

**METODOLOGÍA PARA LA ESTIMACIÓN DE MAPAS DE  
PROFUNDIDAD EN IMÁGENES *LIGHT FIELD*  
MEDIANTE EL USO DE TÉCNICAS DE SEGMENTACIÓN**

JUAN DIEGO SANTAMARIA BAREÑO



UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS  
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE  
TELECOMUNICACIONES  
BUCARAMANGA  
2017

METODOLOGÍA PARA LA ESTIMACIÓN DE MAPAS DE  
PROFUNDIDAD EN IMÁGENES *LIGHT FIELD*  
MEDIANTE EL USO DE TÉCNICAS DE SEGMENTACIÓN

JUAN DIEGO SANTAMARIA BAREÑO

*Trabajo de grado para optar al título de  
Ingeniero Electrónico*

Director  
SAID DAVID PERTUZ ARROYO  
Ph.D Ciencias de la Computación

Codirectores:  
OSCAR MAURICIO REYES  
Dr. Ingeniería Electrónica  
MIGUEL ANGEL MARQUEZ  
Ingeniero de Sistemas

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS  
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE  
TELECOMUNICACIONES  
BUCARAMANGA

2017

*Dedicado a*

Primero que todo a Dios por los dones que me dio para lograr este objetivo.

A mi madre Martha Bareño de quien aprendí que el trabajo duro tiene su recompensa. A mis padres José Álvarez y Leonel Santamaría que nunca dudaron de mis capacidades a pesar de las dificultades que llegue a presentar y me brindaron consejos que me ayudaron a crecer tanto personal como profesionalmente.

A mi nona querida Cecilia Duarte, por su apoyo incondicional a pesar de la distancia.

A Marylight Garnica, esa persona especial a la que doy gracias por aguantarme entre tanto momento de estrés que tuve y por ayudarme a calmarme para lograr ver la solución a varios problemas.

A cada uno de mis compañeros y amigos presentes durante mi carrera (la lista es larga pero ellos saben quienes son), gracias por los momentos de diversión fuera de clases y su apoyo durante mi etapa de aprendizaje.

*"Nunca consideres el estudio como una obligación, sino como una oportunidad para penetrar en el bello y maravilloso mundo del saber."*

---

Albert Einstein

**Juan Diego Santamaria Bareño**

## *Agradecimientos*

Principalmente a mi tutor de tesis el profesor Said Pertuz, un maestro en todo el sentido de la palabra. Gracias por su apoyo y confianza depositada en mí para realizar este proyecto y espero que podamos coincidir en futuros trabajos.

A Miguel Marquez y el profesor Oscar Reyes por su disposición de tiempo para guiarme con su codirección durante el proyecto y la redacción del mismo. Mis mas sinceros agradecimientos.

A las secretarias Ingrid y "Yolandita" por sus momentos de risa y su apoyo incondicional.

A todos mis familiares que me apoyaron tanto directa como indirectamente a realizar mis estudios.

A mis compañeros y amigos con los que pase tiempos agradable en la universidad.

Y por último pero no menos importante, a la Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones (E<sup>3</sup>T) por permitirme ser parte del proceso de formación profesional que ellos brindan.

¡ Muchas gracias !

# Índice general

<b>INTRODUCCIÓN</b>	<b>12</b>
<b>1 MARCO TEÓRICO</b>	<b>14</b>
1.1 FUNCIÓN PLENÓPTICA . . . . .	14
1.2 ESTIMACIÓN DE PROFUNDIDAD . . . . .	15
1.2.1 Estimación de profundidad en imágenes 2D . . . . .	15
1.2.2 Estimación de profundidad con imágenes de campo de luz . . . . .	16
1.3 SEGMENTACIÓN . . . . .	20
<b>2 METODOLOGÍA</b>	<b>23</b>
2.1 DECODIFICACIÓN DE LAS IMÁGENES DE CAMPO DE LUZ . . . . .	24
2.2 SEGMENTACIÓN DE LAS IMÁGENES DE CAMPO DE LUZ . . . . .	24
2.3 ESTIMACIÓN DE PROFUNDIDAD DE LAS IMÁGENES DE CAMPO DE LUZ . . . . .	25
2.4 FILTRADO MAPA DE PROFUNDIDAD . . . . .	25
<b>3 RESULTADOS EXPERIMENTALES Y ANÁLISIS</b>	<b>29</b>
3.1 BASES DE DATOS . . . . .	29
3.2 CÁMARA LYTRO . . . . .	30
3.3 ANÁLISIS ALGORITMOS DE ESTIMACIÓN DE PROFUNDIDAD . . . . .	31
3.4 ANÁLISIS DE ALGORITMOS DE SEGMENTACIÓN PARA IMÁGENES DE CAMPO DE LUZ . . . . .	33
3.5 RESULTADOS DE LAS PRUEBAS . . . . .	34
3.5.1 Resultados en imágenes de las bases de datos . . . . .	36
3.5.2 Resultados imágenes cámara Lytro . . . . .	38
3.6 ANÁLISIS CUALITATIVO . . . . .	38
<b>4 CONCLUSIONES</b>	<b>43</b>



## Índice de figuras

Figura 1	Visión del ojo humano . . . . .	14
Figura 2	Parametrización 4D light field . . . . .	17
Figura 3	Mosaico de imagen de campo de luz . . . . .	17
Figura 4	Diagrama básico del funcionamiento de una cámara plenóptica .	18
Figura 5	Ejemplo de la captura de una imagen <i>light field</i> . . . . .	21
Figura 6	Proceso de segmentación . . . . .	22
Figura 7	Diagrama de flujo de la metodología propuesta . . . . .	23
Figura 8	Decodificación imagen en el sensor. . . . .	24
Figura 9	Proceso de filtrado del mapa de profundidad. . . . .	27
Figura 10	Muestra montaje . . . . .	31
Figura 11	Estimación de Profundidad Tradicional . . . . .	32
Figura 12	Métodos de Segmentación. . . . .	33
Figura 13	Resultados experimentales . . . . .	35
Figura 14	Imagen ejemplo de la base de datos . . . . .	36
Figura 15	Tiempos de estimación de profundidad . . . . .	37
Figura 16	Prueba vistas angulares . . . . .	38
Figura 17	Resultados metodología <i>light field gantry</i> . . . . .	39
Figura 18	Continuación resultados metodología <i>Light Field Gantry</i> . . . .	40
Figura 19	Comparación parámetro Q . . . . .	41
Figura 20	Resultados metodología cámara lytro . . . . .	41
Figura 21	Limitación en base al mapa de profundidad . . . . .	44
Figura 22	Limitación en sobre-segmentación . . . . .	44

## Índice de tablas

Tabla 1	Imágenes de campo de luz usadas de la Universidad de Stanford	30
Tabla 2	Imágenes de campo de luz usadas del MIT Media Lab . . . . .	30
Tabla 3	Parámetros Estimación de Profundidad . . . . .	32
Tabla 4	Parámetros Segmentación . . . . .	34
Tabla 5	Resultados de la encuesta . . . . .	42

# ***OBJETIVOS***

## **OBJETIVO GENERAL**

Implementar una metodología que permita mejorar la estimación de profundidad en imágenes *light field* incorporando técnicas de segmentación

## **OBJETIVOS ESPECÍFICOS**

- ❖ Revisar el estado del arte de algoritmos de segmentación y de algoritmos de estimación de profundidad en imágenes de *light field* tomando como criterio su impacto, uso en otras investigaciones y su disponibilidad.
- ❖ Realizar una caracterización de parámetros de los algoritmos de segmentación y estimación de profundidad seleccionados previamente basados en las recomendaciones del autor y un análisis propio.
- ❖ Desarrollar e implementar un algoritmo que permita la unificación de la información obtenida de la segmentación y la estimación de profundidad.
- ❖ Evaluar el desempeño de la metodología propuesta con respecto a los métodos de estimación de profundidad seleccionados.

# INTRODUCCIÓN

En visión artificial la estimación de profundidad comprende un conjunto de técnicas y algoritmos que permiten estimar la distancia entre un objeto y una cámara. La estimación de profundidad es un importante campo de investigación en varias aplicaciones tanto comerciales como científicas de la visión computacional, tales como, reconocimiento de objetos [2], detección de obstáculos [11], interpretación de escenas [23], entre otros.

En algunas aplicaciones de la visión computacional es importante obtener información detallada de la escena para generar una reconstrucción 3D de alta calidad. Sin embargo, la estimación de profundidad con imágenes convencionales 2D es muy limitada debido a que dichas imágenes corresponden a una proyección perspectiva de la escena sobre un plano. Esta proyección causa que toda la información relacionada a la profundidad se pierda [8]. Por lo tanto, trabajos recientes se han enfocado en la estimación de profundidad usando imágenes multi-dimensionales, tales como las imágenes de campos de luz, o también conocidas por su nombre en inglés como *Light Field* [3,5,9,18–21]. Las imágenes de campo de luz se basan en una parametrización de una función que describe el comportamiento de la luz en el espacio, la *función plenóptica* [12]. Formalmente, la función plenóptica se define como un campo 7-dimensional  $P(x, y, z, \theta, \phi, \lambda, t)$ , donde  $(x, y, z)$  representan las dimensiones espaciales,  $(\theta, \phi)$  representan las dimensiones angulares,  $\lambda$  es la longitud de onda y  $t$  es el tiempo. En secciones posteriores se detallará a fondo éste tipo de representación.

Los resultados de las investigaciones realizadas en estimación de profundidad mediante imágenes de campo de luz [3, 5, 9, 18–21] presentan resultados más efectivos en comparación a los métodos tradicionales con imágenes 2D. Sin embargo, estos métodos no toman en cuenta la información adicional de radiancia que se puede obtener de una imagen como los bordes, la textura y, los colores. Este proyecto se basa en la hipótesis

de usar la información adicional de los bordes de los objetos de la escena para obtener una estimación de profundidad más precisa. La herramienta con la cual se logra este objetivo es la segmentación de imágenes.

Con este proyecto, se pretende indagar en el campo de las imágenes multi-dimensionales como las imágenes de campo de luz, estudiar los conceptos básicos de estimación de profundidad y segmentación en dichas imágenes, y aplicarlos para brindar un pie de apoyo sólido para futuros trabajos de investigación que se lleven a cabo en dicho campo.

En este trabajo se da a conocer un marco teórico en el que se da una breve explicación de los conceptos básicos sobre imágenes de campo de luz, estimación de profundidad, y segmentación. Luego, se da a conocer la metodología propuesta en éste proyecto y sus resultados experimentales en las bases de datos utilizadas. Finalmente, se da una análisis cualitativo de los resultados con sus respectivas conclusiones, posibles extensiones y trabajos futuros que se pueden derivar de este proyecto.

# Capítulo 1

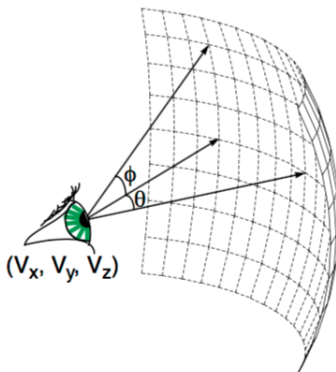
## MARCO TEÓRICO

### 1.1. FUNCIÓN PLENÓPTICA

Toda la información relacionada a lo que se puede observar en un entorno se concentra en rayos de luz de diferente intensidad. Si se pudiera ubicar una cámara de tamaño diminuto, tal que se ubique en un punto de intersección de dichos rayos de luz, se podría observar que esos rayos de luz forman una imagen. De forma similar funciona el ojo humano; tal que la estructura de un grupo de rayos es revelada en la localización de la pupila (figura. 1).

Por consiguiente, toda la distribución de la intensidad  $P$  atravesando un punto en específico se puede parametrizar en coordenadas espaciales  $P(x, y)$  ó, coordenadas esféricas  $P(\theta, \phi)$  para el caso de las imágenes bi-dimensionales. En el caso de fotografías a color o imágenes espectrales se considera la longitud de onda  $\lambda$  con lo cual, la parametrización resultante sería de la forma  $P(x, y, \lambda)$ . Un video a color agrega a la

Figura 1: Visión del ojo humano. Imagen tomada de [14]



parametrización el tiempo  $t$  obteniendo  $P(x, y, \lambda, t)$ . Una imagen de campo de luz tiene en cuenta las dimensiones tanto espaciales como angulares  $P(x, y, \theta, \phi)$ , entre otras parametrizaciones entre la cual, el ejemplo más completo que define a la función plenóptica como una función que describe los rayos de luz que componen una escena, es el video holográfico, el cual permite la reconstrucción desde cualquier punto de vista, en todo momento, posición y longitud de onda  $P(x, y, z, \theta, \phi, \lambda, t)$ .

## 1.2. ESTIMACIÓN DE PROFUNDIDAD

La profundidad de una escena puede ser definida como la distancia que hay desde un objeto a un punto de vista específico. En el caso de la visión computacional, la profundidad es la distancia entre el objeto y el sensor. La estimación de profundidad es una herramienta importante en aplicaciones como la reconstrucción 3D en la cual, como su nombre lo dice, se busca tener una reconstrucción 3D de una escena a partir de imágenes 2D. Dicha reconstrucción tiene sus ventajas y desventajas. Debido a esto varias investigaciones [1, 13, 17] se han realizado para implementar metodologías que presenten la mayor eficiencia en la estimación de profundidad como lo son la visión estéreo, sensores 2.5D, imágenes de campo de luz, entre otras.

**1.2.1 Estimación de profundidad en imágenes 2D** Una imagen 2D corresponde a un mapeo de la función plenóptica en el cual, solo se toman en cuenta las dimensiones espaciales  $(x, y)$  y el plano  $z$  se comprime y toma un valor constante. Este mapeo y compresión del plano  $z$  facilita el procesamiento de las imágenes 2D debido a que se tiene menor cantidad de información a procesar. Sin embargo, en aplicaciones en las cuales se requiere una visualización o reconstrucción 3D de una escena, las imágenes 2D presentan un inconveniente debido al proceso de captura de la imagen el cual se puede interpretar como una serie de proyecciones geométricas que limitan la cantidad de información disponible de la escena. Debido a esto, se han implementado técnicas de reconstrucción 3D entre las cuales se encuentran la visión estéreo [13] y la estimación con sensores activos [1] con el fin de optimizar el proceso. Estas técnicas en la actualidad son usadas en aplicaciones tales como, visión computacional, los video juegos, entre otras. A pesar de que estas técnicas son muy utilizadas, su principal inconveniente radica en que usualmente se requieren una mayor cantidad de capturas de la misma escena ó, sensores especiales los cuales pueden llegar a ser costosos. Para abordar éste problema,

recientemente han surgido dispositivos capaces de obtener mayor información de una escena como lo son las cámaras plenópticas las cuales capturan un tipo de imagen multi-dimensional conocida como *Light Field* [12].

### 1.2.2 Estimación de profundidad con imágenes de campo de luz

Las imágenes *light field* son una representación de la función plenóptica cuyo modelo fue propuesto por primera vez en 1996 por [12]. Este modelo se basa en tres hipótesis principales:

1. Al momento de capturar una imagen, la longitud de onda  $\lambda$  se puede simplificar en tres canales (R, G, B)

$$[P_c] = \int_{\lambda_2}^{\lambda_1} P(x, y, z, \theta, \phi, \lambda, t) d\lambda, \quad (1.1)$$

donde  $P_c$  denota el canal de color que se quiere obtener, para  $c \in \{R, G, B\}$  los límites  $\lambda_1$  y  $\lambda_2$  cambian según el canal que se desee obtener (R, G, o B) se pueden representar matemáticamente como la integración de la función plenóptica a través de un rango específico de la longitud de onda.

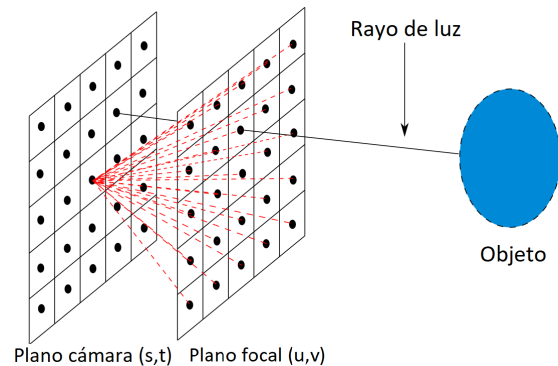
2. Considerando que el espacio que rodea la escena es aire, se asume que la radiancia a través de un rayo de luz permanece constante [12]. Basados en esta hipótesis, se puede reducir la función plenóptica en una dimensión. Esta simplificación de una dimensión permite reducir el tamaño final de la imagen a trabajar y facilita la reconstrucción de la función de radiancia de sus muestras [12].
3. Las escenas a capturar son imágenes, por lo tanto se puede despreciar el efecto del tiempo  $t$  en la función plenóptica.

Teniendo en cuenta las tres consideraciones anteriores, la función plenóptica se puede simplificar a una función 4D en la cual se define la imagen *light field* como:

$$LF = P(x, y, \theta, \phi). \quad (1.2)$$

El modelo de la imagen *light field* propuesto en [12] consta de dos planos paralelos llamados plano de la cámara y plano focal. Como se puede observar en la figura 2, los planos son discretizados para tener un número finito de rayos de luz, una vez hecho esto, todos los puntos discretizados en el plano focal se conectan a un punto discretizado del

Figura 2: Parametrización 4D light field.



plano imagen, obteniendo como resultado una imagen 4D en la que se encuentra tanto la información espacial como angular de la escena (figura.3).

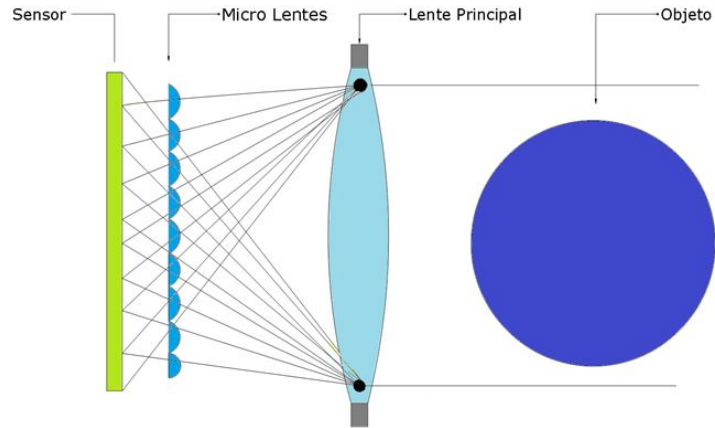
Figura 3: Resultado práctico de la parametrización. La imagen de la izquierda muestra la información angular ( $\theta, \phi$ ) de la escena, la imagen extraída a la derecha contiene la información espacial ( $x, y$ ) de la escena.



En las cámaras plenópticas, las imágenes *light field* se obtienen mediante un juego de lentes secundarios los cuales consisten en un arreglo matricial de micro lentes como se puede observar en la figura 4. Cada microlente corresponde a la proyección de un mismo punto del espacio desde diferentes puntos de vista. Un ejemplo de una imagen *light field* se puede observar en la figura 5.

En la literatura varios trabajos se han desarrollado para la estimación de profundi-

Figura 4: Diagrama básico del funcionamiento de una cámara plenóptica



dad usando diferentes técnicas con imágenes light field. Entre los trabajos relevantes de mayor impacto se encuentran aquellos en los que se plantea la idea de explotar las señales de correspondencia y desenfoco en la imagen EPI<sup>1</sup> y combinarlas con el fin de que se complementen una a otra [19]. El trabajo mencionado anteriormente se extendió en [20] con un nuevo algoritmo que combina el desenfoco, la correspondencia y el sombreado. Otro trabajo importante es el desarrollado por [21] que consiste en usar un modelado de oclusión para obtener información detallada de contornos de los objetos. Otros de los trabajos relevantes del estado del arte en esta área son los propuestos por [9] y [3]. Estos algoritmos trabajan con el método de *stereo vision* lo cual permite obtener una representación más eficiente de la profundidad aprovechando la información dada por éste tipo de geometría. La ventaja de estos algoritmos radica en que se necesita ajustar ciertos parámetros que dependen del tipo de escena y cámara utilizada y, de manera automática, realiza la estimación de profundidad basados en unos parámetros de costo los cuales definen una tolerancia para la cual determinar las mejores disparidades para obtener una eficiente estimación de profundidad.

Enfocando el análisis en los algoritmos de tipo *stereo vision* como lo son los propuestos por [3,9], se debe tener en cuenta que estos métodos trabajan bajo tres criterios que son: *determinar función de costo*, *agregación del costo* y *estimación de disparidad*

---

<sup>1</sup>Imagen Epipolar. Esta imagen se basa en la geometría epipolar la cual es una geometría de la visión estéreo (stereo vision). Esta geometría consiste en que dos cámaras observando una escena 3D desde dos posiciones distintas tienen una relación geométrica que queda totalmente establecida a partir de los parámetros intrínsecos (longitud focal y centro de proyección) y extrínsecos (orientación relativa) de las cámaras.

en los cuales la diferencia está principalmente en el cálculo de la función de costo que es de la forma  $C = \sum \Psi$ , donde  $\Psi$  es el método que cambia de acuerdo al método o variables utilizadas.

Para el caso del algoritmo de [3], el parámetro  $\Psi$  se puede expresar como:

$$\Psi = \sum_{i=1}^{i=c} \Gamma_{d_\theta}(E_\theta[L_c(x - \theta d_\theta, y, \theta, \phi)]I_x \dots - L_c(x - \theta d_\theta, y, \theta, \phi)), \quad (1.3)$$

donde se realiza una comparación entre cada pixel en la imagen *light field*  $L_c(x, y, \theta, \phi)$  en un índice de canales de color  $c^2$  con todos los pixeles agregando una disparidad a lo largo del eje angular horizontal  $\theta$  o vertical  $\phi^3$  definida como  $L_c(x - \theta d_\theta, y, \theta, \phi)$  para  $d_\theta = \{d_{\theta min} \dots d_{\theta max}\}$ , donde  $d_{\theta min}$  y  $d_{\theta max}$  son las disparidades máximas y mínimas esperadas. Estos valores dependen exclusivamente de las condiciones de captura de la escena como lo son la estructura interna de la cámara como su disposición de micro-lentes y características de la escena capturada.  $E_\theta$  es el estimador de varianza a utilizar entre una de las dos coordenadas angulares. Se utilizó la media y la mediana en el eje angular horizontal ya que presentó mejores resultados.  $I_x$  es una matriz identidad la cual se encarga de repetir todo el proceso de comparación con  $E_\theta[L_c(x - \theta d_\theta, y, \theta, \phi)]$  en todo el eje angular horizontal  $\theta$ .  $\Gamma_{d_\theta}$  es la función que determina el costo de un pixel para una disparidad dada. Para una mayor profundización sobre éste método, consultar la referencia [3].

Para el caso del algoritmo de [9], el parámetro  $\Psi$  hace uso de la transformada bi-dimensional de Fourier para realizar un desplazamiento preciso de los sub-pixeles de la imagen<sup>4</sup>  $I(x)$  de la forma:

$$I'(x) = I(x + \Delta x) = \mathcal{F}\{I(x)\}e^{2\pi i \Delta x}, \quad (1.4)$$

donde al aplicar el desplazamiento se calculan dos costos, la suma de las diferencias absolutas SAD y la suma de las diferencias de gradientes GRAD. Basados en la ecuación 1.5, la suma de las diferencias absolutas construye un costo de acople mediante la comparación de la imagen central de sub-apertura  $I(x, s_c)$  con las demás imágenes de

---

<sup>2</sup>Solo se toman en cuenta los espacios de color donde los canales de color son equivalentes entre sí como por ejemplo el canal RGB.

<sup>3</sup>Para éste caso, durante todo el trabajo se tomo el eje angular horizontal ya que presentó mejores resultados.

<sup>4</sup>Para esta parte, se denota  $x = (x, y)$  como las coordenadas cartesianas y  $s = (\theta, \phi)$  como las coordenadas angulares.

sub-apertura  $I(x, s)$  generando desde un punto de vista canónico un mapa de disparidad, donde  $V$  contiene las coordenadas angulares  $s$  del pixel excepto para la vista central  $s_c$  y  $R_x$  las coordenadas cartesianas  $x$  y  $\tau_1$  es un valor de truncado de la función.

$$C_A(x, l) = \sum_{s \in V} \sum_{x \in R_x} \min(|I(x, s_c) - I(x + \Delta x(s, l), s)|, \tau_1) \quad (1.5)$$

Ahora, teniendo en cuenta la ecuación 1.6, la suma de las diferencias de gradientes controla la importancia relativa de las diferencias de los dos gradientes direccionales basado en las coordenadas relativas  $s$  donde  $\tau_2$  es una constante de truncado de la función y  $\text{Diff}_x(x, s, s_c, l) = |I_x(x, s_c) - I_x(x + \delta x(s, l), s)|$  denota las diferencias entre el gradiente x-direccional de la imagen de sub-apertura. De igual manera se define  $\text{Diff}_y$  en los gradientes y-direccionales.

$$C_G(x, l) = \sum_{s \in V} \sum_{x \in R_x} \beta(s) \min(\text{Diff}_x(x, s, s_c, l), \tau_2) + (1 - \beta(s)) \min(\text{Diff}_y(x, s, s_c, l), \tau_2) \quad (1.6)$$

Para una mayor profundización sobre éste método, consultar la referencia [9].

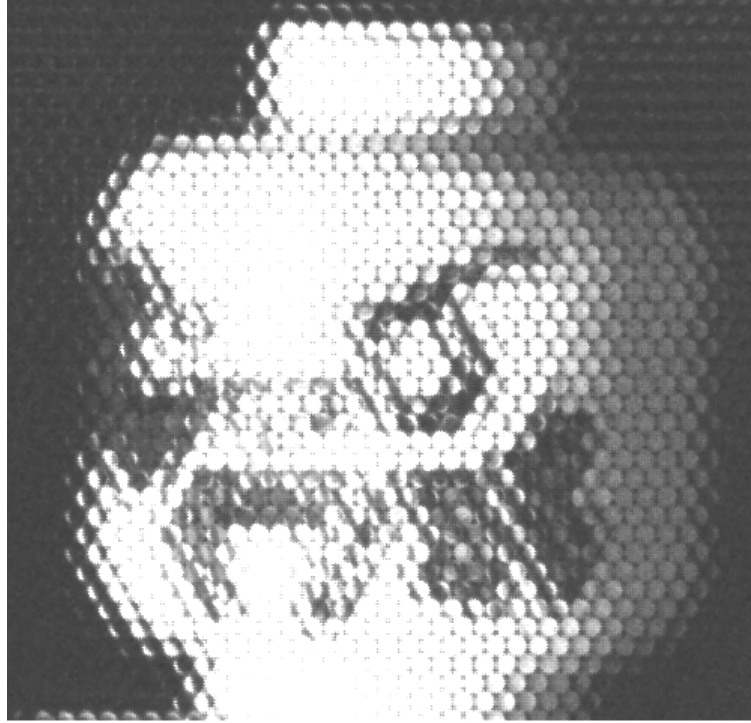
Actualmente los métodos de estimación de profundidad con imágenes *light field* presentan excelentes resultados comparados con su estimación con imágenes bi-dimensionales. Sin embargo, los métodos existentes de estimación no toman por completo las ventajas de la información de radiancia, como lo son los bordes, color, y textura.

### 1.3. SEGMENTACIÓN

En éste trabajo, se propone intergrar los métodos de estimación de profundidad en imágenes Light Field anteriormente descritos con técnicas de segmentación. Durante el proceso de segmentación, se toma una imagen  $\mathbf{I} \in \mathbb{R}^{m \times n}$ . Mediante un proceso de agrupamiento, la imagen es dividida en regiones  $\mathbf{R} = \{r_1, r_2, \dots, r_k\}$  donde  $k$  es el número de regiones o segmentos deseados. Dentro de los métodos de agrupamiento, se destaca el algoritmo de K-means. El principio fundamental de éste algoritmo se basa en dividir mediante iteraciones una imagen en  $k$  regiones. El proceso de iteración para el K-means es el siguiente:

1. Seleccionar el numero grupos K a realizar.

Figura 5: Recorte de una imagen *light field* capturada en el sensor. La cámara permite un tamaño de exportación a 2D de 4 MP. con una apertura de lentes constante de  $f/2.0$



2. Asignar a cada pixel de la imagen el grupo que tenga la menor distancia entre el pixel y el centro del grupo.
3. Re-calcular el centro del grupo realizando la media de todos los píxeles del grupo.
4. Repetir los pasos 2 y 3 hasta que haya convergencia.

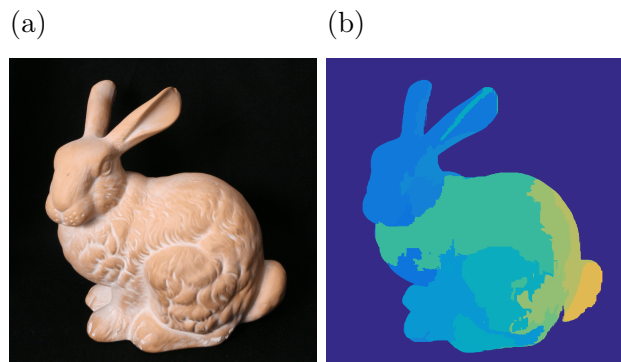
Una vez definido el proceso de agrupamiento, la segmentación de imágenes se puede expresar como una función de la forma:

$$\bar{\mathbf{I}} = \text{Seg}(\mathbf{I}, p_1, p_2, \dots, p_n), \quad (1.7)$$

donde  $\bar{\mathbf{I}} \in \mathbb{R}^{m \times n}$  es la imagen segmentada en  $\mathbf{R}$  regiones,  $\text{Seg}(\cdot)$  es el algoritmo de segmentación dependiendo del método a utilizar [7, 10, 15] y  $p_n$  son los parámetros necesarios para llevar a cabo la segmentación.

Al identificar zonas con tonalidades similares en sus píxeles, el algoritmo de segmentación se encarga de agruparlos en una cantidad  $k$  de regiones, para el ejemplo mostrado

Figura 6: Proceso de segmentación. (a) Imagen original. (b) Imagen segmentada utilizando [15].



en la figura 6, se utilizó un  $k = 32$ .

Actualmente la segmentación es una herramienta fundamental en la visión computacional donde su principal aplicación es la de reconocimiento de objetos mediante la detección de bordes. Mediante la detección de bordes se puede lograr un reconocimiento preciso sobre la geometría del objeto segmentado, gracias a éste reconocimiento preciso, se logra una clasificación de objetos efectiva lo cual es fundamental para la reconstrucción 3D.

## Capítulo 2

# METODOLOGÍA

Teniendo en cuenta la problemática que presenta la estimación de profundidad en imágenes de campo de luz que se presentó en la sección 1.2.2, en éste proyecto se propone una metodología para la estimación de profundidad que incluya técnicas de segmentación. La metodología propuesta para éste trabajo se puede resumir con el diagrama mostrado en la figura 7. En el desarrollo de la metodología se usa la información obtenida de la segmentación de la imagen de campo de luz y algoritmos tradicionales de estimación de profundidad.

Para llevar a cabo esta metodología, se tienen en cuenta cuatro etapas: decodificación de las imágenes de campo de luz, segmentación, estimación de de profundidad y filtrado del mapa de profundidad.

Figura 7: Diagrama de flujo de la metodología propuesta. La imagen de campo de luz se procesa en dos partes, primero se calcula su estimación de profundidad con métodos tradicionales. Por otra parte, se seleccionan los valores  $(x, y, \theta, \phi)$  tal que se obtenga la vista central para pasar a la segmentación de dicha vista. Finalmente se hace un filtrado del mapa de profundidad usando como guía la información de la segmentación para obtener un nuevo mapa de profundidad.

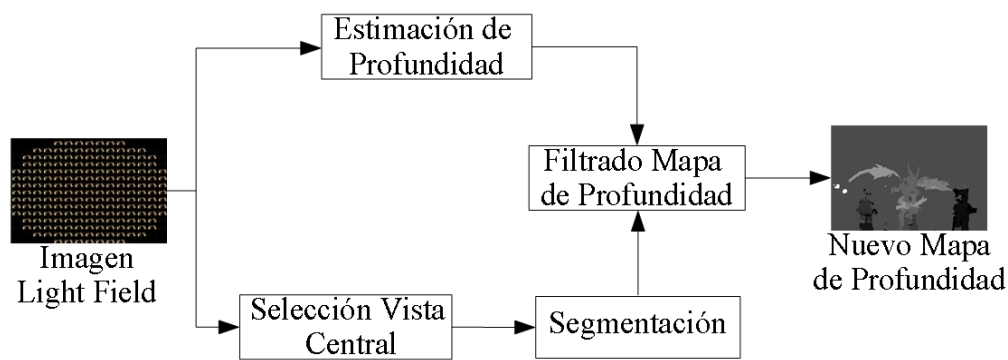
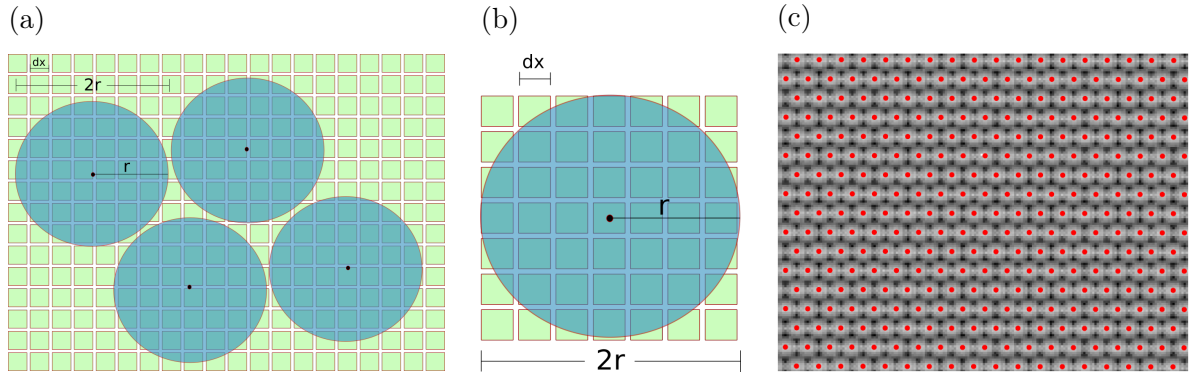


Figura 8: (a) Imagen 2D en el sensor. Durante la decodificación se obtiene la imagen en el sensor sin rectificar ya que los centros de los micro-lentes no coinciden con los centros de los píxeles de la imagen. Para corregir eso, en (b) se realizan transformaciones de rotación y escalado para que todos los centros de los micro-lentes coincidan con el centro de los píxeles de la imagen. Al final en (c) se puede observar la alineación de los lentes con el centro de los píxeles en donde  $dx$  es el tamaño del píxel y  $r$  es el radio de cada micro lente.



## 2.1. DECODIFICACIÓN DE LAS IMÁGENES DE CAMPO DE LUZ

El proceso de decodificación (figura 8) se puede dividir en dos pasos: un mapeo 2D a 4D y una rectificación del mapeo. En el primer paso, las proyecciones adquiridas por el sensor 2D son re-ordenadas con el fin de obtener la imagen de campo de luz 4D con distorsión. En el segundo paso, la imagen campo de luz distorsionada es corregida para compensar las no-linealidades del proceso de formación de la imagen. El proceso de decodificación se ilustra brevemente en la 8 y su resultado final se puede observar previamente en la figura 5. Para un análisis más profundo en la decodificación de las imágenes, en éste trabajo se utilizó el método de decodificación presentado en [6].

## 2.2. SEGMENTACIÓN DE LAS IMÁGENES DE CAMPO DE LUZ

En éste paso se busca obtener una buena segmentación teniendo como criterio la detección del contorno de los objetos en la escena. Tal como se explicó en la sección 1.3, definiendo la función de segmentación de la forma  $\bar{\mathbf{I}} = \text{Seg}(\mathbf{I}, p_1, p_2, \dots, p_n)$ , se obtiene

una segmentación de la imagen de campo de luz  $\bar{\mathbf{I}}$  con información sobre el contorno de los objetos. Sin embargo, al momento de aplicar los métodos de segmentación a la imagen de campo de luz, se debe tener en cuenta que una imagen de campo de luz pertenece a un espacio multi-dimensional ( $LF \in \mathbb{R}^{x,y,\theta,\phi}$ ). Por lo tanto, es necesario llevar a cabo una proyección a un espacio bi-dimensional para poder aplicar dichos métodos y obtener una segmentación  $\bar{\mathbf{I}} \in \mathbb{R}^{m \times n}$ . Para éste propósito, se codifica la imagen de campo de luz como un conjunto de imágenes 2D con diferentes vistas angulares [12]. Específicamente se opta por una imagen  $\mathbf{I} = \mathbf{LF}_{x,y}$  donde por convenciencia se seleccionan los parámetros  $(x, y, \theta, \phi)$  tal que sea la vista central de la imagen de campo de luz seleccionada. Hecho lo anterior, se utilizan métodos de segmentación clásicos de imágenes 2D. En éste trabajo, se realizó un estudio de diferentes algoritmos de segmentación en el estado del arte con el objetivo de seleccionar los algoritmos con mejor desempeño para la aplicación de interés (ver sección 1.3).

## 2.3. ESTIMACIÓN DE PROFUNDIDAD DE LAS IMÁGENES DE CAMPO DE LUZ

Para esta parte, tal como se detallo en la sección [1.2.2], se utilizaron algoritmos de estimación de profundidad de tipo *stereo vision* los cuales permiten obtener una representación más eficiente de la profundidad aprovechando la información dada por éste tipo de geometría. Para éste caso consideraron específicamente los algoritmos propuestos por [3, 9]. El funcionamiento de dichos algoritmos se presentó en la seccion [1.2.2], para mayor profundización en los métodos, consultar las referencias respectivas.

## 2.4. FILTRADO MAPA DE PROFUNDIDAD

Como último paso para obtener la nueva estimación de profundidad se realiza un filtrado del mapa de profundidad obtenido previamente. Este filtrado se realiza tomando como guía la información obtenida por la segmentación de la vista central de la imagen de campo de luz. Utilizando dicha guía se espera obtener una nueva estimación de profundidad en la cual se mantenga la información de profundidad con la mejora adicional de una mejor detección de los contornos en los objetos de la imagen capturada.

Para éste proceso, se define la nueva estimación de profundidad como una función

$f(\mathbf{D}, M_k)$ , donde  $f(\cdot)$  es la función que realiza el filtrado del mapa de profundidad obtenido utilizando como guía la segmentación obtenida de la vista central. Para llevar a cabo éste proceso, definimos un conjunto de regiones  $\mathbf{R} \in \mathbb{R}^{m \times n}$  que representan las regiones segmentadas de la imagen tal que,  $R \in \{r_1, r_2, \dots, r_k\}$ , donde  $r_k$  corresponde al  $k$ -ésimo objeto en la imagen segmentada y  $n$  es el número de objetos. A partir de éste conjunto de regiones, se define una máscara de segmentación  $M_k$  tal que:

$$\mathbf{M}_{k(m,n)} = \begin{cases} 1 & \text{si } \mathbf{I}_{(m,n)} \in r_k \\ 0 & \text{otros.} \end{cases} \quad (2.1)$$

Ahora, la función  $f(\mathbf{D}, M_k)$  se define como:

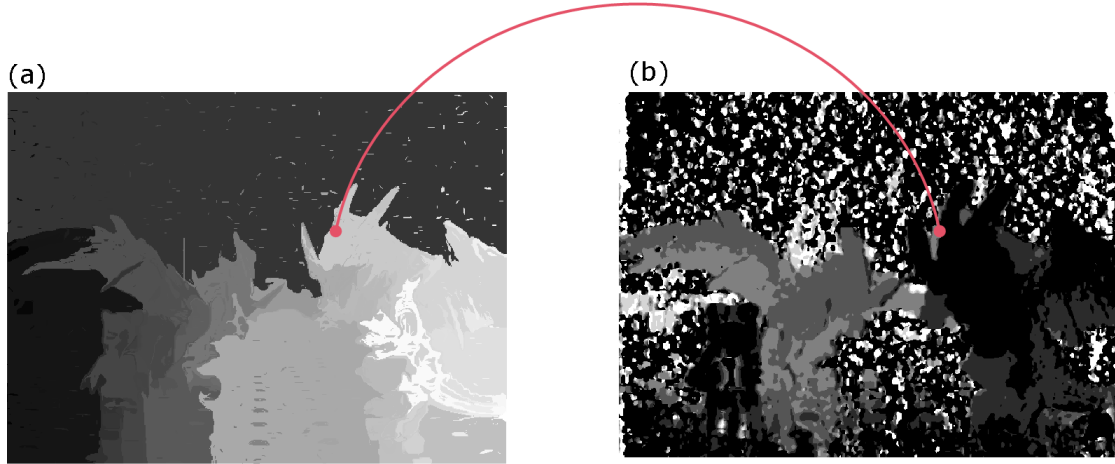
$$f(\mathbf{D}, M_k) = \frac{1}{\eta(r_k)} \left( \sum_{(m,n) \in r_k} \mathbf{D}_{(m,n)} M_{k(m,n)} \right), \quad (2.2)$$

donde  $\eta_k$  es el número de pixeles de la  $k$ -ésima región. Mediante éste proceso, se busca en cada pixel de la máscara de segmentación su correspondiente pixel en la estimación de profundidad de tal forma que se computa el valor medio de la información de profundidad en los pixeles de cada región  $r_k$  (figura 9). Por lo tanto, la ecuación (2.2) permite obtener la estimación final de profundidad mediante la superposición de cada región:

$$\bar{\mathbf{D}} = \sum_{k=1}^n f(\mathbf{D}, M_k). \quad (2.3)$$

La metodología propuesta se puede resumir mediante el pseudo-código 1. Éste algoritmo trabaja mediante el uso de la imagen de campo de luz llevando a cabo un proceso de codificación mediante los archivos de calibración *CFiles* que se encuentran en la cámara. Una vez decodificada la imagen, se procede con la estimación de profundidad  $\mathbf{D}$  obtenida con métodos tradicionales tales como los propuestos por [3, 9] y la segmentación de la vista central de la imagen obtenida previamente mediante los algoritmos de segmentación publicamente disponibles. Se crea una máscara de regiones de la imagen segmentada  $\bar{\mathbf{I}} \in \mathbb{R}^{m \times n}$ . Se define una variable  $n \leftarrow \eta(r_k)$  que toma el número de pixeles en cada  $k$ -ésima región. La nueva estimación de profundidad  $\bar{\mathbf{D}} \in \mathbb{R}^{m \times n}$  se inicializa en zeros. Luego mediante un ciclo *for* en el que se recorren todas las regiones, se buscan los pixeles de la imagen segmentada que pertenezcan a dicha región para ubicarlos en una

Figura 9: (a) Imagen segmentada con  $r_k$  regiones. Acorde a la ecuación (2.3) la función busca por la posición de un píxel  $(m, n)$  en (b) para computar su valor medio con el píxel correspondiente en una región  $r_k$  en (a).




---

**Algoritmo 1** Metodología Propuesta

---

- 1: *Entradas:*  $\mathbf{LF} \in \mathbb{R}^{x,y,\theta,\phi}$
  - 2: *Salida:*  $\bar{\mathbf{D}} \in \mathbb{R}^{m \times n}$
  - 3: *Inicio:*
  - 4:  $\mathbf{D} \leftarrow \text{Prof}(\mathbf{LF})$ ; donde  $\mathbf{D} \in \mathbb{R}^{m \times n}$  es la estimación de profundidad
  - 5:  $\bar{\mathbf{I}} \leftarrow \text{Seg}(\mathbf{LF}_{\text{cv}})$ ; donde  $\bar{\mathbf{I}} \in \mathbb{R}^{m \times n}$  es la imagen segmentada
  - 6:  $\mathbf{R} \leftarrow \text{regiones}(\bar{\mathbf{I}})$ ; donde  $R$  es un conjunto con las regiones de la segmentación
  - 7:  $n \leftarrow \eta(r_k)$ ; conjunto de regiones  $r_k \in \mathbf{R}$
  - 8:  $\bar{\mathbf{D}} \leftarrow \text{zeros}(\text{size}(\mathbf{D}))$ ; se inicia la nueva estimación de profundidad como una máscara vacía
  - 9: **for**  $k = 1 : n$  **do**
  - 10:      $M_k \leftarrow \bar{\mathbf{I}}_{(m,n)} \in r_k$ ; se genera una máscara para cada región en la segmentación
  - 11:      $\bar{\mathbf{D}}(M_k) \leftarrow f(\mathbf{D}, M_k)$ ; se realiza el proceso de filtrado del mapa de profundidad basados en la ecuación 2.2
  - 12: *Fin*
-

nueva máscara  $\mathbf{M}_k$  que posteriormente se agrega a la máscara  $\bar{\mathbf{D}}$  mediante la función  $f(\mathbf{D}, M_k)$ . La función  $f(\mathbf{D}, M_k)$  calcula el valor promedio de profundidad en una región para luego agregar dicha información a la máscara  $\bar{\mathbf{D}}$  creada previamente.

## *Capítulo 3*

# ***RESULTADOS EXPERIMENTALES Y ANÁLISIS***

En éste capítulo se detalla el análisis de los algoritmos de estimación de profundidad y segmentación tradicionales utilizados para llevar a cabo la prueba de la metodología, se detallará desde la captura de las imágenes de campo de luz, la selección de las bases de datos, selección y parámetros de los algoritmos y finalmente se darán los resultados obtenidos de la metodología tanto con imágenes de campo de luz de bases de datos como imágenes originales capturadas en laboratorio.

Como se mencionó anteriormente en la sección anterior de metodología, las imágenes de campo de luz usadas corresponden a dos tipos diferentes: imágenes de campo de luz obtenidas de una base de datos pública e imágenes propiamente capturadas.

### **3.1. BASES DE DATOS**

Como una primera prueba, se utilizaron imágenes de campo de luz obtenidas de una base de datos publica proporcionadas por la Universidad de Stanford<sup>5</sup> y del MIT Media Lab del Instituto Tecnológico de Massachusetts<sup>6</sup>. En el caso de la base de datos de la Universidad de Stanford, las imágenes fueron capturadas con una cámara Canon Digital Rebel XTi con un lente Canon de 10-22 mm. Las imágenes de la base de datos del MIT Media Lab fueron generadas sintéticamente mediante POV-Ray [16],

---

<sup>5</sup>The (New) Stanford Light Field Archive. Disponible Online: <http://lightfield.stanford.edu/lfs.html>

<sup>6</sup>Synthetic Light Field Archive. Disponible Online: <http://web.media.mit.edu/~gordonw/SyntheticLightFields>

Tabla 1: Imágenes de campo de luz usadas de la Universidad de Stanford

Base de datos Stanford		
Nombre Imagen	Vistas Angulares Originales	Resolución (pxs)
<i>Lego Bulldozer</i>	17x17	1536x1152
<i>Amethyst</i>	17x17	768x1024
<i>The Stanford Bunny</i>	17x17	1024x1024
<i>Jelly Beans</i>	17x17	1024x512
<i>Tarot Cards</i>	17x17	1024x1024

Tabla 2: Imágenes de campo de luz usadas del MIT Media Lab

Base de datos <i>MIT Media Lab</i>		
Nombre Imagen	Vistas Angulares Originales	Resolución (pxs)
<i>Dragon and Bunnies</i>	5x5	840x593
<i>Messerschmitt</i>	7x7	840x593
<i>Red Dragon</i>	7x7	512x384
<i>Green Dragon</i>	7x7	512x384

un software libre que permite crear escenarios tridimensionales. Para cada imagen los parámetros de simulación de distancia de vista y plano focal varían. Las especificaciones de las imágenes en cada base de datos se observan en las tablas 1 y 2.

### 3.2. CÁMARA LYTRO

La captura de imágenes con la cámara Lytro se llevo a cabo en el laboratorio de Óptica del grupo HDSP de la Universidad Industrial de Santander. El motivo de capturar las imágenes en el laboratorio es debido a que se cuenta con una mesa perforada con separación de 1 pulgada entre cada perforación 10, con lo que se puede determinar en forma más precisa el funcionamiento de los algoritmos de profundidad y de la metodología propuesta. Los objetos de la escena a capturar se ubicaron a diferentes profundidades tal y como se observa en la figura 10.

Figura 10: Las figuras se ubicaron a diferentes distancias del lente principal de la cámara.



### 3.3. ANÁLISIS ALGORITMOS DE ESTIMACIÓN DE PROFUNDIDAD

Para esta parte se seleccionaron algoritmos que cumplieran los siguientes requisitos: impacto en el área de investigación, ser algoritmos de fecha reciente de máximo 3 años y como ultimo requisito, que sean públicamente disponibles para su re-utilización. De acuerdo a estos criterios, los algoritmos usados son los propuestos por [3,9]. El funcionamiento de éstos algoritmos se presentó en la sección [1.2.2].

El primer paso para poder utilizar los algoritmos es la caracterización de los parámetros a usar, debido a esto, para estos algoritmos se realizó un barrido de valores en los parámetros para determinar los valores óptimos. En el caso del algoritmo de [9] se llegó a la conclusión de que los mismos valores de los parámetros recomendados por los autores son óptimos respecto a la información de profundidad obtenida, por lo cual se optó por trabajar con el algoritmo por defecto. Pero a diferencia, el algoritmo de [3] presentó mejores resultados con diferentes valores a los recomendados por los autores para las imágenes utilizadas. Los parámetros seleccionados se muestran en la tabla 3.

Un ejemplo de los resultados de la estimación de profundidad tradicional con los parámetros seleccionados se puede observar en la figura 11.

Tabla 3: Parámetros utilizados para la estimación de profundidad.

Parámetros [3]	
$d_{min}$	-1
$d_{max}$	2
$angularop$	's'
$outer$	9
$varest$	4
$aggregation$	1
$N$	6
$M$	1

Figura 11: Resultados de la estimación de profundidad tradicional. (a) imagen original, (b) método de [3], y (c) método de [9]

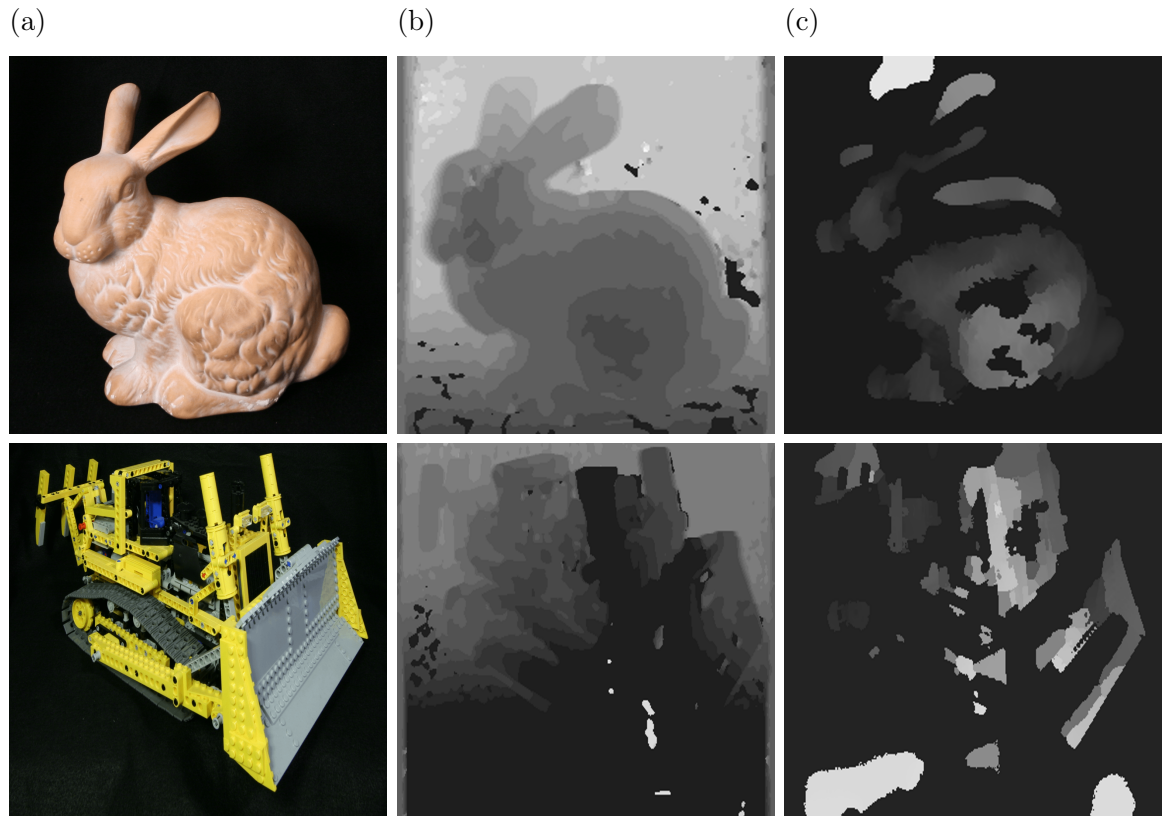
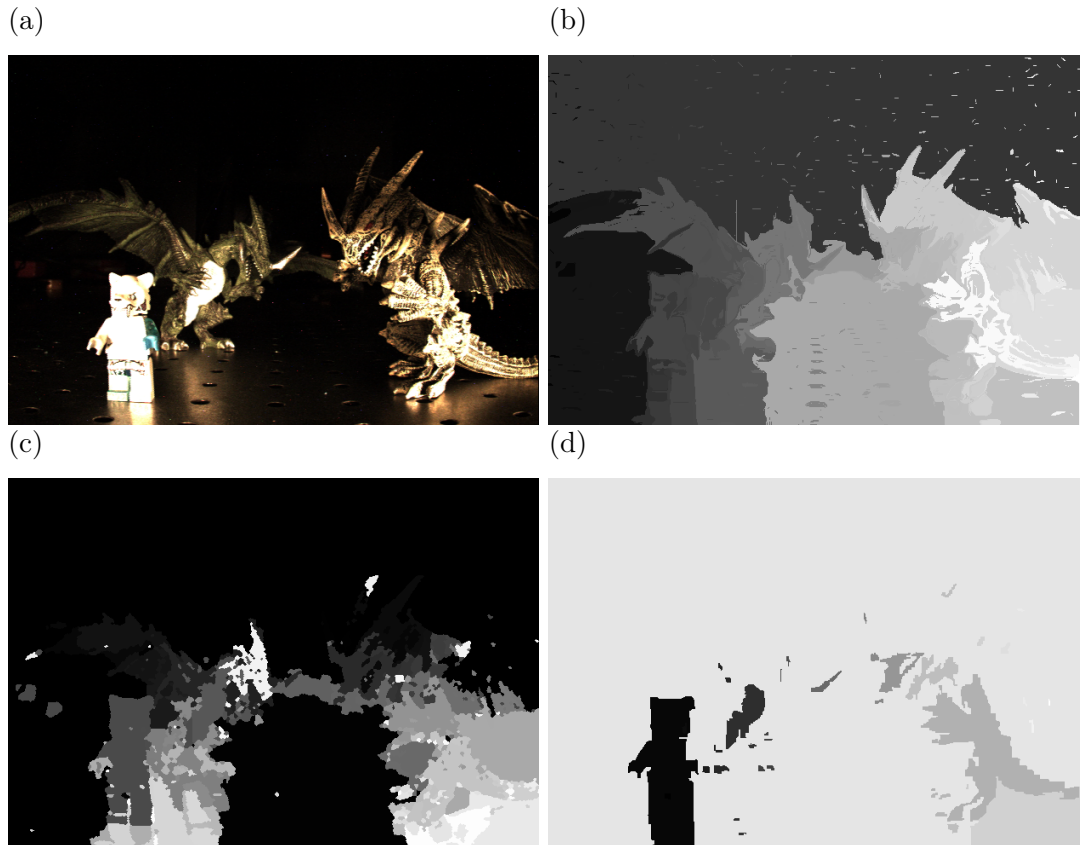


Figura 12: (a) Imagen de campo de luz capturada en el laboratorio. (b) Segmentación utilizando el algoritmo de Felzenszwalb y Huttenlocher. (c) Segmentación utilizando el algoritmo de Kovesi. (d) Segmentación utilizando el algoritmo de Nock y Nielsen



### 3.4. ANÁLISIS DE ALGORITMOS DE SEGMENTACIÓN PARA IMÁGENES DE CAMPO DE LUZ

Mediante la segmentación se desea obtener una información precisa de los bordes de los objetos. Para esta parte se seleccionaron los algoritmos que mejores resultados presentaban en la detección de bordes los cuales fueron [7, 10, 15]. Debido a que los métodos de segmentación solo funcionan para imágenes bi-dimensionales, la codificación de la imagen de campo de luz como un conjunto de imágenes 2D se realiza mediante un *toolbox* de procesamiento de imágenes de campo de luz proporcionado por [6]. La segmentación de una imagen de campo de luz mediante los métodos seleccionados se observa en la figura 12. Cabe aclarar que cada método de segmentación tiene diferentes

Tabla 4: Parámetros utilizados en cada uno de los algoritmos de segmentación utilizados.

Algoritmos de Segmentación		
Felzenszwalb	$\sigma = 3$ $min\ size = 100$	Valor de <i>threshold</i> para el área a segmentar. Mínimo tamaño de componentes de segmentación.
Kovesi	$k = 100$ $m = 1$ $seRadius = 1$ $colopt = median$	Número de Super píxeles deseados. Factor de ponderación entre el color y las diferencias espaciales. Radio mínimo de las regiones. Las más pequeñas se fusionan con la región adjacente Indica como se debe calcular el centro del color del grupo.
Nock	$Q = 8$	Número de regiones para segmentar. Debe ser un factor de $2^n$ .

parámetros. En el caso de los tres algoritmos utilizados, se seleccionaron los valores de dichos parámetros basados en recomendaciones de los autores y pruebas realizadas previamente mediante un barrido de valores entre un rango recomendado por los autores. Los valores de los parámetros utilizados se muestran en la tabla 4.

### 3.5. RESULTADOS DE LAS PRUEBAS

Debido a que se utilizaron dos tipos de imágenes, las disponibles públicamente y las capturadas por la cámara Lytro, se divide esta parte de resultados en dos partes. Durante el desarrollo de las pruebas cabe resaltar que se llevaron a cabo las posibles combinaciones entre los métodos de estimación de profundidad disponibles ([3, 9]) y los algoritmos de segmentación disponibles ([7, 10, 15]) con el fin de observar cual de esas combinaciones presentaba los mejores resultados. A manera de ilustración, en la figura 13 se puede observar algunos ejemplos de aplicar el método propuesto, donde en las imágenes de la primera fila y segunda fila pertenecientes a la base de datos de Stanford y MIT respectivamente, los mejores resultados en promedio se obtuvieron con la combinación del algoritmo de estimación de profundidad propuesto por [3] y los algoritmos de segmentación propuestos por [10, 15]. En la imagen de la tercera fila perteneciente a las imágenes capturadas en laboratorio, el mejor resultado fue la combinación del algoritmo de estimación de profundidad propuesto por [3] y el algoritmo de segmentación propuesto por [7]. .

Figura 13: Resultados experimentales en las bases de datos utilizadas. (a) Vista central en color de la escena, (b) método de [3], (c) método de [9], y (d) método propuesto. Las imágenes utilizadas corresponden a diferentes bases de datos donde la primera fila corresponde a la base de datos de Stanford, la segunda fila al MIT y la tercera fila a las imágenes originales capturadas en laboratorio.

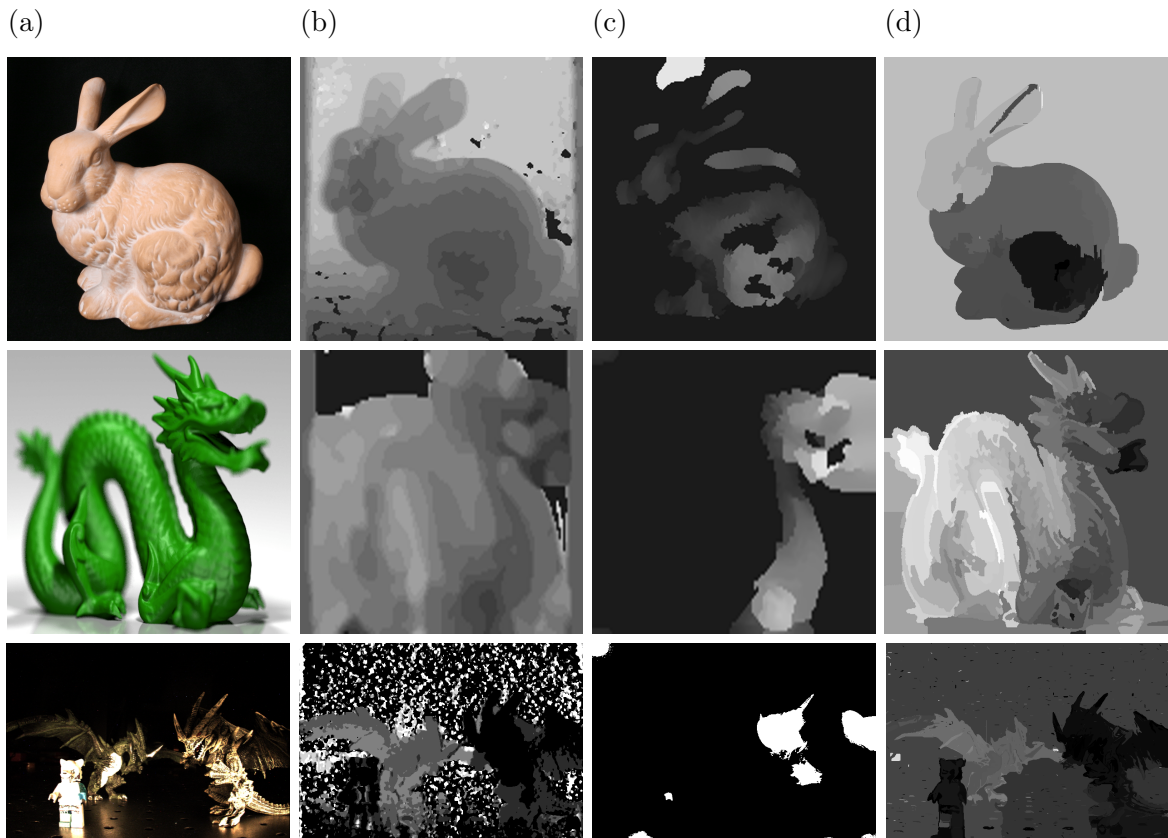
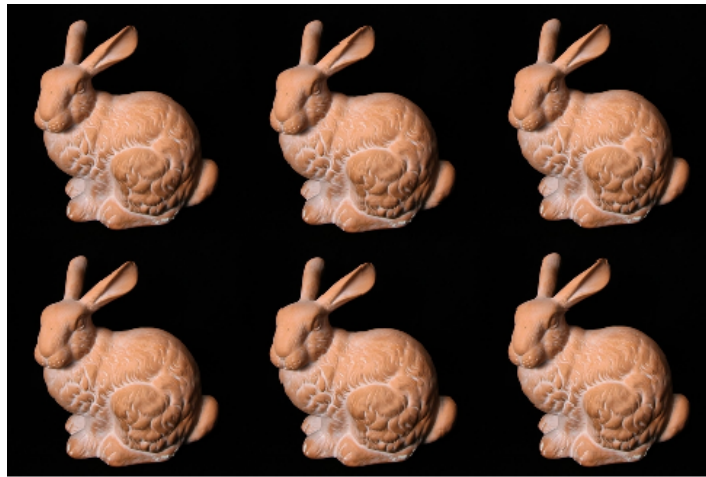


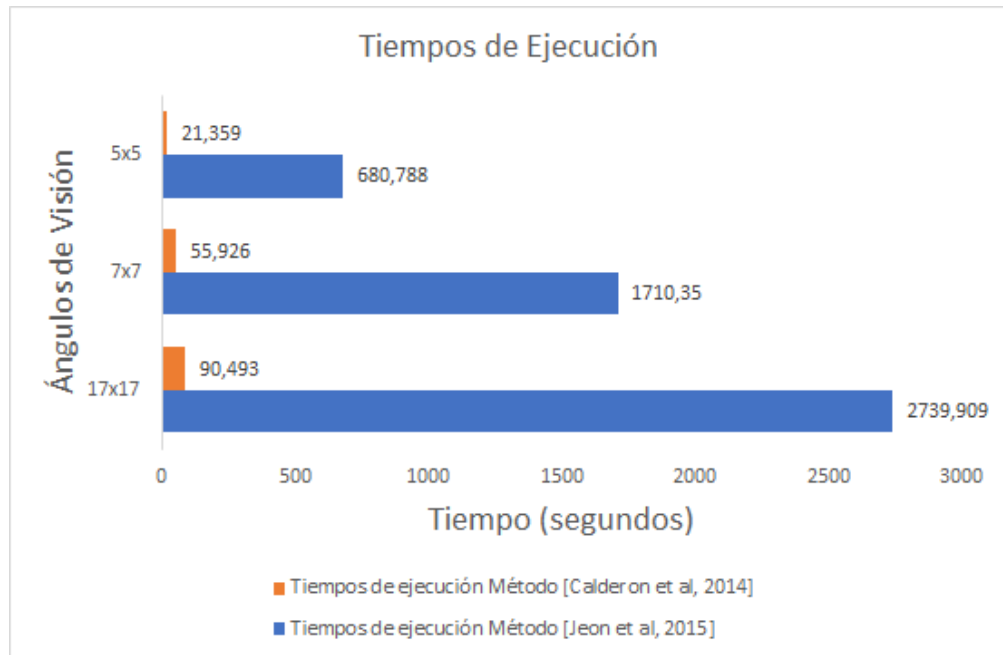
Figura 14: Las bases de datos presentan múltiples imágenes, cada imagen representa una vista angular diferente desde la cámara hacia el objeto



**3.5.1 Resultados en imágenes de las bases de datos** Estas imágenes cuentan con la característica de ser de tipo arreglo matricial. Este tipo de imagen consiste en la captura de una escena con una cámara común desde diferentes ángulos de vista mediante el movimiento de la cámara en direcciones verticales y horizontales. El movimiento de la cámara se lleva a cabo para replicar el efecto de los micro-lentes que posee una cámara Lytro tal como se dio a conocer en la sección 1.2.2 en la figura 4. Como resultado del movimiento de la cámara, las imágenes de las bases de datos son como se observa en la figura 14.

Durante el desarrollo de las pruebas, se observó que el proceso de obtener los mapas de profundidad con los métodos tradicionales conllevaba un alto costo computacional al procesar imágenes con 17x17 vistas angulares, por tal motivo se realizaron pruebas de la metodología para un número menor de vistas con el objetivo de observar si el número de vistas de la imagen afectaba de manera positiva o negativa a la metodología en general. Estas pruebas se llevaron a cabo para vistas angulares de 5x5, 7x7 y 17x17 las cuales eran las más comunes entre las bases de datos utilizadas. El computador de referencia para las pruebas fue un ASUS K53S con un procesador Intel(R) Core(TM) i5-2450M con una velocidad de CPU a 2.50GHz con 6Gb de memoria RAM y tarjeta de video dedicada NVIDIA GeForce 610M de 2Gb. Un resumen de los tiempos de ejecución de la metodología se observan en la figura 15 los cuales se obtuvieron mediante la función tic-toc de MATLAB.

Figura 15: Tiempos de estimación de profundidad. Para éste cálculo solo se tuvo en cuenta la duración del computo del mapa de profundidad mediante los métodos tradicionales debido a que el proceso de segmentación ocupaba en su totalidad menos de 10 segundos en comparación.



En la figura 16 se puede apreciar los resultados de la prueba de usar diferente cantidad de vistas angulares. Cabe resaltar que al usar menor cantidad de vistas angulares, obtenemos una mejor representación comparada con una mayor cantidad de vistas. En el caso donde se usaron 5x5 vistas angulares, se presentó una mejor detección completa del objeto además de reducir significativamente la duración del proceso debido a que se procesa una menor cantidad de información. Debido a éste resultado, se concluye en usar 5x5 vistas angulares para todas las pruebas hechas en las imágenes de tipo matricial.

Los resultados de la metodología propuesta utilizando los parámetros de las tablas 3 y 4 se observan en las figuras 17 y 18.

En las figuras 17 y 18 se puede detallar que la metodología propuesta cumple el objetivo planteado ya que presenta una alta eficiencia en la detección de bordes de la imagen manteniendo la información correspondiente a la profundidad.

**3.5.2 Resultados imágenes cámara Lytro** Para estas pruebas se utilizaron las imágenes capturadas previamente en el laboratorio con la cámara Lytro Illum. Recordando los parámetros de la tabla 4, específicamente el parámetro  $Q$  del algoritmo de [15], se tuvo que hacer una modificación en el valor del parámetro debido a que en las pruebas de las imágenes tipo matricial, solo se tenía un objeto en la mayoría de imágenes que era sencillo, por lo que al usar 8 regiones de segmentación era más que suficiente. El problema surge con las imágenes capturadas con la cámara Lytro ya que al utilizar una cantidad mayor de objetos con mayor de detalle en comparación con las imágenes en las bases de datos, por tal motivo, el parámetro  $Q = 8$  no fue suficiente para presentar una buena segmentación. Debido a esto, se realizó un nuevo barrido del parámetro  $Q$  obteniendo como mejor resultado en la detección de los objetos combinado con la metodología propuesta un valor de  $Q = 1024$ , la comparación del cambio de parámetro se observa en la figura 19.

### 3.6. ANÁLISIS CUALITATIVO

Durante el desarrollo de éste trabajo, se logran obtener mapas de profundidad con imágenes de campo de luz obtenidas de una base de datos publica en los que se mejora la detección de bordes en los objetos. Debido a la problemática de no poseer una métrica para corroborar los resultados obtenidos debido a no contar con un sensor infrarojo de medición de distancia especializado para éste tipo de aplicaciones, se llevó a cabo una encuesta con 92 participantes. En la encuesta, los participantes debían elegir el método

Figura 16: Resultados de aplicar la metodología propuesta con un número diferente de vistar de 5x5 hasta 17x17. (a) Imagen original, (b) 5x5 vistas angulares, (c) 7x7 vistas angulares, (c) 17x17 vistas angulares.

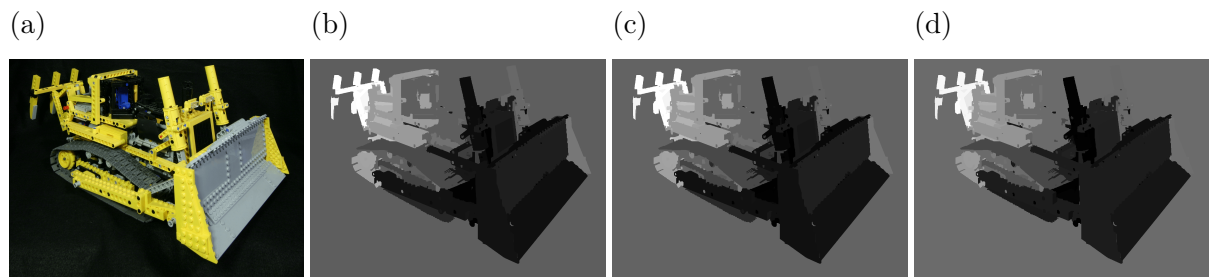


Figura 17: Resultados de la metodología. La primera columna corresponde a las imágenes originales, la segunda columna corresponde a la estimación de profundidad utilizando [3], la tercera columna corresponde a la estimación de profundidad utilizando [9], y la cuarta columna corresponde a la metodología propuesta.

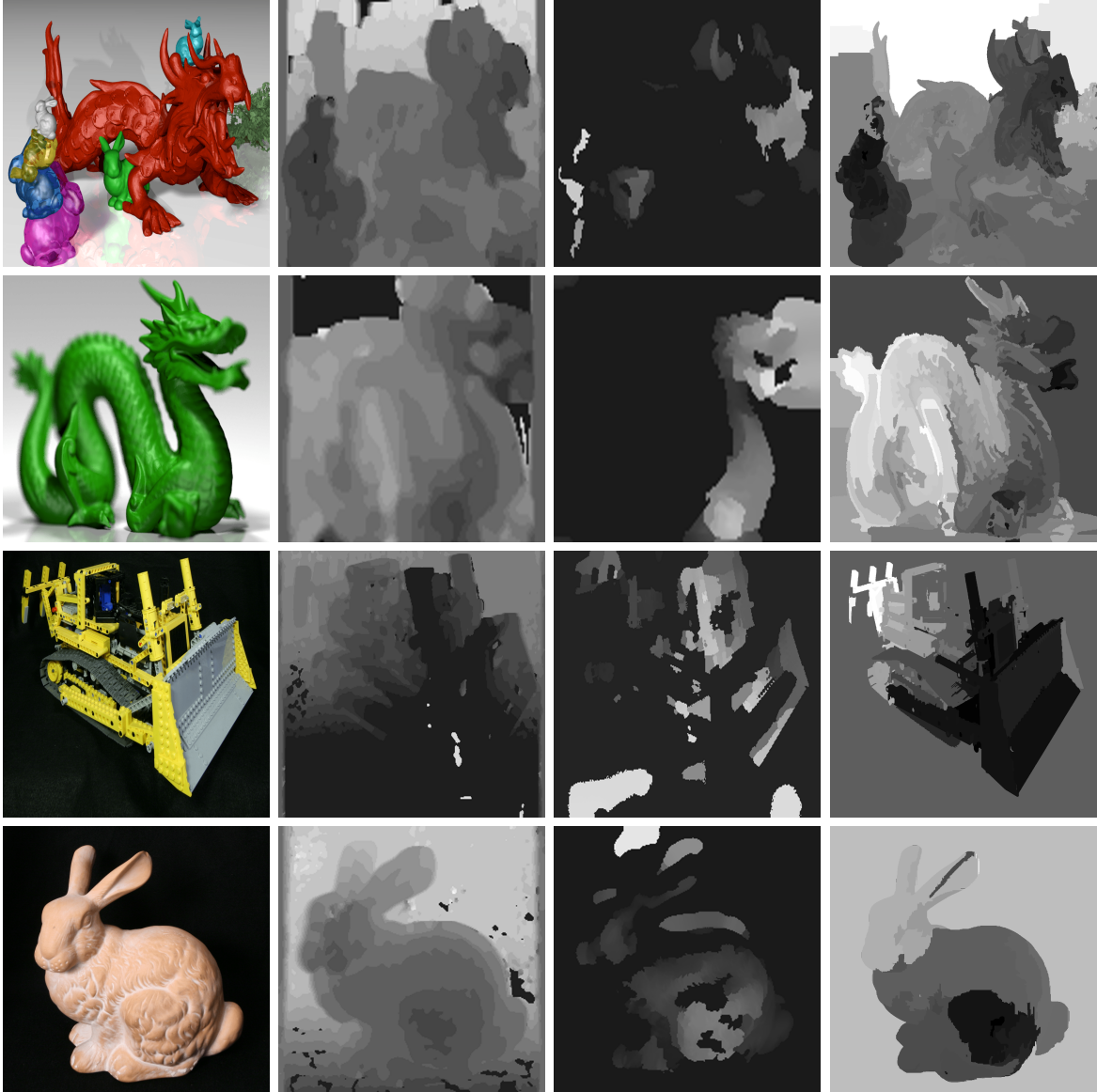


Figura 18: Continuación resultados de la metodología. La primera columna corresponde a las imágenes originales, la segunda columna corresponde a la estimación de profundidad utilizando [3], la tercera columna corresponde a la estimación de profundidad utilizando [9], y la cuarta columna corresponde a la metodología propuesta.

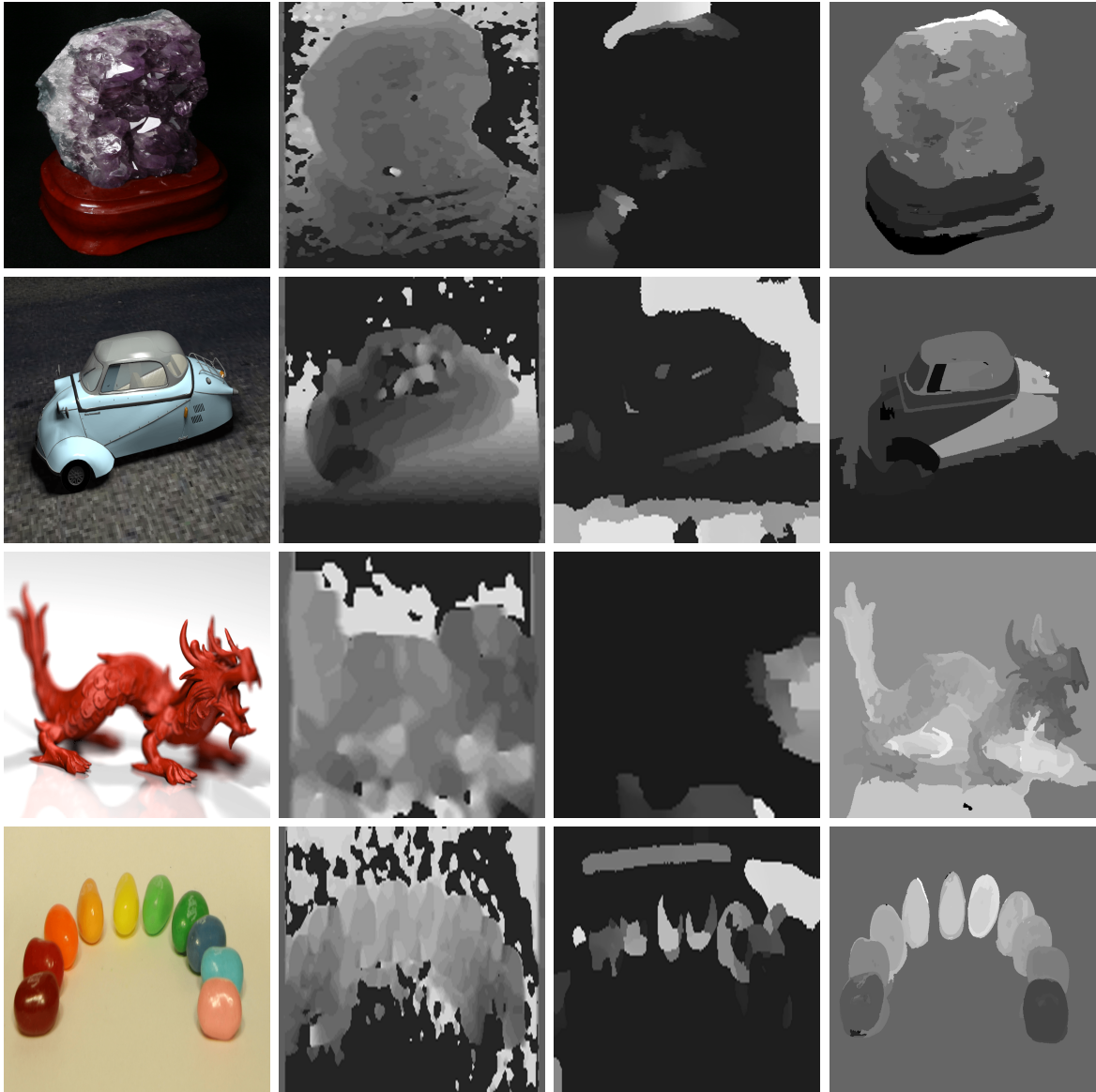


Figura 19: Comparación del parámetro  $Q$ . (a)  $Q = 8$  y (b)  $Q = 1024$

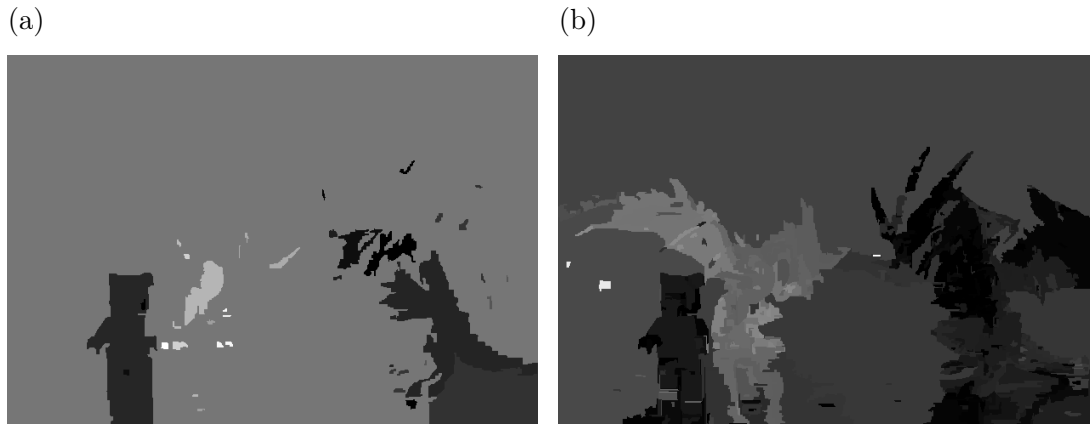


Figura 20: Resultados de la metodología. La primera columna corresponde a las imágenes originales, la segunda columna corresponde a la estimación de profundidad utilizando [3], la tercera columna corresponde a la estimación de profundidad utilizando [9], y la cuarta columna corresponde a la metodología propuesta.

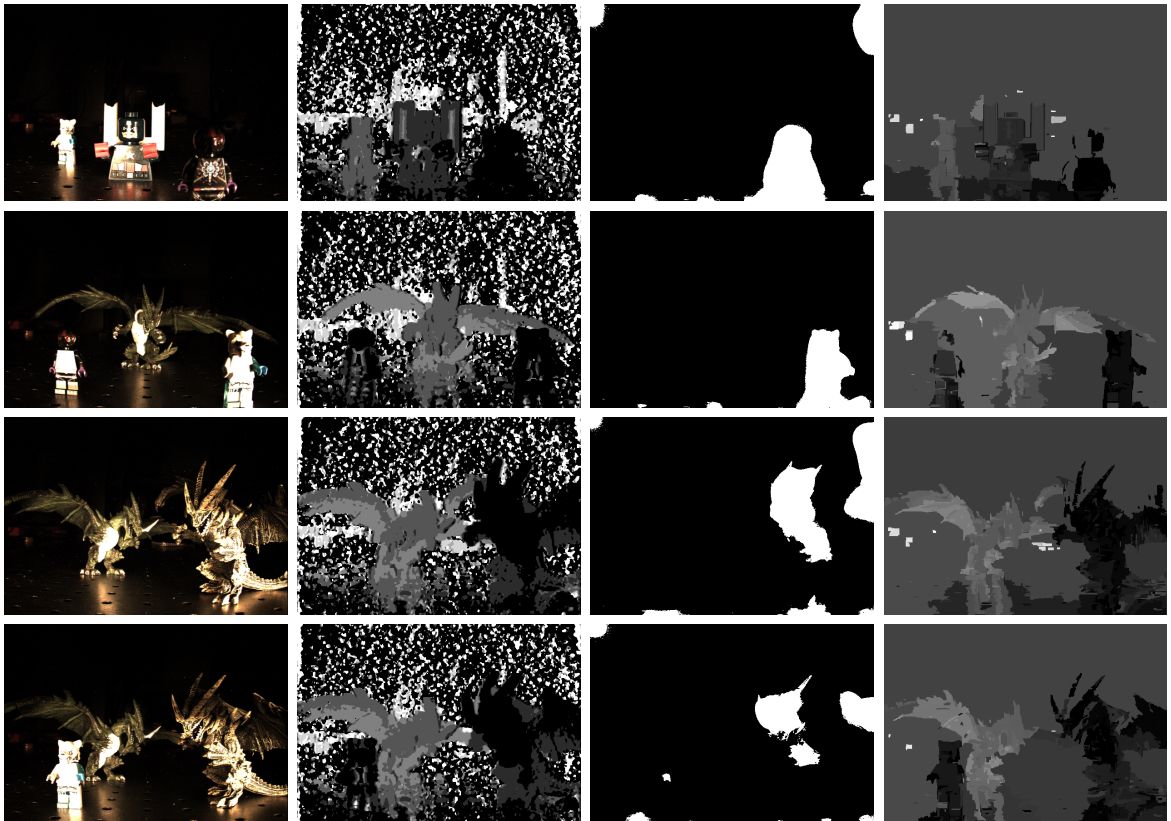


Tabla 5: Resultados de la encuesta

Comparación de Métodos		
Nombre de la escena	Calderon	Metodologia propuesta
Dragon and Bunnies	10.9 %	89.1 %
Green Dragon	6.5 %	93.5 %
Lego Bulldozer	4.3 %	95.7 %
Stanford Bunny	48.9 %	51.1 %
Amethyst	23.9 %	76.1 %
Messerschmitt	17.4 %	82.6 %
Red Dragon	5.4 %	94.6 %
Jelly Beans	4.3 %	95.7 %

que ellos consideraran mejor en una evaluación subjetiva respecto a la imagen original. Los resultados de la encuesta se muestran en la tabla 5.

## *Capítulo 4*

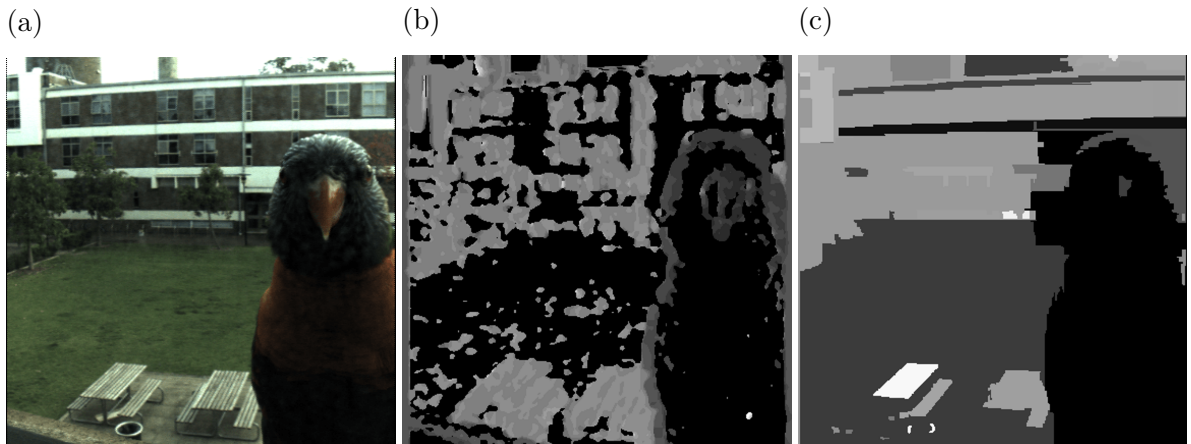
# *CONCLUSIONES*

En éste trabajo se propuso un método para estimación de profundidad en imágenes de campo de luz. El método se basa en el uso eficiente de la información de radiancia brindada por técnicas de segmentación como lo es la detección de contornos en los objetos. Se pudo comprobar el funcionamiento de ésta metodología y su eficiente desempeño como una base para futuros trabajos en el área de la estimación de profundidad y procesamiento de imágenes multi-dimensionales.

Durante las pruebas, se logró el objetivo principal que consistía en una mejora de la estimación de profundidad. Dicha mejora se ve reflejada en la detección del contorno de los objetos en la escena lo cual permite un reconocimiento más eficiente del objeto. A pesar de obtener resultados prometedores, durante la validación de la metodología se presentaron varias limitaciones. Por ejemplo, se pudo observar que esta metodología depende demasiado de la estimación de profundidad original. Si la estimación de profundidad original no es efectiva, esta metodología no puede compensar esos problemas ya que el objetivo principal de éste método es mejorar la estimación de los bordes para la detección de objetos en el mapa de profundidad. Un ejemplo de éste problema se muestra en la figura 21. Esa imagen de campo de luz fue capturada por una cámara Lytro Illum y fue obtenida de una base de datos pública [4]. Esto se debe a que la metodología aporta un método de estimación local y se podría complementar con una etapa previa de estimación global que entregue una mejor estimación de profundidad original.

Otro problema con éste método es la sobre-segmentación. Si la imagen está sobre-segmentada, el mapa de profundidad presenta sobre-segmentación también (figura 22).

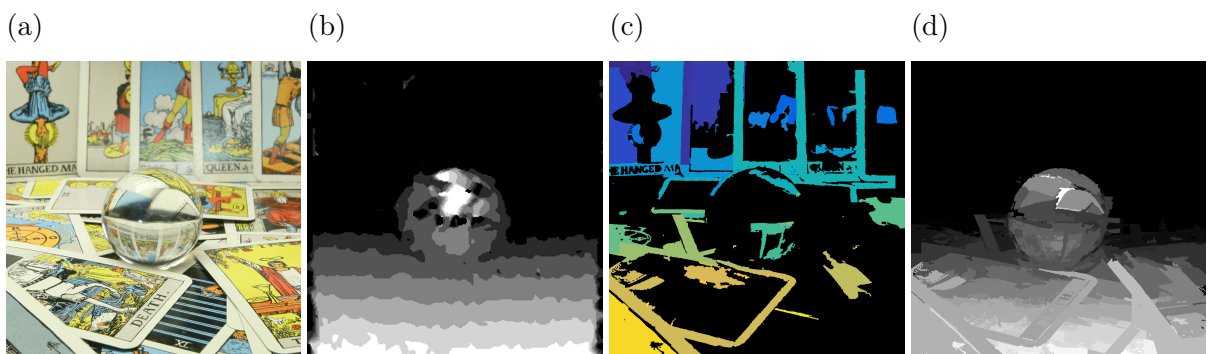
Figura 21: La estimación de profundidad por el método tradicional en (a) presenta un problema al diferenciar los planos de profundidad (b) que no se pueden corregir con la metodología (c).



En el caso de objetos transparentes como la bola de cristal en la figura 22, la sobre segmentación se presenta en la mayoría de la imagen debido a las cartas con alto detalle.

Para solucionar el problema de sobre-segmentación que se presenta con los objetos transparentes, se puede extender éste trabajo usando métodos como el presentado por [22] que se enfoca en la segmentación para objetos cristalinos. Además, en éste trabajo solo se explotaron 4 dimensiones de la función plenóptica. Como trabajo futuro se puede realizar un estudio de esta metodología en imágenes hiper-espectrales. Además, también cabe resaltar que éste trabajo se puede extender hacia un análisis más detallado hacia el desarrollo de una métrica que permita medir la eficiencia de los algoritmos de estimación

Figura 22: La imagen original (a) presenta sobre-segmentación en (c). Cuando se filtra con el mapa de profundidad tradicional en (b), la sobre-segmentación sigue presente en el nuevo mapa de profundidad (d).



de profundidad.

## ***BIBLIOGRAFÍA***

- [1] ALHWARIN, F., FERREIN, A., AND SCHOLL, I. *IR Stereo Kinect: Improving Depth Images by Combining Structured Light with IR Stereo*. Springer International Publishing, Cham, 2014, pp. 409–421.
- [2] BOUKHERS, Z., SHIRAHAMA, K., LI, F., AND GRZEGORZEK, M. Object detection and depth estimation for 3d trajectory extraction. In *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)* (June 2015), pp. 1–6.
- [3] CALDERON, F. C., PARRA, C. A., AND NIÑO, C. L. Depth map estimation in light fields using an stereo-like taxonomy. In *2014 XIX Symposium on Image, Signal Processing and Artificial Vision* (Sept 2014), pp. 1–5.
- [4] DANSEREAU, D. Lftoolbox samples. Accessed: 2016-03-15.
- [5] DANSEREAU, D., AND BRUTON, L. Gradient-based depth estimation from 4d light fields, May 2004.
- [6] DANSEREAU, D. G., PIZARRO, O., AND WILLIAMS, S. B. Decoding, calibration and rectification for lenselet-based plenoptic cameras., Jun 2013.
- [7] FELZENSZWALB, P. F., AND HUTTENLOCHER, D. P. Efficient graph-based image segmentation., Sept. 2004.
- [8] HARTLEY, R. I., AND ZISSERMAN, A. *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press, ISBN: 0521540518, 2004.
- [9] JEON, H. G., PARK, J., CHOE, G., PARK, J., BOK, Y., TAI, Y. W., AND KWEON, I. S. Accurate depth map estimation from a lenslet light field camera. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015), pp. 1547–1555.

- [10] KOVESI, P. Image segmentation using slic superpixels and dbscan clustering, April 2013. Accessed: 2016-02-15.
- [11] KUMAR, S., GUPTA, D., AND YADAV, S. Sensor fusion of laser and stereo vision camera for depth estimation and obstacle avoidance, 2010.
- [12] LEVOY, M., AND HANRAHAN, P. Light field rendering., 1996.
- [13] MALIK, A. S., MALIK, A. S., CHOI, T. S., AND NISAR, H. *Depth Map and 3D Imaging Applications: Algorithms and Technologies*, 1st ed. IGI Global, Hershey, PA, USA, 2011.
- [14] MCMILLAN, L., AND BISHOP, G. Plenoptic modeling: An image-based rendering system, 1995.
- [15] NOCK, R., AND NIELSEN, F. Statistical region merging., Nov. 2004.
- [16] PERSISTENCE OF VISION PTY, L. Persistence of vision pty. ltd. (2004) and persistence of vision (tm) raytracer and persistence of vision pty. ltd., williamstown, victoria, australia.
- [17] SAXENA, A., CHUNG, S. H., AND NG, A. Y. 3-d depth reconstruction from a single still image. *International Journal of Computer Vision* 76, 1 (Jan 2008), 53–69.
- [18] SHAO, W., SHENG, H., AND LI, C. *Segment-Based Depth Estimation in Light Field Using Graph Cut*. Springer International Publishing, Cham, 2015, pp. 248–259.
- [19] TAO, M. W., HADAP, S., MALIK, J., AND RAMAMOORTHY, R. Depth from combining defocus and correspondence using light-field cameras.
- [20] TAO, M. W., SRINIVASAN, P. P., HADAP, S., RUSINKIEWICZ, S., MALIK, J., AND RAMAMOORTHY, R. Shape estimation from shading, defocus, and correspondence using light-field angular coherence. *IEEE Transactions on Pattern Analysis and Machine Intelligence PP*, 99 (2016), 1–1.
- [21] WANG, T.-C., EFROS, A. A., AND RAMAMOORTHY, R. Depth estimation with occlusion modeling using light-field cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence PP*, 1 (2016), 1–1.

- [22] XU, Y., NAGAHARA, H., SHIMADA, A., AND TANIGUCHI, R. Transcut: Transparent object segmentation from a light-field image. corr, 2015.
- [23] ZHUO, W., SALZMANN, M., HE, X., AND LIU, M. Indoor scene structure analysis for single image depth estimation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015), pp. 614–622.