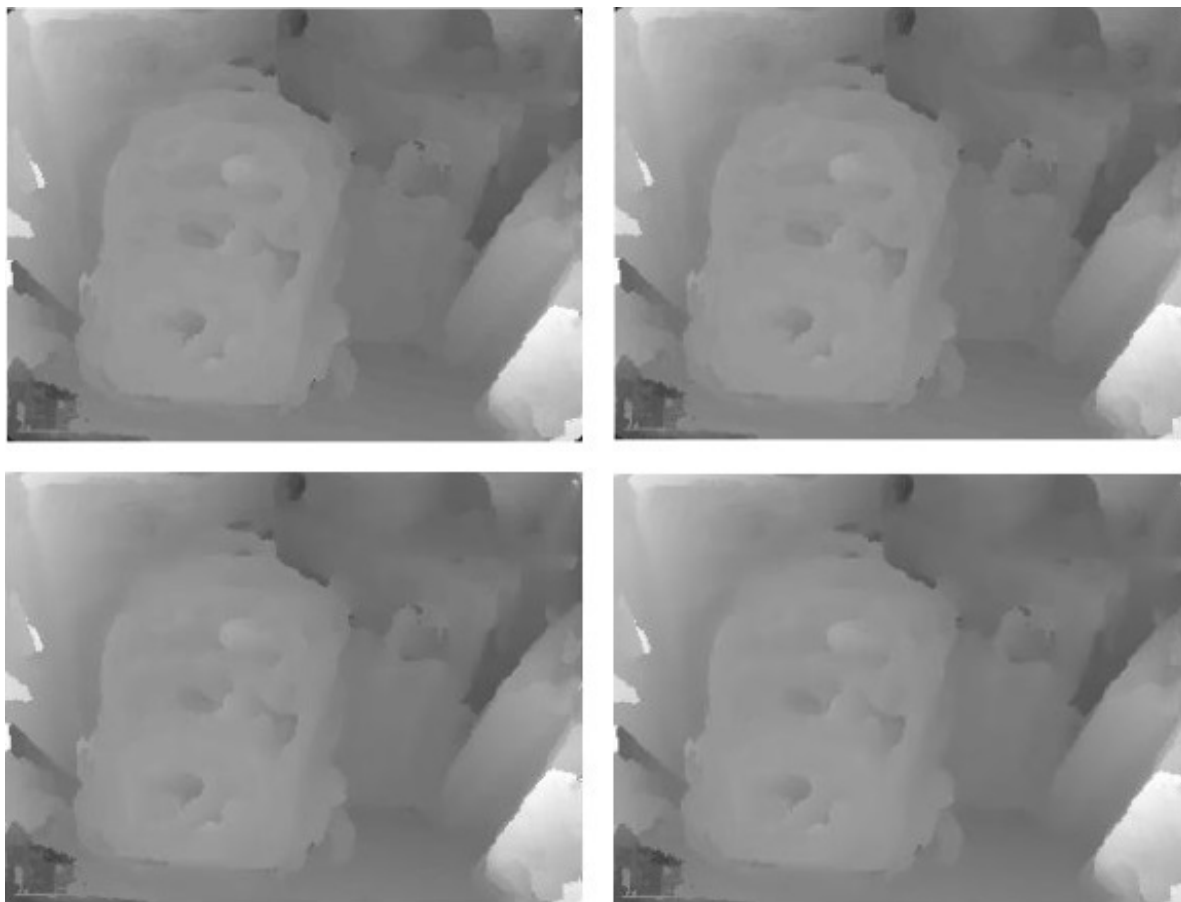


Figura 3.12: Esta figura se corresponde con la secuencia *Mochila*. En ella se muestra la evolución del mapa de profundidades al aplicar el regularizador. Se puede observar que durante la evolución el mapa de profundidad no varía, esto es debido a que el parámetro λ es muy alto y por lo tanto la imagen no se regulariza.

λ	Error [mm]
$2,4 * 10^{-7}$	122,59
$2,4 * 10^{-6}$	121,83
$2,4 * 10^{-6}$	113,56
$2,4 * 10^{-4}$	70,24
0,0024	68,57
0,024	68,71
0,24	69,47
2,4	69,26

Cuadro 3.3: Valores correspondientes al error obtenido al variar el valor del parámetro λ . El valor resaltado en negrita es el valor más bajo. El valor estandar con el que se han realizado las reconstrucciones es de $2,4 * 10^{-4}$. Con valores más pequeños de λ el error aumenta, lo contrario ocurre con valores más grandes.

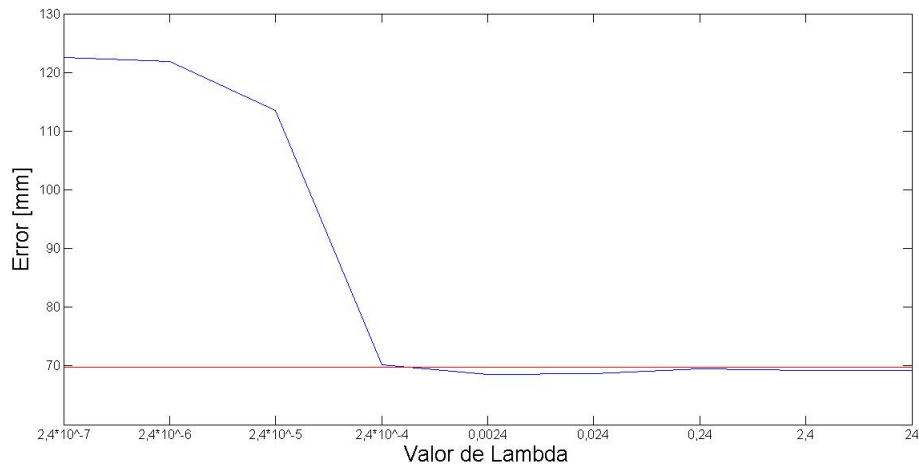


Figura 3.13: Esta gráfica representa los errores obtenidos al variar el parámetro λ con respecto al valor utilizado originalmente, en la secuencia *Estantería*.

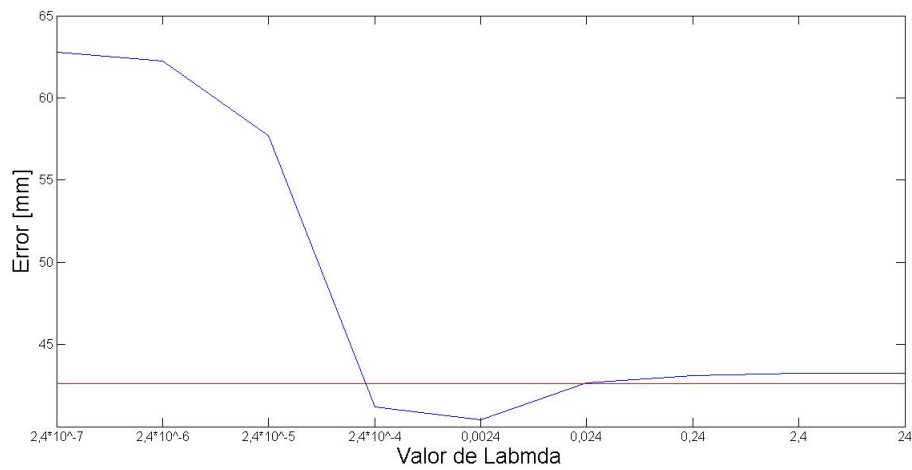


Figura 3.14: Esta gráfica representa los errores obtenidos al variar el parámetro λ con respecto al valor utilizado originalmente, en la secuencia *Mochila*.

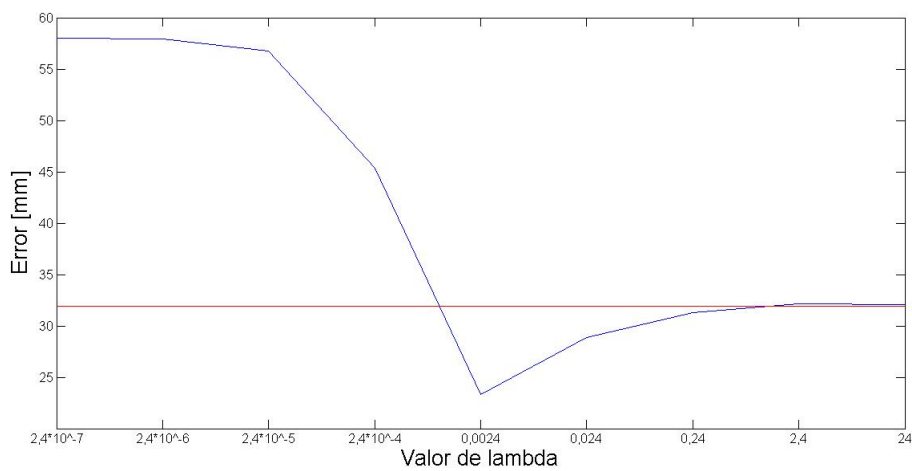


Figura 3.15: Esta gráfica representa los errores obtenidos al variar el parámetro λ con respecto al valor utilizado originalmente, en la secuencia *Librería*.

Capítulo 4

Conclusiones

El objetivo principal de este proyecto ha sido el estudio de un método de reconstrucción 3D denso de escenas a partir de imágenes y como la modificación de algunos de sus parámetros influye en la precisión.

El proyecto se ha centrado en la reconstrucción del mapa de la escena. Para ello, se ha minimizado la ecuación de la energía global que tiene dos sumandos, el término fotométrico y el término regularizador. Ambos se han analizado por separado, y modificando algunos de sus parámetros se han obtenido mejoras en los resultados.

En lo relativo al término fotométrico se puede concluir que:

- Se ha aumentado el tamaño del identificador que se utiliza para calcular el error fotométrico y así estimar la profundidad. El tamaño óptimo del identificador es 11x11. En cuanto a la métrica del error, lo habitual es calcular la norma del vector y en este caso se han obtenido la media y la mediana de los valores. El método más preciso es el cálculo de la mediana. Aplicando estas modificaciones la mejora es del 40 % en error mediano, 38,5 % en error medio y 31 % en error máximo.

En cuanto al término regularizador se ha evaluado la influencia que tiene en las reconstrucciones, y como afecta la variación del parámetro λ de la ecuación de la energía global. De los resultados se concluye que:

- El regularizador actúa de manera diferente para cada secuencia de imágenes. En algunas secuencias puede provocar resultados negativos si regulariza en exceso la imagen. Para corregir esto, se ha estudiado el parámetro λ que es el peso relativo de los términos de la ecuación de la energía global, tras lo cual se puede concluir que para cada secuencia de imágenes hay un λ óptimo diferente.

Cuando el regularizador afecta positivamente en la reconstrucción, los resultados mejoran notablemente, por ello sería interesante ser capaces de predecir el valor de λ óptimo

para cada secuencia de imágenes. De esta manera el regularizador siempre mejoraría la reconstrucción.

Bibliografía

- [1] A. Concha Belenguer. Using Superpixels in Monocular SLAM, EINA. 2013.
- [2] Ana Belén Moreno José Francisco Vélez y Ángel Sánchez. *Visión por computador*. 2003.
- [3] Richard A. Newcombe, Steven J. Lovegrove, y Andrew J. Davison. Dtam: Dense tracking and mapping in real-time. *Department of Computing, Imperial College London, UK*, 2011.
- [4] Alain Pagani. 2014. Lecture 3D Computer Vision, German Research Center for Artificial Intelligence.
- [5] G.Klein R.O.Castle y D.W.Muray. Video-rate localization in multiple maps for wearable augmented reality. *Proc 12th IEEE International Symposium on Wearable Computers, Pittsburg PA, Sept 28- Oct 1*, 2008.
- [6] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2010.
- [7] K. Vilá Ubieta. Reconstrucción 3d de modelos utilizando técnicas de visión artificial, ICAI. 2009.