



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN Y DESARROLLO DE TECNOLOGÍA DIGITAL



**RECONOCIMIENTO DE OBJETOS TRIDIMENSIONALES EN IMÁGENES
MONOCULARES UTILIZANDO FILTROS DE CORRELACIÓN ADAPTATIVOS**

Que para obtener el grado de
Doctorado en Ciencias en Sistemas Digitales

Presenta:

KENIA PICOS ESPINOZA

Bajo la dirección de:

Dr. Víctor Hugo Díaz Ramírez

Comité tutorial:

Dr. Julio César Rolón Garrido

Dr. Moisés Sánchez Adame

Dr. Juan José Tapia Armenta

Dr. Leonardo Trujillo Reyes

Enero, 2017

Tijuana, B.C., México



INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de Tijuana, B.C. siendo las 9:30 horas del día 8 del mes de diciembre del 2016 se reunieron los miembros de la Comisión Revisora de Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de CITEDI

para examinar la tesis titulada:

Reconocimiento de objetos tridimensionales en imágenes monoculares utilizando filtros de correlación adaptativos.

Presentada por el alumno:

PICOS

Apellido paterno

ESPINOZA

Apellido materno

KENIA

Nombre(s)

Con registro:

A	1	4	0	8	1	7
---	---	---	---	---	---	---

aspirante de:

DOCTORADO EN CIENCIAS EN SISTEMAS DIGITALES

Después de intercambiar opiniones, los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Director de Tesis

DR. VÍCTOR HUGO DÍAZ RAMÍREZ

DR. JUAN JOSÉ TAPIA ARMENTA

DR. JULIO CÉSAR ROLÓN GARRIDO

DR. MOISÉS SÁNCHEZ ADAME

DR. LEONARDO TRUJILLO REYES

PRESIDENTE DEL COLEGIO DE PROFESORES

DR. JULIO CÉSAR ROLÓN GARRIDO

S.E.P.
INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN Y DESARROLLO
DE TECNOLOGÍA DIGITAL
DIRECCIÓN



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de Tijuana, Baja California, el día 12 del mes de diciembre del año 2016, la que suscribe **Kenia Picos Espinoza** alumna del Programa de DOCTORADO EN CIENCIAS EN SISTEMAS DIGITALES, con número de registro **A140817** adscrita al CENTRO DE INVESTIGACIÓN Y DESARROLLO DE TECNOLOGÍA DIGITAL, manifiesta que es el autor intelectual del presente trabajo de Tesis bajo la dirección del **Dr. Víctor Hugo Díaz Ramírez** y cede los derechos del trabajo titulado **Reconocimiento de objetos tridimensionales en imágenes monoculares utilizando filtros de correlación adaptativos**, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso de la autora y/o directores del trabajo. Este puede ser obtenido escribiendo a las siguientes direcciones Av. Instituto Politécnico Nacional No. 1310 Col Nueva Tijuana, Tijuana, Baja California, México, correo electrónico de contacto: **posgrado@citedi.mx**. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

A handwritten signature in blue ink, appearing to be 'Kenia Picos Espinoza', is written over the printed name below it.

Kenia Picos Espinoza

Dedicatoria

A Ulises, por su amor y apoyo incondicional.

A Dios, porque al que cree todo le es posible.

Agradecimientos

A mi director de tesis, Dr. Víctor Hugo Díaz Ramírez por haberme permitido formar parte de su equipo de investigación durante estos años y por sus valiosos consejos.

Al Dr. Leonardo Trujillo, Dr. Julio César Rolón, Dr. Juan José Tapia y Dr. Moisés Sánchez Adame por aportar a mi desarrollo profesional.

A los profesores Dr. Antonio Sanz Montemayor y Dr. Juan José Pantrigo por recibirme en su universidad durante mi estancia de investigación, gracias por su valiosa amistad.

Al Dr. Rigoberto Juárez Salazar por su disposición y paciencia al compartir sus conocimientos.

Al Centro de Investigación y Desarrollo de Tecnología Digital (CITEDI-IPN) del Instituto Politécnico Nacional, a la Comisión de Operación y Fomento de Actividades Académicas del IPN (COFAA), a la Fundación TELMEX, y al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo económico brindado para realizar mis estudios de posgrado.

A mis compañeros y amigos de CITEDI con los que pude compartir gratos momentos.

A mi familia, en especial a mi padre David Picos y mi madre Bertha Espinoza por su apoyo incondicional.

A mi esposo Ulises Orozco, que ha estado conmigo desde el inicio de este reto, gracias por siempre motivarme y creer en mí, eres mi mayor bendición, sin tí no lo hubiera logrado.

A Dios por darme la fuerza y valor para seguir adelante.

Resumen

En este trabajo se presenta un algoritmo para resolver el problema del reconocimiento de objetos tridimensionales (3D) en imágenes monoculares. El algoritmo es capaz de resolver problemas reales de reconocimiento de objetos y seguimiento de pose en 3D de manera eficiente. Para resolver la estimación de pose se utiliza la técnica de acoplamiento de plantillas a través de un banco de filtros de correlación, en conjunto con filtros de partículas, para estimar los parámetros de pose, dados en términos de localización, orientación y escala del objeto. Se asume que las imágenes de entrada están conformadas por un objeto de interés incrustado sobre un fondo, y están degradadas por ruido aditivo. De igual manera, la apariencia del objeto puede presentar cambios debido a fuentes incidentes de iluminación, distorsiones geométricas y oclusiones parciales del objeto. Los parámetros de pose se estiman mediante la detección del mejor acoplamiento entre la escena de entrada y la plantilla que contiene la vista paramétrica del objeto generada mediante gráficos por computadora. La búsqueda de pose es guiada mediante un filtro de partículas, que se basa en la cinemática del objeto mientras se desplaza dentro de la zona de detección. El algoritmo presenta robustez ante ruido aditivo y es adaptativo ante cambios de pose. Se presentan resultados experimentales con imágenes reales y sintéticas para demostrar la confiabilidad en el reconocimiento y la precisión en la estimación de los parámetros de pose del objeto. Se comparan los resultados con los métodos existentes en el estado del arte.

Palabras clave: reconocimiento de objetos 3D, filtros de correlación adaptativos, estimación de pose.

Abstract

An algorithm for three-dimensional (3D) object recognition in monocular images is presented. The algorithm is able to solve 3D object recognition and pose tracking problems with efficiency. The proposed algorithm uses a bank of correlation filters guided by particle filters to estimate the 3D pose parameters of a target, given by location, orientation and scaling. The input scene is assumed to be composed by an object of interest embedded into a cluttered background, and corrupted with additive noise. The appearance of the object can be modified due to incident light sources, geometrical distortions, and partial occlusions. The pose parameters of the target are estimated by finding the best match between the current view of the object and a parametric image template generated via computer graphics. The pose search is guided by particle filters, which take into account the kinematics of the target while moves within the detection zone. The proposed algorithm exhibits good robustness in noisy scenes, and an adaptation capability to the input scene perturbations. Experimental results obtained with synthetic and real images show that the proposed approach is able to estimate the pose parameters of a target with high accuracy when is compared to similar existing techniques.

Keywords: 3D object recognition, adaptive correlation filters, pose estimation.

Lista de Acrónimos

2D	Bidimensional
3D	Tridimensional
API	Interfaz de aplicaciones de programación, del inglés <i>application programming interface</i>
BRDF	Función de distribución bidireccional de reflectancia, del inglés <i>bidirectional reflectance distribution function</i>
CMF	Filtro de acoplamiento clásico, del inglés <i>classical matched filter</i>
CUDA	Arquitectura de computación de dispositivos unificados, del inglés <i>compute unified device architecture</i>
DC	Capacidad de discriminación, del inglés <i>discrimination capability</i>
FT	Transformada de Fourier, del inglés <i>Fourier transform</i>
GMF	Filtro de acoplamiento generalizado, del inglés <i>generalized matched filter</i>
GOF	Filtro óptimo generalizado, del inglés <i>generalized optimum filter</i>
GPU	Unidad de procesamiento gráfico, del inglés <i>graphics processing unit</i>
IFT	Transformada de Fourier inversa, <i>inverse Fourier transform</i>
LAE	Error absoluto de localización, del inglés <i>location absolute error</i>
LE	Error de localización, del inglés <i>location error</i>
LS	Búsqueda local, del inglés <i>local search</i>
LSPF	Filtro de partículas de búsqueda local, del inglés <i>local search particle filter</i>

MCRT	Método de trazado de ruta de Monte Carlo, del inglés <i>Monte Carlo path tracing</i>
MCRT	Método de trazado de rayos de Monte Carlo, del inglés <i>Monte Carlo ray tracing</i>
OAE	Error absoluto de orientación, del inglés <i>orientation absolute error</i>
OE	Error de orientación, del inglés <i>orientation error</i>
POE	Relación de pico a energía de salida, del inglés <i>peak-to-output energy ratio</i>
SDF	Funciones discriminantes sintéticas, del inglés <i>synthetic discriminant functions</i>
SNR	Relación señal a ruido, del inglés <i>signal-to-noise ratio</i>
VPL	Punto virtual de luz, del inglés <i>virtual point light</i>

Contenido

Resumen	i
Abstract	ii
1 Introducción	1
1.1 Motivación	5
1.2 Preguntas de investigación	6
1.3 Hipótesis	6
1.4 Objetivo	6
1.5 Productos de investigación	7
1.6 Organización del documento	10
2 Modelos y fundamentos del reconocimiento de objetos	12
2.1 Condiciones de las escenas en el reconocimiento de objetos	13
2.2 Generación de escenas mediante gráficos por computadora	19
2.3 Filtros de correlación	26
2.4 Filtros de partículas	31
3 Método propuesto para el reconocimiento de objetos con filtros de correlación adaptativos	35

3.1	Método para el reconocimiento de objetos 3D en escenas con condiciones de ruido, fondo e iluminación incidente	37
3.2	Algoritmo para el reconocimiento de objetos mediante cómputo paralelo heterogéneo	42
3.3	Algoritmo para el reconocimiento de objetos 3D con filtros de correlación y optimización de búsqueda local	44
3.4	Algoritmo para la estimación y seguimiento de pose 3D con filtros de correlación adaptativos y filtros de partículas	48
4	Resultados	58
4.1	Evaluación del algoritmo con distintos modelos de iluminación	59
4.2	Evaluación del algoritmo en una implementación computacional paralela heterogénea	63
4.3	Evaluación de desempeño en reconocimiento de pose 3D a partir de imágenes monoculares	69
4.4	Evaluación de desempeño para el seguimiento de pose 3D utilizando filtros de partículas	82
5	Conclusiones	91
5.1	Trabajo futuro	94
	Referencias	95

Lista de Figuras

2.1	Ilustración de diferentes modelos de imagen. (a) Objeto de referencia, (b) modelo aditivo, (c) modelo multiplicativo, y (d) modelo disjunto.	15
2.2	Representación gráfica de los modelos de iluminación. (a) Modelo de Lambert, (b) modelo de Phong, (c) modelo de Phong-Blinn, y (d) modelo de Cook-Torrance.	16
2.3	Representación de la función bidireccional de reflectancia.	19
2.4	Método de <i>ray tracing</i> desde el punto \mathbf{x} y su incidencia hacia diferentes puntos de la escena.	20
2.5	Visualización de un objeto con material translúcido con (a) fondo, y (b) otros objetos en la escena.	21
2.6	Resultados del método de generación de texturas sobre un objeto 3D utilizando patrones de (a) madera, (b) Voronoi, (c) ruido cúbico, (d) ajedrez, (e) mármol y (f) esferas.	22
2.7	Ilustración del método de Monte Carlo implementado en GPU con NVIDIA Optix en función del tiempo. (a) 250 ms, (b) 750 ms, (c) 1 s, (d) 2 s, (e) 5 s, (f) 10 s, (g) 100 s, y (h) 1000s.	24
2.8	Resultados del algoritmo de mapeo de fotones con enfoque progresivo implementado en GPU con NVIDIA Optix, en los intervalos de tiempo de (a) 100 ms, (b) 1 s, y (c) 10 s.	25

2.9	Diagrama a bloques de un sistema de reconocimiento de objetos mediante filtros de correlación.	30
2.10	Diagrama a bloques de un filtro de partículas.	32
3.1	Escena definida dentro de una zona de detección Ω que contiene un objeto con posición y orientación desconocidas.	37
3.2	Transformación de la geometría del objeto.	38
3.3	Representación geométrica de un mapeo de proyección de perspectiva.	39
3.4	Proyección de un plano de imagen de una escena 3D.	41
3.5	Diagrama a bloques de la implementación paralela con el algoritmo propuesto.	43
3.6	Diagrama a bloques del algoritmo propuesto con filtros de correlación adaptativos mediante búsqueda local.	45
3.7	Ejemplo de la construcción de un banco filtros mediante plantillas sintéticas del objeto.	46
3.8	Algoritmo propuesto para el reconocimiento y seguimiento de pose de un objeto 3D mediante filtros de correlación adaptativos guiados por filtros de partículas.	50
3.9	Modelo de señal representado para una escena 3D y los parámetros de pose del objeto.	51
3.10	Diagrama a bloques propuesto para un filtro de partículas con un enfoque de predicción dinámica.	52
3.11	Generación de imágenes de referencia para diferentes vistas del objeto.	53
4.1	Objeto de referencia con (a) material difuso y (b) su región de soporte.	59

4.2	Ejemplos de escenas utilizadas para el reconocimiento de objetos y el plano de correlación obtenido para diferentes materiales basados en los modelos: (a) Lambert, (b) Phong, (c) Blinn-Phong, (d) Ward, y (e) Cook-Torrance.	60
4.3	Desempeño en el reconocimiento de objetos en términos de DC para diferentes materiales, ante condiciones de ruido de (a) 50 dB, (b) 30 dB, y (c) 10 dB SNR.	62
4.4	Ejemplo de cuadros de entrada con la estimación real (línea sólida, \rightarrow) y estimada (línea punteada, $--\rightarrow$) de los ángulos de orientación.	65
4.5	Desempeño del algoritmo paralelo propuesto en términos de DC con respecto (a) al número de cuadros, y (b) al número de filtros utilizados en el banco.	66
4.6	Error de localización del algoritmo paralelo propuesto con respecto (a) al número de cuadros, y (b) al número de filtros utilizados en el banco.	66
4.7	Error de orientación del algoritmo paralelo propuesto con respecto (a) al número de cuadros, y (b) al número de filtros utilizados en el banco.	67
4.8	Tiempo de ejecución de la implementación paralela del algoritmo propuesto.	69
4.9	Ejemplos de los objetos 3D utilizados en los experimentos con diferentes poses. (a) bunny, (b) dragon y (c) teapot.	70
4.10	Ejemplos de escenas sintéticas utilizadas en los experimentos. (a)-(c) <i>bunny</i> , (d)-(f) <i>dragon</i> , (g)-(i) <i>teapot</i> . Las soluciones candidatas se representan en amarillo, y la solución final se representa en verde.	72
4.11	Desempeño del algoritmo propuesto en términos de (a) LE y (b) OE, en escenas con ruido, respecto al número de filtros de correlación.	73

4.12	Eficiencia de la detección del algoritmo propuesto en términos de la DC y el número de iteraciones en el proceso de refinamiento, para bancos con 4^3 , 5^3 , 6^3 y 7^3 filtros de correlación.	73
4.13	Resultados obtenidos del desempeño del reconocimiento de pose en escenas ruidosas. (a) $\sigma^2 = 0.25$, (b) $\sigma^2 = 0.5$, (c) $\sigma^2 = 0.75$, y (d) $\sigma^2 = 1$	74
4.14	Desempeño del algoritmo propuesto en términos de (a) LE y (b) OE en escenas ruidosas con respecto al número de iteraciones en la etapa de refinamiento.	75
4.15	Desempeño del algoritmo propuesto cuando el 25% del objeto se encuentra ocluido.	76
4.16	Desempeño del algoritmo propuesto cuando el 50% del objeto se encuentra ocluido.	77
4.17	Evaluación del algoritmo para la estimación de pose del objeto parcialmente ocluido, utilizando los modelos 3D (a) <i>bunny</i> , (b) <i>dragon</i> , (c) <i>teapot</i>	78
4.18	Diferentes vistas de los objetos impresos en 3D utilizados en los experimentos. (a–c) <i>bunny</i> , (d–f) <i>dragon</i> , (g–i) <i>teapot</i>	79
4.19	Resultados obtenidos con el algoritmo propuesto para el reconocimiento de pose en escenas reales.	80
4.20	Ejemplo de la operación del algoritmo propuesto en la etapa de refinamiento en diferentes iteraciones para una (a) escena sintética, y una (b) escena real.	80
4.21	Vistas de los modelos 3D utilizados en los experimentos. (a) <i>bunny</i> , (b) <i>dragon</i> , (c) <i>elephant</i> , y (d) <i>teapot</i>	83
4.22	Desempeño del algoritmo propuesto en términos de (a) LE y (b) OE, en relación al número de cuadros en la escena y número de partículas utilizadas.	84

4.23	Resultados de estimación de pose con diferentes niveles de ruido aditivo: (a) $\sigma^2 = 0.1$, (b) $\sigma^2 = 0.4$, y (c) $\sigma^2 = 1$	85
4.24	Comparación del algoritmo propuesto con diferentes métodos de estimación de pose en escenas ruidosas, en términos de (a) LAE, y (b) OAE, respectivamente.	86
4.25	Comportamiento de partículas durante el proceso de estimación de pose 3D en términos de OE con respecto a los tres ángulos de rotación, en el cuadro de escena (a) 1, (b) 2, (c) 4, (d) 5, (e) 10, y (f) 15.	87
4.26	Concentración de partículas a través del proceso de refinamiento en términos de DC , I y la función mono-objetivo propuesta J , en los cuadros de escena (a) 1, (b) 2, (c) 4, (d) 5, (e) 10, y (f) 15.	88

CAPÍTULO 1

Introducción

El reconocimiento de objetos mediante visión por computadora ha sido investigado ampliamente en las últimas décadas. Esto se debe a que hoy en día existe un gran número de aplicaciones en donde se requieren sistemas automáticos para la identificación de objetos en apoyo a la toma de decisiones [1, 2]. Algunas de las aplicaciones más importantes del reconocimiento de objetos pueden abarcar desde sistemas de vigilancia, inspección industrial, reconocimiento en imágenes médicas, realidad aumentada, y asistencia en sistemas para la navegación.

Para diseñar un sistema de reconocimiento de objetos efectivo, es necesario considerar que existen diversos factores que deterioran el desempeño del sistema cuando se utilizan imágenes capturadas en escenas reales. La presencia de ruido, objetos adicionales en el fondo, modificaciones geométricas, oclusiones parciales y la influencia de fuentes de iluminación incidente añaden mayor complejidad al problema de reconocimiento [3, 4]. Los factores mencionados modifican considerablemente la apariencia del objeto de interés con respecto al punto de observación, lo que dificulta su identificación y disminuye el desempeño del sistema [5].

La tarea de estimar con precisión la pose de un objeto a partir de imágenes monoculares es una de las áreas de interés más importantes en el reconocimiento

de objetos tridimensionales [6, 7]. El reconocimiento de pose consiste en estimar descriptores geométricos de un objeto, tales como las coordenadas de localización y los ángulos de orientación, en escenas que contienen a un objeto incrustado en un espacio tridimensional, mediante el procesamiento de imágenes monoculares.

El problema de reconocimiento de pose 3D es complejo de resolver debido a que la pose del objeto en un espacio tridimensional puede variar con múltiples grados de libertad [8, 9]. Es notable que una combinación particular de los parámetros de pose del objeto de interés resulta en una vista única del objeto en la escena observada. Entonces, el espacio de búsqueda de los parámetros de pose a través de imágenes de la escena observada tiende a ser extremadamente alto. Por lo tanto, se requiere el diseño de un algoritmo eficiente para la detección y la estimación precisa de la pose de un objeto de interés en un espacio 3D, mediante el procesamiento de imágenes en escenas monoculares y con una complejidad computacional aceptable.

En la literatura reciente, el reconocimiento de la pose de un objeto 3D ha sido resuelto a través de diferentes enfoques. Un enfoque típico consiste en la búsqueda de características específicas en la imagen, que son extraídos de la escena observada mediante un pre-procesamiento [10]. Estas características pueden ser, por ejemplo, líneas, formas, color y puntos de interés. Además, de que las características en una imagen son resultado de una proyección 2D a partir de un modelo 3D conocido del objeto [2, 11]. Este enfoque ha demostrado ser muy eficiente en diversas aplicaciones [1]. Sin embargo, cuando la escena se degrada debido a factores tales como el ruido aditivo y un fondo desordenado, el desempeño de los algoritmos por extracción de características por lo general es muy bajo, por lo tanto, la estimación de pose resultante no es precisa.

Otro método efectivo para el reconocimiento de pose 3D se realiza mediante métodos de segmentación 2D [9, 12]. En estos métodos, el objetivo es realizar el

acoplamiento del contorno del objeto en la escena observada con el contorno de la proyección 2D del objeto en una pose en particular [13].

Esta técnica es rápida y produce buenos resultados en términos generales. Sin embargo, el desempeño de este método depende de la eficiencia del método de segmentación. Si el método de segmentación es sensible al ruido o a objetos falsos incrustados en el fondo, el contorno se podría extraer de manera imprecisa, y en consecuencia se podría comprometer el desempeño general del algoritmo. Además, debido a que en esta técnica solamente se utiliza el contorno del objeto, se descarta información que pudiera ser útil para estimar la pose 3D del objeto, por ejemplo, el color, la distribución de intensidades y la textura.

Sandhu et al. [14] introdujo un método efectivo para el reconocimiento de pose 3D basado en regiones con funciones iniciales. Este método optimiza el modelo de referencia del objeto con el fin de estimar los parámetros de pose. El algoritmo combina información de un conjunto disponible de modelos 3D que pertenecen a la misma clase de objeto.

El campo del reconocimiento de objetos mediante filtros de correlación es una opción atractiva para llevar a cabo la detección de objetos en escenas con condiciones de ruido aditivo y disjunto [8]. El filtrado por correlación consiste en calcular el acoplamiento entre la imagen de la escena observada y de una plantilla que contiene la información del objeto a detectar. Los filtros de correlación han demostrado obtener estimaciones precisas de la detección del objeto en una escena [15, 16]. Además, los filtros de correlación poseen formulaciones matemáticas que permiten su implementación en procesadores opto-digitales, o en dispositivos digitales programables.

Al realizar la correlación entre dos señales, se ejecuta un sistema lineal que produce un plano de intensidades que representa el grado de acoplamiento entre ellas. El valor máximo del plano de correlación de salida indica las coordenadas estimadas de la localización del objeto en la escena. Los filtros de correlación se diseñan con el

fin de reconocer un objeto en una proyección 2D con respecto a la imagen. Cada imagen de referencia del objeto es conocida como plantilla, la cual contiene una pose específica del objeto. Además, en aplicaciones de reconocimiento de patrones tolerantes a distorsiones, se emplean múltiples vistas del objeto mediante un banco de filtros [17]. Al realizar el acoplamiento de plantillas puede obtenerse una alta precisión en la estimación de los parámetros de localización y orientación del objeto.

En este contexto, el reconocimiento de pose es un problema que se puede resolver mediante un banco de filtros de correlación. Debido a la gran cantidad de vistas que un objeto puede presentar dentro de una escena, se plantea el reconocimiento de pose como un problema de búsqueda. Para esto, se utiliza un enfoque de filtros adaptativos con una metodología capaz de optimizar la estimación de pose de un objeto en términos de la precisión en la localización y la orientación con respecto a la escena de entrada.

En este trabajo se presenta un algoritmo de reconocimiento de pose de un objeto rígido en un espacio tridimensional utilizando filtros de correlación adaptativos a partir de imágenes monoculares. Debido a que se conoce previamente un modelo 3D del objeto de interés, un conjunto de imágenes sintéticas se generan mediante técnicas de gráficos por computadora. Cada plantilla se construye con una vista paramétrica del objeto dada por una proyección 2D del modelo 3D con parámetros de pose establecidos. Se ejecuta el acoplamiento de plantillas, y se utiliza el filtro con la mejor correspondencia entre la plantilla y la escena de entrada. Finalmente, se aplica un proceso iterativo de refinamiento mediante un método híbrido de optimización basado en filtros de partículas.

El algoritmo propuesto ha demostrado ser robusto en imágenes con ruido, fondo, distorsiones geométricas, fuentes incidentes de iluminación y oclusiones parciales del objeto, entregando así, una alta precisión en la estimación de la solución, en términos de la localización, la orientación y el escalamiento.

1.1 Motivación

Debido a la gran demanda de sistemas de visión por computadora en diversos campos de la ingeniería, surge la necesidad de desarrollar sistemas robustos para la detección, clasificación y seguimiento de objetos en escenas 3D, que puedan operar en un tiempo de ejecución computacional eficiente.

En la actualidad, existen diversos métodos para solucionar el reconocimiento de patrones de manera efectiva [11, 18, 19]. El reconocimiento de patrones mediante filtros de correlación representa una solución eficiente en términos de robustez al ruido y precisión en la localización del objeto, en contraste con otras técnicas conocidas [20, 21]. Además, se ha demostrado que al incorporar modelos de predicción del estado del objeto en escenas dinámicas, es posible aumentar la flexibilidad de los métodos basados en correlación, conservando sus ventajas [21, 22].

Ante la gran demanda de sistemas para aplicaciones en tiempo real, los dispositivos con arquitecturas de cómputo paralelo juegan un papel importante [20, 23]. Con el empleo de procesadores gráficos se pueden obtener resultados con alta precisión, invirtiendo una menor cantidad de tiempo de procesamiento [24].

El problema del reconocimiento de objetos 3D continua siendo un problema abierto, debido a que es muy complejo diseñar un sistema que sea robusto a diferentes visualizaciones que un objeto puede presentar en una escena. El diseño de un sistema de reconocimiento de objetos 3D con un buen desempeño en escenarios complejos afectados por ruido, cambios de pose, y diferentes fuentes de iluminación, representa la principal motivación para el desarrollo de esta investigación ya que presenta una gran oportunidad hacia nuevos desarrollos innovadores que impacten otras áreas de la ciencia e ingeniería.

1.2 Preguntas de investigación

Las preguntas de investigación que motivan a esta tesis doctoral se presentan a continuación.

- ¿Cuáles son los factores principales que determinan la eficiencia de un sistema de reconocimiento de objetos?
- ¿Cuál es el desempeño de un sistema de reconocimiento de objetos basado en filtros de correlación para escenas en 3D ante diferentes condiciones de ruido, fondo, iluminación, y oclusiones parciales del objeto?
- ¿Cuál es el método computacional más eficiente para el reconocimiento de objetos 3D a partir de imágenes monoculares?

1.3 Hipótesis

El uso de filtros de correlación adaptativos en el problema de reconocimiento de objetos tridimensionales ante condiciones de ruido, distorsiones geométricas, iluminación no uniforme y oclusión, permitirá lograr la estimación de los parámetros de localización y pose de un objeto de interés observado en imágenes monoculares.

1.4 Objetivo

Objetivo general

Desarrollar un algoritmo basado en filtros de correlación adaptativos, para el reconocimiento de objetos tridimensionales y estimación de pose en imágenes monoculares, que sea robusto a la presencia de ruido aditivo y disjunto, cambios de iluminación, distorsiones geométricas, y oclusiones parciales en la escena.

Objetivos particulares

- Identificar los factores principales que determinan la estimación de un sistema de reconocimiento de objetos basado en filtros de correlación.
- Caracterizar el desempeño del sistema de reconocimiento de objetos ante diferentes condiciones de la escena, tales como ruido, oclusión, y fuentes incidentes de iluminación.
- Desarrollar un algoritmo para el reconocimiento de objetos 3D que incorpore cómputo paralelo heterogéneo utilizando arquitecturas de alto desempeño.

1.5 Productos de investigación

En este trabajo de investigación se han desarrollado aportaciones originales para resolver problemas de detección, reconocimiento, estimación de pose y seguimiento de objetos en escenas 3D. En el periodo 2014-2017 se han realizado diferentes actividades y productos de investigación, los cuales se enlistan a continuación.

- **Picos K.**, Díaz-Ramírez V. H., Kober, V., Montemayor, A. S., Pantrigo, J., Accurate 3D pose recognition from monocular images using template matched filtering, 2016, Optical Engineering (revista JCR).
- **Picos K.**, Díaz-Ramírez V. H., Pose detection of a 3D object using template matched filtering. Proc. SPIE 9970, Optics and Photonics for Information Processing X, 99700S, 2016. (artículo de congreso internacional).
- **Picos K.**, Juarez-Salazar, R., Díaz-Ramírez V. H., Fringe-projection method for three-dimensional digitization of human faces. Proc. SPIE 9970, Optics and Photonics for Information Processing X, 99701Q, 2016 (artículo de congreso internacional).

- Orozco-Rosas, U., **Picos K.**, Montiel, O., Sepúlveda, R., Díaz-Ramírez V. H., Obstacle recognition for path planning in autonomous mobile robots. Proc. SPIE 9970, Optics and Photonics for Information Processing X, 99700X, 2016 (artículo de congreso internacional).
- **Picos K.**, González-Fraga, J. A., Díaz-Ramírez V. H., Correlation filtering for three-dimensional object detection from point clouds scenes. Congreso Internacional de Ciencias Computacionales CiComp, Ensenada, B.C., 2016 (congreso nacional).
- **Picos K.**, Díaz-Ramírez V. H., Montemayor, A. S., Pantrigo, J., A parallel CPU/GPU algorithm for 3D pose estimation using CUDA and OpenMP, GPU Technology Conference, San Jose, California, USA, 2016 (póster de congreso internacional).
- **Picos K.**, Díaz-Ramírez V. H., A 3D approach for object recognition in illuminated scenes with adaptive correlation filters. Proc. SPIE 9598, Optics and Photonics for Information Processing IX, 95981E, 2015 (artículo de congreso internacional).
- **Picos K.**, Díaz-Ramírez V. H., 3D Object recognition considering light conditions, GPU Technology Conference, San Jose, California, USA, 2015 (póster de congreso internacional).
- **Picos K.**, Tapia J. J., Díaz-Ramírez V.H., Real-time 3D Video Processing using Multi-stream GPU Parallel Computing, Research in Computing Sciences, vol. 80, 2014 (revista nacional arbitrada, Latindex).
- **Picos K.**, Tapia J. J., Díaz-Ramírez V.H., Real-time 3D Video Processing using Multi-stream GPU Parallel Computing, Congreso Internacional de Ciencias Computacionales CiComp, Ensenada, B.C., 2014 (congreso nacional).

- **Picos K.**, Díaz-Ramírez V. H., Juan J. Tapia, A roadmap to global illumination in 3D scenes: solutions for GPU object recognition applications. Proc. SPIE 9216, Optics and Photonics for Information Processing VIII, 921607, 2014 (congreso internacional).
- Díaz-Ramírez V. H., **Picos K.**, Kober V., Target Tracking in Nonuniform Illumination Conditions using Locally-adaptive Correlation Filters, 2014, Optics Communications (revista JCR).
- **Picos K.**, Object Tracking under non-uniform illumination conditions, GPU Technology Conference, San Jose, California, USA, 2014 (congreso internacional).

Otras actividades relevantes de formación profesional:

- Estancia de investigación en la Universidad Rey Juan Carlos (URJC), Madrid, España, de enero a marzo de 2015.
- Colaboración con el grupo de investigación de la Escuela Técnica Superior de Ingeniería Informática de la URJC, Madrid, España, 2015-2016.
- Colaboración con el Departamento de Ciencias Computacionales del CICESE, Ensenada, Baja California, México, 2014-2016.
- Asistencia a cursos disciplinarios de Cómputo Paralelo de Altas Prestaciones y *Soft Computing*, por la URJC, Madrid, España, 2015.
- Participación en el proyecto *Autonomos* de vehículos autónomos a escala 1:10 dentro del marco del año dual México-Alemania, 2016.

1.6 Organización del documento

El presente trabajo está organizado de la siguiente manera. En el Capítulo 2, se establecen los fundamentos teóricos que sustentan esta investigación. Se abordan los conceptos fundamentales acerca del reconocimiento de objetos y se describen los modelos de imagen más importantes. Se analizan las diversas condiciones que se presentan en las escenas reales, tales como ruido, distorsiones geométricas, oclusiones, y cambios de apariencia del objeto debido a fuentes de iluminación. En este capítulo se estudian los filtros de correlación a partir de los modelos de imagen analizados. De igual manera, se aborda el método de filtrado de partículas con el fin de resolver el reconocimiento y seguimiento de objetos.

En el Capítulo 3, se presenta el método de solución propuesto para el reconocimiento de objetos en 3D. Se exponen las propuestas realizadas para la estimación de parámetros que modifican la pose de un objeto, tales como localización, orientación, escala, así como el estudio de la iluminación incidente dentro de la escena. El algoritmo propuesto utiliza filtros de correlación adaptativos para resolver el reconocimiento de objetos 3D en una metodología planteada en una metodología de cuatro etapas. Primero, se plantea un método para el reconocimiento de objetos con escenas en condiciones de ruido fondo y cambios de iluminación incidente. Además, se propone un algoritmo utilizando filtros adaptativos mediante computación paralela heterogénea. De igual manera, se diseña un algoritmo para el reconocimiento de objetos 3D a partir de imágenes monoculares mediante acoplamiento de plantillas y optimización de búsqueda local. Por último se desarrolla una propuesta metodológica para la estimación y seguimiento de pose de un objeto dinámico utilizando un enfoque adaptativo mediante filtros de partículas.

En el Capítulo 4, se presentan los resultados obtenidos con el método propuesto en términos de métricas de calidad y desempeño computacional. En este capítulo se analiza el desempeño de los algoritmos desarrollados para resolver el reconocimiento

de objetos en escenas ante diferentes condiciones de la escena. Se analiza el desempeño con el algoritmo paralelo desarrollado en una arquitectura heterogénea. Así también, se evalúan los resultados obtenidos mediante filtros adaptativos de correlación.

Finalmente, el Capítulo 5 resumen las conclusiones obtenidas de este trabajo de tesis. Se presentan las recomendaciones para trabajos investigación futura para el reconocimiento de objetos en 3D.

CAPÍTULO 2

Modelos y fundamentos del reconocimiento de objetos

El reconocimiento automático de objetos en un espacio tridimensional es un área de investigación multidisciplinaria importante para la toma de decisiones en aplicaciones de la ingeniería, tales como los sistemas de visión inteligentes, la robótica, la manufactura, la asistencia a vehículos autónomos, los sistemas de vigilancia, entre otros [6, 25].

El problema del reconocimiento de objetos consiste en detectar objetos de interés en una escena observada, estimar su estado con respecto a su posición y orientación, y asignar cada objeto detectado a una categoría. El reconocimiento de objetos puede realizarse a través de varias técnicas, tales como sistemas basados en características [26–28], utilizando técnicas metaheurísticas y de aprendizaje [29, 30], o a través de filtros de correlación [15, 16].

Sin embargo, el problema de reconocimiento de pose 3D es complejo de resolver debido a que la pose del objeto puede variar con múltiples grados de libertad dentro del espacio tridimensional [8, 9]. Además, la apariencia del objeto en una escena puede

degradarse por diversos factores, tales como ruido aditivo, fondo, influencia de fuentes incidentes de luz, y oclusiones parciales del objeto [3, 5, 31].

En este capítulo se presentan los fundamentos teóricos para abordar el problema de reconocimiento de objetos. Se incluye un análisis de los modelos matemáticos de señal más importantes para representar las imágenes con distintas condiciones. De igual manera, se presentan las herramientas necesarias para la generación de escenas sintéticas mediante gráficos por computadora. Por último, se explica la teoría fundamental para el reconocimiento de objetos mediante filtros de correlación utilizando acoplamiento de plantillas y filtros de partículas.

2.1 Condiciones de las escenas en el reconocimiento de objetos

Una escena capturada de un ambiente real contiene diversas condiciones que pueden comprometer el desempeño de cualquier sistema de reconocimiento de objetos. Las imágenes de entrada pueden contener ruido aditivo, un fondo desordenado, distorsiones geométricas, oclusiones parciales del objeto, y diversas fuentes incidentes de iluminación.

Es posible representar analíticamente dichas condiciones mediante modelos de señal. Los modelos de señal son formulaciones matemáticas para describir una imagen en términos de los elementos que la componen. En el reconocimiento de objetos, los modelos de señal básicos que se utilizan comúnmente son el aditivo, el disjunto, y el multiplicativo [32]. Ejemplos de estos modelos se muestran en la Fig. 2.1.

El modelo aditivo (Fig. 2.1(b)) describe una imagen $f(x, y)$ que contiene a un objeto $t(x, y)$ en una posición desconocida (x_0, y_0) , y donde la escena es afectada por

la presencia de ruido $n(x, y)$, de la siguiente manera

$$f(x, y) = t(x - x_0, y - y_0) + n(x, y). \quad (2.1)$$

Comúnmente, las coordenadas desconocidas (x_0, y_0) en la Ec. (2.1) son consideradas variables aleatorias con distribución uniforme. El modelo multiplicativo (Fig. 2.1(c)) es utilizado comúnmente para representar modificaciones de intensidad debido a cambios en la iluminación de la escena. Este modelo puede expresarse como

$$f_d(x, y) = f(x, y)d(x, y), \quad (2.2)$$

donde $f_d(x, y)$ indica la imagen degradada y $d(x, y)$ una función de degradación. El modelo disjunto (Fig. 2.1(d)) es útil para representar una escena en donde un objeto se encuentra incrustado dentro de un fondo $b(x, y)$, como

$$f(x, y) = t(x - x_0, y - y_0) + b(x, y)\bar{w}(x - x_0, y - y_0) + n(x, y), \quad (2.3)$$

donde $\bar{w}(x, y)$ es una función binaria que representa la región de soporte inversa del objeto en la posición (x_0, y_0) , dada por

$$\bar{w}(x, y) = \begin{cases} 0, & \text{dentro del área de } t(x, y) \\ 1, & \text{fuera del área de } t(x, y). \end{cases} \quad (2.4)$$

De igual manera, es necesario considerar que las condiciones de iluminación de la escena juegan un papel importante. En aplicaciones de visión por computadora, la luz incidente influye en la apariencia de un objeto, lo que produce un cambio en el desempeño en un sistema de reconocimiento de objetos [22]. La apariencia de un objeto depende de la interacción entre la fuente de iluminación y el material de su superficie [33].

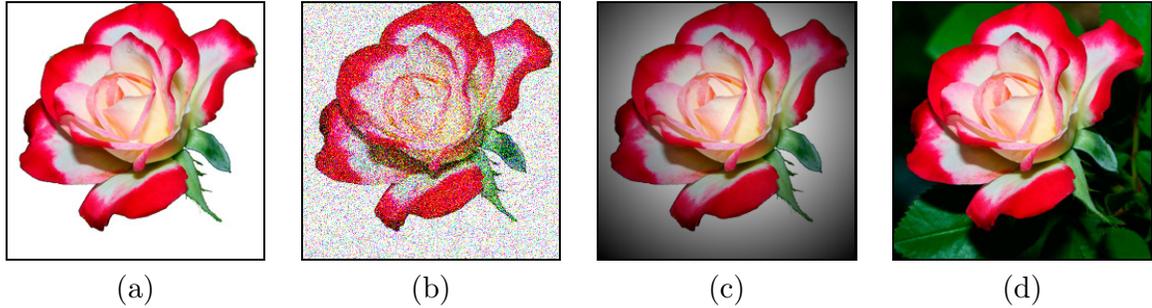


Figura 2.1: Ilustración de diferentes modelos de imagen. (a) Objeto de referencia, (b) modelo aditivo, (c) modelo multiplicativo, y (d) modelo disjunto.

La estimación computacional de la iluminación en una escena, es una tarea de notable complejidad. Para ello, se estudian modelos determinísticos caracterizados por un conjunto de funciones matemáticas simplificadas que simulan el comportamiento de la luz en la superficie de los objetos, sin incluir todos los parámetros de las leyes de la física óptica [34].

Los modelos de iluminación describen, de una manera simplificada, la interacción de la luz con los materiales de la superficie de los objetos. La geometría descrita en la Fig. 2.2 ilustra un punto P en un plano, que es iluminado por un rayo incidente de luz ω_i , mismo que se refleja en la dirección del observador ω_r .

El caso más simple se representa por el modelo difuso, conocido como modelo de Lambert [35]. En la Fig. 2.2(a) la luz se dispersa uniformemente en todas las direcciones alrededor del punto P . La cantidad de luz sobre la superficie depende del coseno del ángulo de incidencia θ_i , que es equivalente al producto punto entre el vector normal a la superficie \mathbf{n} y la dirección ω_i . El color de la superficie C (definido por la longitud de onda de la luz que refleja el material), la luz ambiental I_a , y la intensidad del color de la luz I_i son definidos comúnmente por un triplete RGB. La iluminación consiste en la ponderación de coeficientes constantes de luz difusa y ambiental (K_d y K_a , respectivamente).

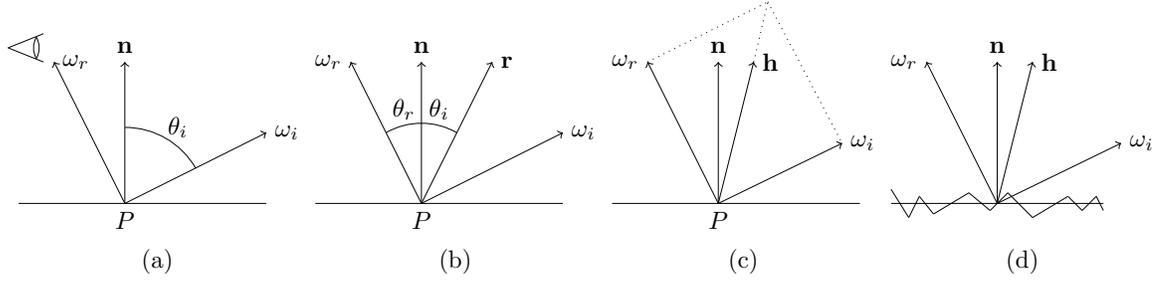


Figura 2.2: Representación gráfica de los modelos de iluminación. (a) Modelo de Lambert, (b) modelo de Phong, (c) modelo de Phong-Blinn, y (d) modelo de Cook-Torrance.

En términos de estas cantidades, la intensidad de la luz reflejada dada una superficie, está dada por [36]

$$I = K_a I_a + K_d C I_i \max((\mathbf{n} \cdot \omega_i), 0). \quad (2.5)$$

Una desventaja del modelo de Lambert es que no considera las reflexiones especulares.

En cambio, el modelo de iluminación de Phong agrega un término especular al modelo difuso, a través de una constante K_s [37]. El color de la reflexión especular C_s , el vector reflejado \mathbf{r} (entre ω_i y \mathbf{n}), y el coeficiente de rugosidad p , son términos que conforman la iluminación especular, como

$$I = K_a I_a + K_d C I_i \max((\mathbf{n} \cdot \omega_i), 0) + K_s C_s I_i \max((\mathbf{r} \cdot \omega_r), 0)^p, \quad (2.6)$$

donde ω_r es el valor unitario en la dirección del observador, como se muestra en la Fig. 2.2(b). Un caso adicional de iluminación especular está dado por el modelo de Phong-Blinn [37, 38], representado en la Fig. 2.2(c). Obsérvese que este modelo incluye un vector intermedio \mathbf{h} entre ω_i y ω_r , representado por la relación $\mathbf{h} = \frac{\omega_i + \omega_r}{\|\omega_i + \omega_r\|}$. Entonces, a partir de la Ec. (2.6), la iluminación descrita por el modelo Phong-Blinn,

está dada por

$$I = K_a I_a + K_d C I_i \max((\mathbf{n} \cdot \boldsymbol{\omega}_i), 0) + K_s C_s I_i \max((\mathbf{n} \cdot \mathbf{h}), 0)^p. \quad (2.7)$$

Otro modelo que considera un término especular está dado por el modelo de iluminación isotrópica de Ward [39]. Éste incluye un término exponencial, dado por

$$I = K_a I_a + K_d C I_i \max((\mathbf{n} \cdot \boldsymbol{\omega}_i), 0) + K_s C_s I_i \exp(-p \tan \theta_i^2), \quad (2.8)$$

donde θ_i es el ángulo entre los vectores \mathbf{n} y \mathbf{h} (dados en la Ec. (2.7)). Comúnmente, la Ec. (2.8) se reescribe en términos del producto punto, como

$$I = K_a I_a + K_d C I_i \max((\mathbf{n} \cdot \boldsymbol{\omega}_i), 0) + K_s C_s I_i \exp\left(-p \frac{1 - (\mathbf{n} \cdot \mathbf{h})^2}{(\mathbf{n} \cdot \mathbf{h})^2}\right). \quad (2.9)$$

El modelo de Cook-Torrance es conocido por entregar resultados precisos al representar superficies metálicas [40]. Este modelo utiliza un término conocido como función de Fresnell $F(\theta)$ y considera que la superficie está compuesta por un arreglo de pequeños espejos para lograr reflexiones intensas [41], como se muestra en la Fig. 2.2(d). El modelo de Cook-Torrance se expresa por

$$I = I_i \max(0, (\mathbf{n} \cdot \boldsymbol{\omega}_i)) \frac{F(\theta) G D_B(\alpha)}{(\mathbf{n} \cdot \boldsymbol{\omega}_i)(\mathbf{n} \cdot \boldsymbol{\omega}_r)}, \quad (2.10)$$

donde $G = \min\left(1, \frac{2(\mathbf{n} \cdot \mathbf{h})(\mathbf{n} \cdot \boldsymbol{\omega}_r)}{(\boldsymbol{\omega}_r \cdot \mathbf{h})}, \frac{2(\mathbf{n} \cdot \mathbf{h})(\mathbf{n} \cdot \boldsymbol{\omega}_i)}{(\boldsymbol{\omega}_r \cdot \mathbf{h})}\right)$ es el factor de atenuación de la geometría [42], y \mathbf{h} es el vector intermedio descrito en la Eq. (2.7). El término $D_B(\alpha)$ es la función de distribución de Beckmann que caracteriza la rugosidad de la superficie [36], dada por $D_B(\alpha) = \frac{1}{m^2(\mathbf{n} \cdot \mathbf{h})^4 \alpha} \exp\left(-\frac{1 - (\mathbf{n} \cdot \mathbf{h})^2}{(\mathbf{n} \cdot \mathbf{h})^2 m^2}\right)$, donde $\alpha = \arccos(\mathbf{n} \cdot \mathbf{h})$, y m es un parámetro de control.

Se ha establecido un modelo generalizado, que a diferencia de los modelos descritos anteriormente, no hace simplificaciones tan drásticas, sino que considera términos de la teoría electromagnética con el fin de obtener resultados más precisos al modelar escenas reales.

En este sentido, el estudio de la luz que incide sobre una escena no sólo se estudia como un rayo, sino también como una onda. A esta representación se le conoce como modelo de iluminación global [43]. En el modelo de iluminación global, la radiancia total en una escena L_o , equivale al flujo radiante¹ por unidad de área por ángulo sólido, medido en $\text{W m}^{-2} \text{sr}^{-1}$, y está dada por la radiancia emitida L_e y reflejada L_r , que se representa por

$$L_o(\mathbf{x}, \omega_r) = L_e(\mathbf{x}, \omega_r) + L_r(\mathbf{x}, \omega_r). \quad (2.11)$$

La función $L_r(\mathbf{x}, \omega_r)$ se puede expresar en términos de la radiancia incidente integrando sobre el hemisferio Ω como

$$L_r(\mathbf{x}, \omega_r) = \int_{\Omega} f_r(\mathbf{x}, \omega_r, \omega_i)(\mathbf{n} \cdot \omega_i)L_i(\mathbf{x}, \omega_i)d\omega_i. \quad (2.12)$$

Por lo tanto, la Ec. (2.11) puede reescribirse como [44]

$$L_o(\mathbf{x}, \omega_r) = L_e(\mathbf{x}, \omega_r) + \int_{\Omega} f_r(\mathbf{x}, \omega_r, \omega_i)(\mathbf{n} \cdot \omega_i)L_i(\mathbf{x}, \omega_i)d\omega_i, \quad (2.13)$$

donde $f_r(\mathbf{x}, \omega_r, \omega_i)$ equivale a la función bidireccional de reflectancia (BRDF, por sus siglas del inglés *Bidirectional Reflectance Distribution Function*) y está dada en función de la dirección de la iluminación incidente ω_i y reflejada ω_r para cada posición \mathbf{x} , dentro del hemisferio Ω .

¹potencia de radiación electromagnética, medida en watts (W). También definida como la energía que transportan las ondas de luz por unidad de tiempo.

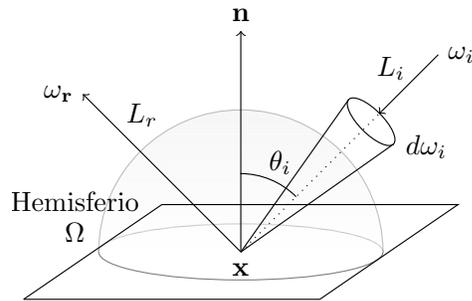


Figura 2.3: Representación de la función bidireccional de reflectancia.

Como se muestra en la Fig. 2.3, la BRDF se puede expresar como la relación entre la iluminación reflejada y la irradiancia², tal como

$$f_r(\mathbf{x}, \omega_r, \omega_i) = \frac{dL_r(\mathbf{x}, \omega_r)}{dE_i(\mathbf{x})} = \frac{dL_r(\mathbf{x}, \omega_r)}{dL_i(\mathbf{x}, \omega_i) \cos \theta_i d\omega_i}, \quad (2.14)$$

donde la irradiancia $E_i(\mathbf{x})$ equivale a [45]

$$E_i(\mathbf{x}) = \int_{\Omega} L_i(\mathbf{x}, \omega_i) (\mathbf{n} \cdot \omega_i) d\omega_i. \quad (2.15)$$

Si el material del objeto a reconocer es conocido, la función $f_r(\mathbf{x}, \omega_r, \omega_i)$ se puede aproximar por medio de los modelos de iluminación previamente definidos, como se muestra en la Fig. 2.2.

2.2 Generación de escenas mediante gráficos por computadora

Con el fin de producir una aproximación precisa de la visualización 3D de un objeto a través de una escena 2D, se requiere procesar una cantidad muy significativa de

²potencia de radiación electromagnética incidente en una superficie por área, medida en W m^{-2} .

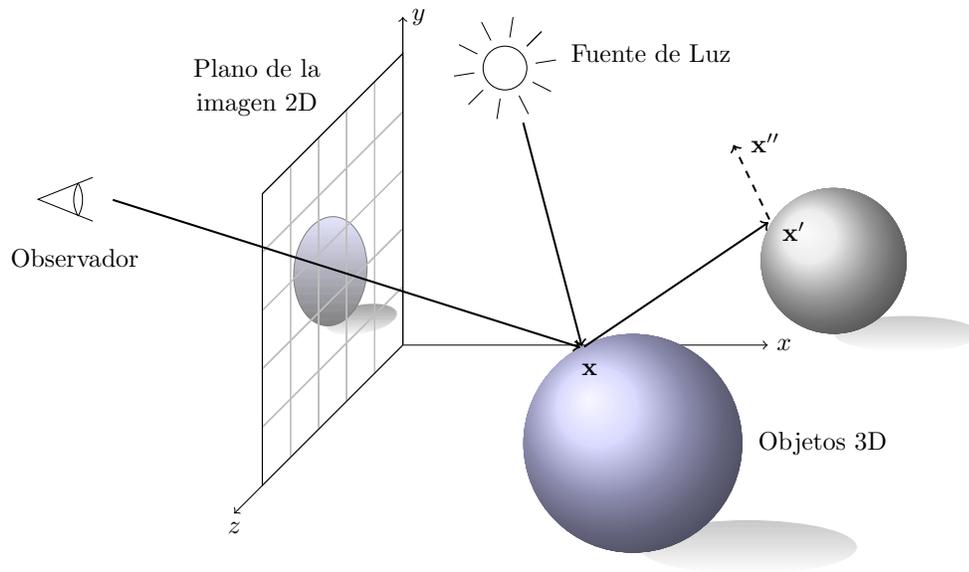


Figura 2.4: Método de *ray tracing* desde el punto x y su incidencia hacia diferentes puntos de la escena.

datos con el fin de modelar la interacción de la luz con el material de la superficie del objeto [46].

El estudio del fenómeno de la interacción de la luz y las superficies continúa siendo un problema abierto, principalmente, en aplicaciones en tiempo real. A partir de las nuevas arquitecturas de procesamiento de gráficos por computadora, se han desarrollado varios métodos de síntesis de imágenes para representar de forma realista la apariencia de la luz incidente sobre un objeto [47–49]. Los algoritmos recientes para el cálculo de la iluminación emplean modelos matemáticos y estrategias computacionales que permiten soluciones eficientes en tiempo real.

En la literatura reciente se estudian los métodos más comunes para resolver la iluminación global por computadora [49, 50]. Para simular la iluminación por gráficos por computadora, la BRDF se expresa en términos de todas las direcciones incidentes sobre un punto de la superficie. La BRDF se evalúa en cada punto de la escena, y el promedio de todos los resultados convergen a la solución real. Por lo general, la

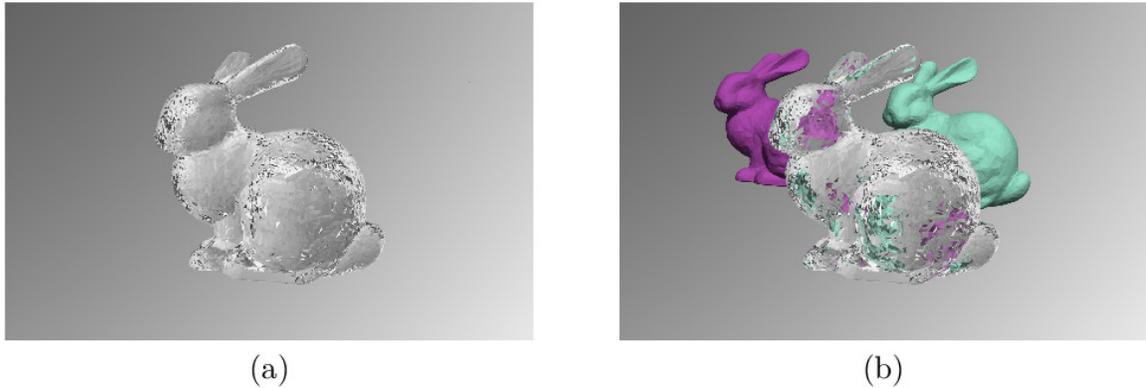


Figura 2.5: Visualización de un objeto con material translúcido con (a) fondo, y (b) otros objetos en la escena.

función resultante se obtiene utilizando el método de trazado de rayos (conocido en inglés como *ray tracing*), el cual se envía un rayo con dirección conocida y en el primer punto en que toca una superficie, se calcula la intensidad de la luz emitida, tal como se muestra en la Fig. 2.4. La radiación de entrada L_i en la posición \mathbf{x} que proviene de una dirección arbitraria ω_i , es la radiación de salida de otro punto de superficie \mathbf{x}' , que se puede encontrar trazando el rayo desde \mathbf{x} en la dirección ω_i . La radiación de salida de \mathbf{x}' se calcula nuevamente por la Ec. (2.13), y la radiación incidente en la posición \mathbf{x}' que proviene de otra dirección arbitraria, puede a su vez calcularse por el trazo desde \mathbf{x}' en esa dirección, y así sucesivamente [43].

Las primeras demostraciones de este enfoque mediante una implementación en GPU se hicieron en 2002 por Purcell et al [51]. Esta implementación se puede ejecutar en múltiples hilos concurrentes y funciones de GPU en paralelo [47, 52].

Algunas aplicaciones más actuales del método de trazado de rayos son capaces de procesar 100×10^6 rayos de luz incoherente por segundo [49, 53, 54]. La dificultad en este método radica en el procesamiento de múltiples incidencias de rayos de luz.

Cuando el material del objeto que se desea representar es translúcido visualmente, los objetos adyacentes que rodean al objeto de interés también influyen en el cálculo

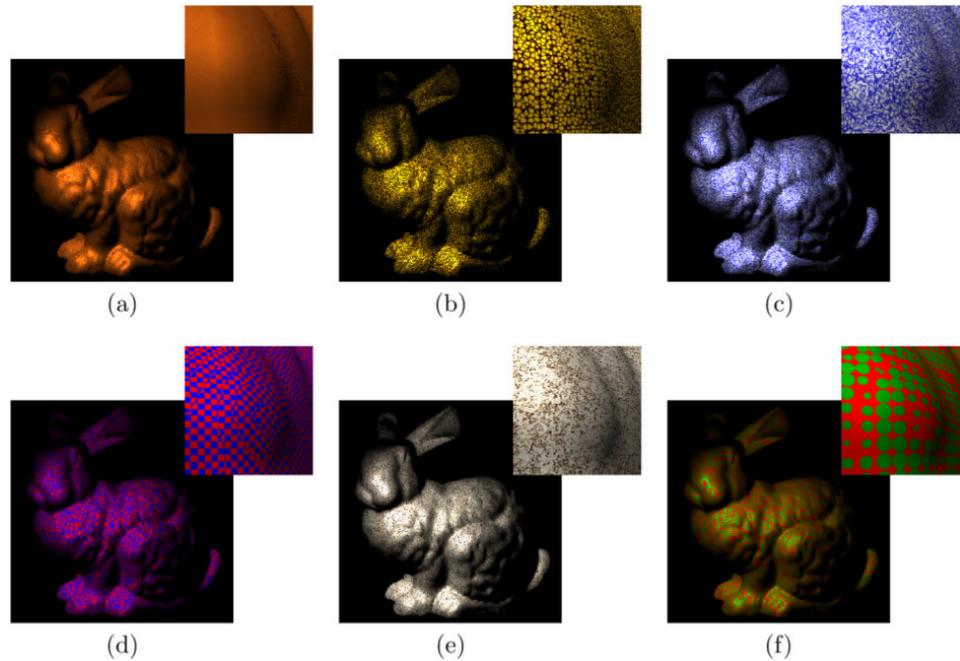


Figura 2.6: Resultados del método de generación de texturas sobre un objeto 3D utilizando patrones de (a) madera, (b) Voronoi, (c) ruido cúbico, (d) ajedrez, (e) mármol y (f) esferas.

de la iluminación [55]. La refracción del material translúcido puede introducir una aberración óptica a la visualización de la escena debido a la propiedad transmisiva de su superficie. En esta visualización ocurre en el espacio 3D dependiendo de la forma del objeto y el índice de refracción del material [56].

En la Fig. 2.5(a) se observa un objeto con material translúcido con un índice de refracción de 1.15 sobre un fondo de color gris. En la Fig. 2.5(b) se muestra el mismo objeto con otros objetos de color detrás. Se puede notar que a pesar de que no hay diferencias en las propiedades de la superficie del objeto, su apariencia cambia considerablemente, y la escena detrás se refracta a través de la superficie translúcida dependiendo de la forma, material, color y textura de los objetos que la componen.

Otra manera de estudiar los cambios de apariencia del objeto es mediante análisis de texturas. Para ésto, se utiliza comúnmente un mapeo de texturas para mejorar los

cambios de la visualización de los objetos. El proceso de añadir textura en una superficie 3D consiste en cubrir un objeto con una imagen en 2D que contiene un patrón específico. Las estrategias más comunes para introducir textura a un objeto son por cuadrículado y por estiramiento, las cuales son adecuadas para formas geométricas conocidas [42], tales como cubos, esferas o pirámides. Un enfoque diferente se lleva a cabo por la técnica de generación de texturas mediante formulaciones matemáticas (conocido en inglés por el término *procedural texturing*) que se calculan sobre la forma 3D del objeto, considerando características tales como el color y el sombreado [57,58]. La Fig. 2.6 muestra la textura de los objetos a través de la técnica de *procedural texturing* [59] la cual calcula diferentes patrones de textura utilizando una implementación paralela en GPU con NVIDIA Optix.

Otra estrategia algorítmica para modelar la iluminación es mediante el método de trazado de rayos de Monte Carlo (por sus siglas en inglés, MCRT). Este método es un proceso estocástico en el cual una gran cantidad de rayos de luz inciden en múltiples direcciones. La aproximación de la Ec. (2.13) se lleva a cabo mediante la integración numérica de la radiación incidente muestreada a través de diferentes puntos en el hemisferio. En cada punto, se traza un rayo con el fin de encontrar el punto de superficie en el cual se origina la luz. El número de muestras se incrementa exponencialmente con el número de reflexiones. En el método MCRT se eligen aleatoriamente las direcciones de las muestras.

Un caso particular del método de Monte Carlo es el MCPT (por sus siglas del inglés, Monte Carlo *path tracing*), el cual incluye rutas generales de luz, esto es, que los rayos reflejados se lanzan en todas direcciones en vez de seguir el camino trazado por la reflexión [43].

El método de Monte Carlo fue implementado para este propósito por primera vez por Cook et al. [60], en donde se promedian los resultados de una colección de rayos reflejados que son muestreados de manera uniforme en un cono incidente en

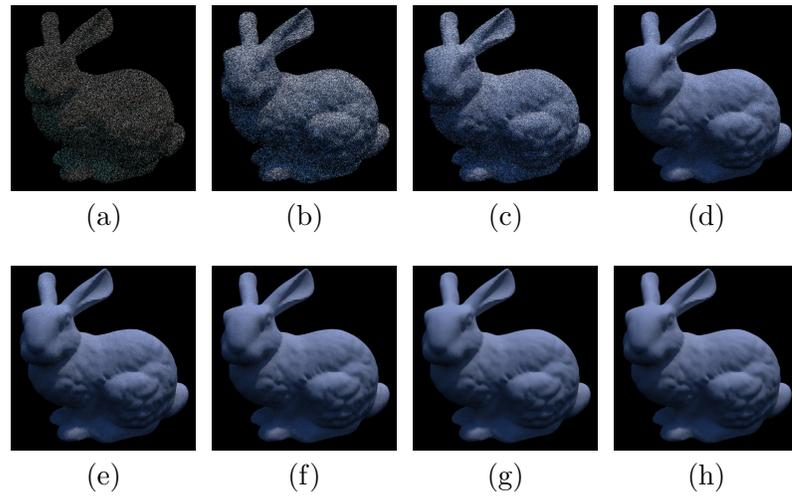


Figura 2.7: Ilustración del método de Monte Carlo implementado en GPU con NVIDIA Optix en función del tiempo. (a) 250 ms, (b) 750 ms, (c) 1 s, (d) 2 s, (e) 5 s, (f) 10 s, (g) 100 s, y (h) 1000s.

la dirección de la reflexión. El desempeño del método de Monte Carlo depende del número de muestras y de qué manera se eligen. La Fig. 2.7 ilustra los resultados obtenidos con una implementación en GPU [61] a diferentes instantes de tiempo.

Un enfoque diferente, conocido como trazado bidireccional de trayectorias [62], consiste en la inicialización de trayectorias del rayo de luz desde su fuente hasta un punto de observación. La propuesta de Henrik Wann Jensen [63], conocida como mapeo de fotones (en inglés, *photon mapping*), se desarrolló con el fin de trazar dichas trayectorias por separado. Este algoritmo opera emitiendo una vasta cantidad de puntos aleatorios, llamados fotones, por una fuente de iluminación. Los rayos de luz rebotan en toda la escena utilizando un método basado en trazado de rayos. Para cada punto de arribo de la luz, un fotón registra un mapa de la trayectoria recorrida. Por consiguiente, se calcula la intensidad de la luz que entra en cada píxel para estimar su densidad. El mapeo de fotones es un método recomendable para el cálculo de cáusticas [63], y es robusto en diferentes condiciones de la escena [64, 65].

Existen implementaciones en GPU del mapeo de fotones que generan resultados muy precisos [66]. En la Fig. 2.8 se ilustra una implementación en GPU de un mapeo

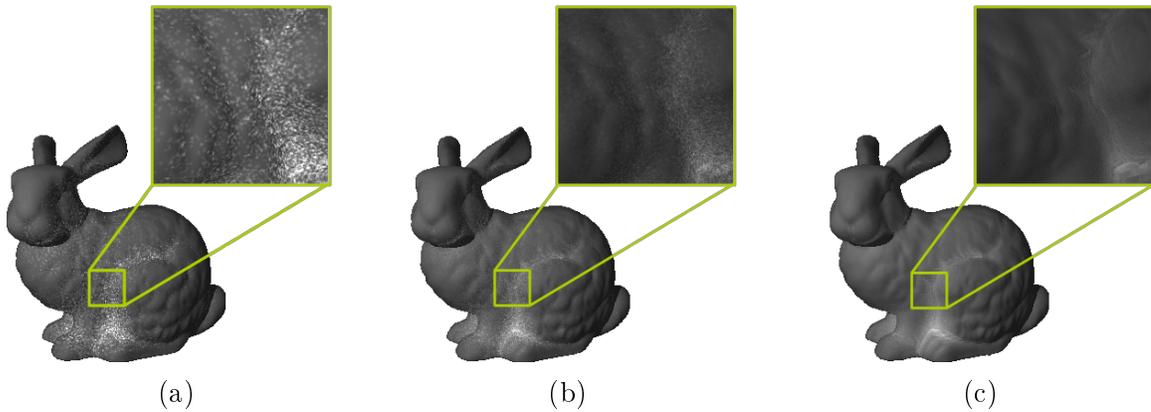


Figura 2.8: Resultados del algoritmo de mapeo de fotones con enfoque progresivo implementado en GPU con NVIDIA Optix, en los intervalos de tiempo de (a) 100 ms, (b) 1 s, y (c) 10 s.

de fotones progresivo utilizando diferentes tiempos de muestreo [67]. Primero, las partículas se lanzan desde la fuente incidente de luz, entonces las intersecciones de la superficie del objeto se registran en un mapa. Luego, este mapa se utiliza para estimar la intensidad de la luz en los puntos que tocan su superficie.

Al igual que el mapeo de fotones, existe otra técnica conocida como radiosidad instantánea, la cual utiliza de igual manera fotones y registra las múltiples trayectorias de luz dentro de la escena [68]. Una estructura de datos se registra con base en los puntos de intersección de estos fotones con la superficie, donde su densidad en cualquier posición es proporcional a la radiosidad en esa ubicación. Sin embargo, no utiliza la estimación de densidades. En cambio, utiliza un proceso de acumulación, en donde cada fotón se analiza como un punto virtual de luz (conocido en inglés como *virtual point light*, VPL), el cual se encarga de emitir luz en la escena [69]. Para calcular la luz indirecta en un píxel, es necesario iluminar cada VPL. Algunos enfoques basados en GPU permiten que cada VPL en el mapa se adapte conforme evoluciona en el tiempo [70], lo cual es una estrategia eficiente para gráficos en tiempo real.

Otra técnica útil para la generación de gráficos por computadora es el elemento finito. El elemento finito es un método clásico para aproximar soluciones en la propagación de la luz en una escena. Un método popular desarrollado para gráficos por computadora fue propuesto por Goral et al. [71]. La idea principal de este enfoque es discretizar varias superficies en la escena como elementos finitos, llamadas porciones (en inglés, *patches*). Posteriormente, se calcula la interacción de los rayos de luz. Este método tiene una complejidad computacional de $\mathcal{O}(n^2)$, donde n representa el número de porciones a calcular. Una aplicación reciente con GPU utiliza una geometría ya predefinida [72]. Este enfoque es adecuado para las escenas de baja complejidad y con sólo pocas distorsiones.

2.3 Filtros de correlación

Los filtros de correlación son un método estadístico para estimar la posición exacta de un objeto de interés en una imagen. En relación a la forma en que son diseñados, los filtros de correlación pueden dividirse en filtros analíticos y filtros compuestos [15]. Los filtros analíticos se diseñan optimizando diferentes criterios de desempeño, a partir del modelo de imagen de la escena.

La principal ventaja de estos filtros es que se representan mediante una expresión matemática de forma cerrada. Por el contrario, los filtros compuestos se diseñan a partir de un conjunto de imágenes de entrenamiento, las cuales deben ser suficientemente descriptivas y representativas de los patrones y sus variantes a ser reconocidos por el filtro, así como de los diferentes patrones que se desean rechazar [73]. El filtro de acoplamiento clásico fue el primer filtro de correlación propuesto [74], y a partir de éste, se han desarrollado un gran número de filtros de correlación para resolver diferentes situaciones dentro del reconocimiento de patrones. Comúnmente, las técnicas

de diseño están orientadas a proveer invariancia a distorsiones de la escena, precisión en la localización y robustez ante el ruido.

El filtro de acoplamiento clásico (CMF, *Classical Matched Filter*) optimiza la relación señal a ruido (SNR) para el modelo aditivo de la Ec. (2.1). El CMF tiene como objetivo generar como salida un valor de intensidad cercano a la unidad en las coordenadas (k, l) del objeto en la escena, y un valor cercano a cero, en cualquier otra parte fuera del área del objeto [15], como

$$c(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(k, l)h(x + k, y + l) dk dl = \begin{cases} 1, & \text{si } (k = x, l = y) \\ 0, & \text{otro caso,} \end{cases} \quad (2.16)$$

donde $t(x, y)$ es la imagen de referencia del objeto, $h(x, y)$ es la respuesta al impulso del filtro y $c(x, y)$ es el plano de correlación. Asumiendo que el valor esperado del ruido es $E\{n(x, y)\} = 0$, la SNR para un filtro de correlación, se expresa como

$$SNR = \frac{\left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T(\mu, \nu) H^*(\mu, \nu) d\mu d\nu \right|^2}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_n(\mu, \nu) |H(\mu, \nu)|^2 d\mu d\nu}, \quad (2.17)$$

donde $T(\mu, \nu)$ y $H(\mu, \nu)$ son la transformada de Fourier de $t(x, y)$ y $h(x, y)$, respectivamente, y $S_n(\mu, \nu)$ representa la densidad espectral de potencia del ruido $n(x, y)$. El numerador y el denominador de la Ec. (2.17) representan la intensidad del valor promedio del pico de correlación, y la varianza del valor del pico de correlación, respectivamente, tomados de un conjunto de muestras [15]. Es importante notar que mientras mayor sea el valor de la SNR, menor será la variación del valor del pico de correlación. Al maximizar la SNR en la Ec. (2.17), se obtiene la respuesta en

frecuencia del filtro de acoplamiento clásico, dada por [74]

$$H(\mu, \nu) = \alpha \frac{T^*(\mu, \nu)}{S_n(\mu, \nu)}, \quad (2.18)$$

donde α es un valor constante.

Otro diseño importante es conocido como filtro de funciones discriminantes sintéticas (SDF, *Synthetic Discriminant Functions*). Este diseño se utiliza para el reconocimiento de patrones multiclase [75]. Sea $\{t_i(x, y), i = 1, 2, \dots, N\}$ un conjunto linealmente independiente de imágenes de entrenamiento, cada una con M píxeles. La respuesta al impulso del filtro $h(x, y)$ se expresa en el dominio espacial como una combinación lineal de las imágenes de entrenamiento

$$h(x, y) = \sum_{i=1}^N a_i t_i(x, y), \quad (2.19)$$

donde $\{a_i, i = 1, 2, \dots, N\}$ son los coeficientes de ponderación que se eligen para satisfacer las siguientes condiciones

$$c_i(x, y) = \langle t_i(x, y), h(x, y) \rangle. \quad (2.20)$$

En la Ec. (2.20), $\{c_i, i = 1, 2, \dots, N\}$ representan los valores de la correlación (en el origen) asignados para cada imagen de entrenamiento [73], y “ $\langle \cdot \rangle$ ” denota la operación de producto punto. Sea \mathbf{R} una matriz con N columnas y M renglones, donde la i -ésima columna está dada por los elementos de la imagen $s_i(x, y)$ ordenados de forma lexicográfica. Sean \mathbf{a} y \mathbf{c} la representación de los vectores columna de los valores a_i y c_i , respectivamente. Las Ecs. (2.19) y (2.20) se pueden escribir en una notación matriz-vector, como

$$\mathbf{h} = \mathbf{R}\mathbf{a}, \quad (2.21)$$

y

$$\mathbf{c} = \mathbf{R}^+ \mathbf{h}, \quad (2.22)$$

donde "+" representa el transpuesto conjugado de la matriz. El (i, j) -ésimo elemento de la matriz $\mathbf{T} = (\mathbf{R}^+ \mathbf{R})$ es el valor en el origen de la correlación entre las imágenes de entrenamiento $t_i(x, y)$ y $t_j(x, y)$. Si la matriz \mathbf{T} es no singular, la solución del sistema de ecuaciones es $\mathbf{a} = (\mathbf{R}^+ \mathbf{R})^{-1} \mathbf{c}$, y la solución del filtro es

$$\mathbf{h} = \mathbf{R}(\mathbf{R}^+ \mathbf{R})^{-1} \mathbf{c}. \quad (2.23)$$

Se observa que el filtro SDF es capaz de reconocer distintas versiones de una clase de objeto [15, 76]. El vector \mathbf{c} es el valor deseable de la salida del pico de correlación, y puede utilizarse como

$$\mathbf{c} = [1, 1, \dots, 1]. \quad (2.24)$$

En la literatura podemos encontrar diversas soluciones para la detección de objetos con ruido no traslapado (ruido de fondo de la imagen) [77–79].

El filtro óptimo generalizado (GOF, *Generalized Optimum Filter*) ha demostrado un alto desempeño en la estimación de la localización del objeto en la escena [77]. En este diseño, se asume que el ruido en la escena está dado por

$$\tilde{n}(x, y) = \tilde{n}_b(x, y) + \tilde{n}_t(x, y), \quad (2.25)$$

donde $\tilde{n}_b(x, y)$ es el ruido del fondo, el cual se encuentra disjunto alrededor del objeto, y $\tilde{n}_t(x, y)$ es el ruido traslapado, el cual degrada directamente al objeto. En el modelo de señal disjunto, una imagen de entrada $f(x, y)$ contiene un objeto localizado en las coordenadas desconocidas (k, l) , sobre ruido no estacionario $\tilde{n}(x, y)$ que incluye el ruido del fondo y ruido traslapado con el objeto. El filtro GOF ha sido diseñado para optimizar la relación pico a energía de salida (POE, del inglés *Peak-to-Output Energy*

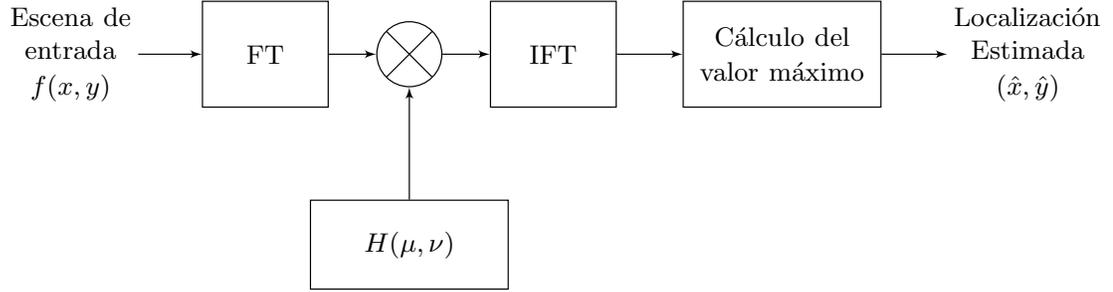


Figura 2.9: Diagrama a bloques de un sistema de reconocimiento de objetos mediante filtros de correlación.

ratio), definida por

$$POE = \frac{|E \{c(k, l)\}|^2}{E \{c(x, y; k, l)\}}, \quad (2.26)$$

donde $c(x, y; k, l)$ indica la salida del filtro en la posición donde se encuentra el objeto.

La función de transferencia del filtro GOF para el modelo disjunto está dada por [77]

$$H_{GOF}^*(\mu, \nu) = \frac{T_m(\mu, \nu)}{|T_m(\mu, \nu)|^2 + \frac{1}{2\pi} |W_1(\mu, \nu)|^2 * N_b^0(\mu, \nu) + \frac{1}{2\pi} |W_t(\mu, \nu)|^2 * N_t^0(\mu, \nu)}, \quad (2.27)$$

donde $T_m(\mu, \nu) = T(\mu, \nu) + m_b W_1(\mu, \nu) + m_t W_t(\mu, \nu)$, el cual está relacionado con los parámetros estadísticos del valor esperado del objeto m_t y del fondo m_b . Los términos $W_1(\mu, \nu)$ y $W_t(\mu, \nu)$ son la transformada de Fourier de $\bar{w}(x, y)$ y $w(x, y)$, respectivamente. Además, $N_b^0(\mu, \nu)$ y $N_t^0(\mu, \nu)$ representan la densidad espectral de potencia del ruido del fondo y ruido traslapado, respectivamente.

El filtro de acoplamiento generalizado (GMF, *Generalized Matched Filter*), ha demostrado ser robusto y preciso en la localización de un objeto dentro de una escena [16]. El filtro de correlación GMF maximiza el criterio de la relación señal a ruido (SNR), y su respuesta en frecuencia está dada por [80]

$$H_{GMF}^*(\mu, \nu) = \frac{T(\mu, \nu) + m_b W_1(\mu, \nu) + m_t W_t(\mu, \nu)}{\frac{1}{2\pi} |W_1(\mu, \nu)|^2 * N_b^0(\mu, \nu) + \frac{1}{2\pi} |W_t(\mu, \nu)|^2 * N_t^0(\mu, \nu)}. \quad (2.28)$$

La Fig. 2.9 ilustra el diagrama a bloques de un sistema de reconocimiento de objetos mediante el filtro GMF descrito en la Ec. (2.28), donde las coordenadas estimadas del objeto dentro de la escena, se obtienen como

$$(\hat{x}, \hat{y}) = \underset{x,y}{\operatorname{argmax}} \{|c(x, y)|^2\}. \quad (2.29)$$

La Ec. (2.28) permite la localización precisa de objetos ante condiciones de ruido aditivo y disjunto en la imagen.

2.4 Filtros de partículas

Los filtros de partículas son una técnica eficiente para estimar el espacio de estados de un objeto a partir de una colección de observaciones que entran al sistema de manera secuencial [81,82]. Los filtros de partículas son algoritmos secuenciales de Monte Carlo basados en poblaciones, en los cuales, se aproximan las distribuciones de los espacios de estado mediante la simulación de medidas aleatorias llamadas partículas [30].

En la Fig. 2.10 se muestra un filtro de partículas descrito por las etapas: *inicialización*, *evaluación*, *difusión* y *selección* [83]. Cada partícula se compone por un estado asociado a un coeficiente de peso [83]. Sea q_i^τ un conjunto de N partículas en un tiempo τ , donde la i -ésima partícula contiene información acerca del estado del sistema α_i^τ , y su valor de ponderación J_i^τ , descrito por

$$q_i^\tau = \{\alpha_i^\tau, J_i^\tau\} = \{(\alpha_0^\tau, J_0^\tau), (\alpha_1^\tau, J_1^\tau), \dots, (\alpha_{N-1}^\tau, J_{N-1}^\tau)\}. \quad (2.30)$$

La etapa de *inicialización* representa la entrada del filtro, donde el vector de estados α_i^τ está dado por los parámetros del sistema $\alpha_1, \alpha_2, \dots, \alpha_N$, como

$$\alpha_i^\tau = [\alpha_{1_i}^\tau, \alpha_{2_i}^\tau, \dots, \alpha_{N_i}^\tau]. \quad (2.31)$$

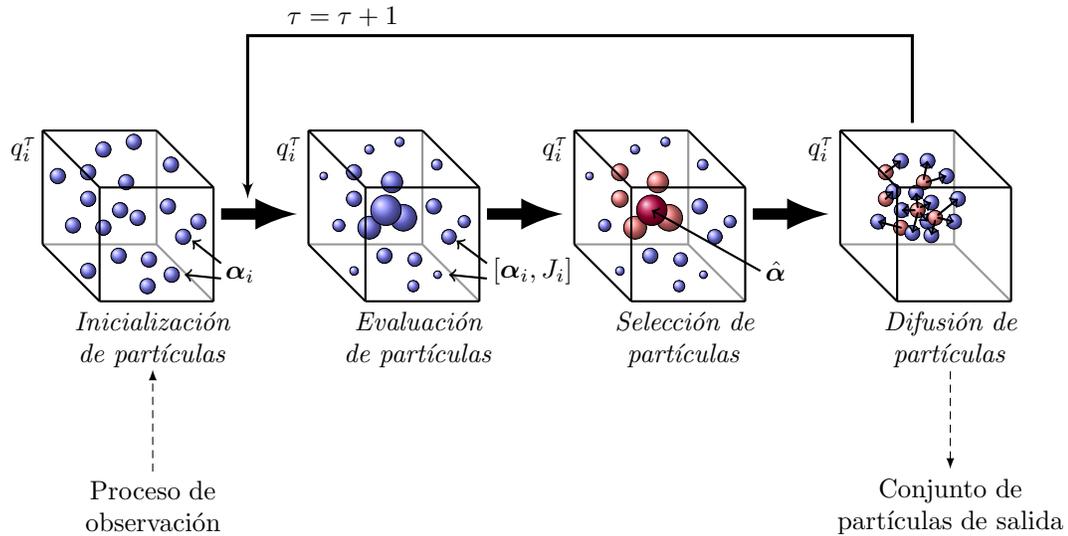


Figura 2.10: Diagrama a bloques de un filtro de partículas.

Después de la inicialización, las partículas se cuantifican mediante un valor de ponderación que es calculado en la etapa de *evaluación*. La idea principal es representar la función de densidad posterior por un conjunto de muestras aleatorias con sus pesos asociados, y así calcular las estimaciones basadas en dicha ponderación. Este valor de ponderación se calcula para cada partícula, y está dado por la función de aptitud

$$f(\alpha_i^\tau) = J_i^\tau. \quad (2.32)$$

La etapa de *selección* define el conjunto de partículas con los mayores valores de ponderación a partir de los resultados de los pesos J^τ en cada partícula. El conjunto de partículas resultante del proceso de selección son soluciones candidatas para la estimación final del estado del objeto. Entonces, dichas partículas se refinan en la etapa de *difusión* con respecto a su vector de estados α_i^τ , las cuales son propagadas (o desechadas) con base a su peso, que a su vez se determina mediante la función de aptitud [82].

El espacio de estados consiste en un proceso de observación $p(\mathbf{Z}^\tau|\boldsymbol{\alpha}^\tau)$, y en un proceso de transición $p(\boldsymbol{\alpha}^\tau|\boldsymbol{\alpha}^{\tau-1})$, donde \mathbf{Z}^τ y $\boldsymbol{\alpha}^\tau$ representan la observación actual y el vector de estados, respectivamente. El término $p(\cdot|\cdot)$ denota probabilidad condicional. Asumiendo que las observaciones arriban secuencialmente a través del tiempo. La probabilidad del estado actual $\boldsymbol{\alpha}^\tau$ a partir de la observación \mathbf{Z}^τ está dada en términos de la función de densidad de probabilidad *a posteriori* $p(\boldsymbol{\alpha}^\tau|\mathbf{Z}^\tau)$, dada por

$$p(\boldsymbol{\alpha}^\tau|\mathbf{Z}^\tau) = \frac{p(\mathbf{Z}^\tau|\boldsymbol{\alpha}^\tau)p(\boldsymbol{\alpha}^\tau|\mathbf{Z}^{\tau-1})}{p(\mathbf{Z}^\tau|\mathbf{Z}^{\tau-1})}. \quad (2.33)$$

La función $p(\boldsymbol{\alpha}^\tau|\mathbf{Z}^{\tau-1})$ se propaga en cada instante de tiempo [30, 83], mediante la ecuación de Chapman-Kolmogorov, dada por

$$p(\boldsymbol{\alpha}^\tau|\mathbf{Z}^{\tau-1}) = \int_{-\infty}^{\infty} p(\boldsymbol{\alpha}^\tau|\boldsymbol{\alpha}^{\tau-1})p(\boldsymbol{\alpha}^{\tau-1}|\mathbf{Z}^{\tau-1})d\boldsymbol{\alpha}^{\tau-1}. \quad (2.34)$$

Debido a que la solución analítica de la Ec. (2.34) puede llegar a ser bastante compleja, en la práctica los filtros de partículas se llevan a cabo a partir de una aproximación [30] de la solución de las Ecs. (2.33) y (2.34). Por lo tanto, los filtros de partículas se pueden considerar como algoritmos que se llevan a cabo mediante el desarrollo en el tiempo de soluciones recursivas. En aplicaciones de reconocimiento de objetos, la función de densidad condicional en el dominio discreto, puede representar la probabilidad de la pose actual de un objeto dentro de una escena, en términos de su localización y su orientación.

En este capítulo se han descrito los modelos de señal más importantes en el estudio del procesamiento de imágenes. Para ello, se han considerado diversas condiciones existentes en escenas capturadas en ambientes reales, tales como ruido aditivo, ruido disjunto, iluminación, entre otras. A partir de la representación matemática de la imagen de entrada, se logra caracterizar filtros a partir de la optimización de una

medida de desempeño específica (por ejemplo, la relación señal a ruido). Estos filtros son sistemas lineales, en cuya respuesta al impulso se procesa de manera tal que produce un valor de correlación alto a la salida cuando la apariencia del objeto en la escena coincide con un patrón de referencia. Se utiliza un filtro por cada vista del objeto que se desea detectar, por lo tanto, se construye un banco de filtros de correlación con el fin de obtener reconocimiento robusto a cambios de apariencia del objeto. Debido a que conocer todas las vistas posibles del objeto se convierte en un problema exhaustivo, en este trabajo de tesis se propone formular el problema de reconocimiento de pose como un problema de búsqueda. Para ello, se estudian los filtros de partículas, los cuales se permiten por medio de procedimientos probabilísticos apoyar la generación del banco de filtros con un enfoque adaptativo.

En el siguiente capítulo se plantea una propuesta algorítmica para la solución del reconocimiento de objetos mediante filtros de correlación adaptativos. En el método propuesto se consideran los modelos de imagen estudiados para analizar escenas con condiciones de ruido, fondo y cambios de iluminación incidente. De igual manera se plantea la propuesta de un algoritmo para la estimación y seguimiento de la pose 3D de un objeto en movimiento a partir de acoplamiento de plantillas en imágenes monoculares.

CAPÍTULO 3

Método propuesto para el reconocimiento de objetos con filtros de correlación adaptativos

En este capítulo se presenta la metodología desarrollada para resolver el reconocimiento de objetos en un espacio 3D mediante filtros de correlación adaptativos. El algoritmo propuesto plantea una solución eficiente para el reconocimiento de objetos 3D en imágenes con diferentes condiciones que degradan la escena de entrada. Las imágenes capturadas de escenarios reales pueden contener ruido aditivo, ruido disjuncto, cambios en iluminación, oclusiones parciales del objeto, y cambios de pose 3D en términos de la localización, orientación y escalamiento del objeto en la escena. De igual manera, en una secuencia de video, los objetos pueden aparecer en movimiento, cuyo desplazamiento ocurra en términos de cambios de la posición y la rotación con respecto al tiempo transcurrido en los cuadros de entrada.

La propuesta de este trabajo de tesis es utilizar las ventajas de los filtros de correlación ante condiciones de ruido y la precisión en la estimación de las coordenadas del objeto. Se genera un banco de filtros de correlación, en el cual, cada uno contiene

imágenes de referencia del objeto. Dichas imágenes de referencia son imágenes 2D capturadas con distintas poses 3D del objeto. Se utiliza un modelo digital 3D del objeto de interés, cuya información se encuentra en un arreglo tridimensional digital. El enfoque de la metodología propuesta se beneficia de la información digital previamente conocida para generar múltiples poses del objeto mediante gráficos por computadora.

Al utilizar los filtros de correlación adaptativos, se obtiene el acoplamiento entre el banco de filtros y la escena. Debido a que utilizar todas las versiones posibles del objeto provoca una cantidad masiva de procesamiento, en esta tesis se propone utilizar algoritmos de búsqueda para utilizar las ventajas ante la exploración de los parámetros de estado con mayor probabilidad de estimación.

El algoritmo propuesto es capaz de reconocer un objeto en una pose desconocida, a través del uso de filtros de correlación con un enfoque adaptativo. Para resolver el reconocimiento de objetos 3D, se ha planteado una metodología en las etapas descritas a continuación.

- Método para el reconocimiento de objetos 3D en escenas con condiciones de ruido, fondo, y cambios de iluminación incidente.
- Propuesta algorítmica para el reconocimiento de objetos utilizando filtros de correlación adaptativos mediante cómputo paralelo heterogéneo.
- Algoritmo para el reconocimiento de objetos 3D a partir de imágenes 2D mediante acoplamiento de plantillas y optimización de búsqueda local.
- Propuesta metodológica para la estimación y seguimiento de la pose 3D de un objeto dinámico, utilizando filtros de correlación adaptativos mediante filtros de partículas.

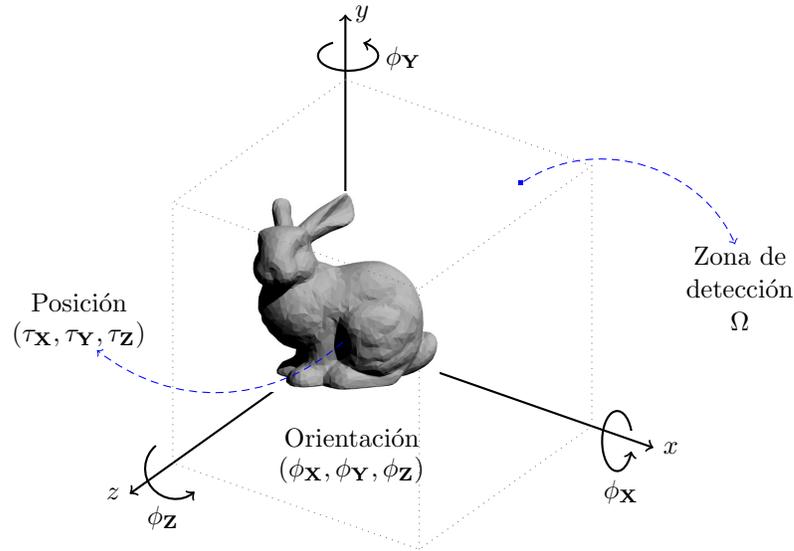


Figura 3.1: Escena definida dentro de una zona de detección Ω que contiene un objeto con posición y orientación desconocidas.

3.1 Método para el reconocimiento de objetos 3D en escenas con condiciones de ruido, fondo e iluminación incidente

Sea $F_\Omega(\mathbf{X})$ la representación de la escena observada definida en un espacio tridimensional $\Omega \subset \mathbb{R}^3$, donde $\mathbf{X} = [X, Y, Z]^T$ denota coordenadas cartesianas. Se asume que la escena contiene un objeto de interés $T_\Omega(\mathbf{X})$ incrustado en un fondo $B_\Omega(\mathbf{X})$ que se encuentra en una localización desconocida (τ_X, τ_Y, τ_Z) , tal como se muestra en la Fig. 3.1. El objeto presenta parámetros de pose desconocidos (ϕ_X, ϕ_Y, ϕ_Z) . La escena tridimensional se puede representar como

$$F_\Omega(\mathbf{X}) = T_\Omega(\mathbf{R}\mathbf{X} - \mathbf{D}) + B_\Omega(\mathbf{X})\overline{W}_\Omega(\mathbf{R}\mathbf{X} - \mathbf{D}), \quad (3.1)$$

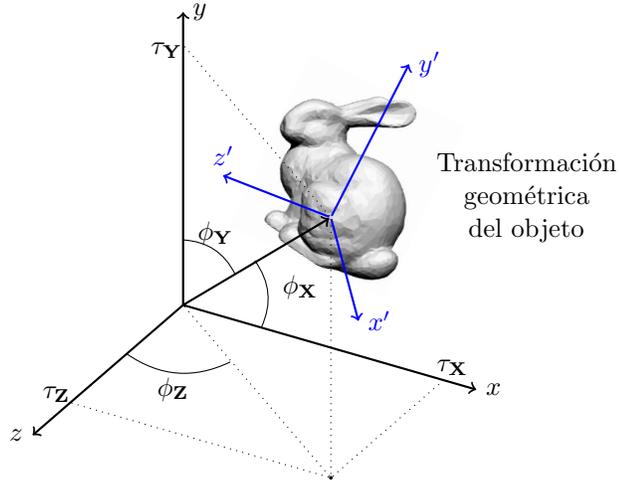


Figura 3.2: Transformación de la geometría del objeto.

donde $\mathbf{D} = [\tau_{\mathbf{X}}, \tau_{\mathbf{Y}}, \tau_{\mathbf{Z}}]^T$ es el vector de localización de las coordenadas del objeto, $\bar{W}_{\Omega}(\mathbf{X})$ es una función binaria definida con un valor de cero dentro del volumen del objeto y valor unitario en otro caso. Como se muestra en la Fig. 3.2, la geometría del objeto también puede modificarse con respecto a la orientación. El término \mathbf{R} representa una matriz de orientación $\mathbf{R} = \mathbf{R}_X \mathbf{R}_Y \mathbf{R}_Z$ compuesta por las matrices de rotación [84]

$$\mathbf{R}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi_X & \sin \phi_X \\ 0 & -\sin \phi_X & \cos \phi_X \end{bmatrix}, \quad (3.2)$$

$$\mathbf{R}_Y = \begin{bmatrix} \cos \phi_Y & 0 & -\sin \phi_Y \\ 0 & 1 & 0 \\ \sin \phi_Y & 0 & \cos \phi_Y \end{bmatrix}, \quad (3.3)$$

$$\mathbf{R}_Z = \begin{bmatrix} \cos \phi_Z & \sin \phi_Z & 0 \\ -\sin \phi_Z & \cos \phi_Z & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.4)$$

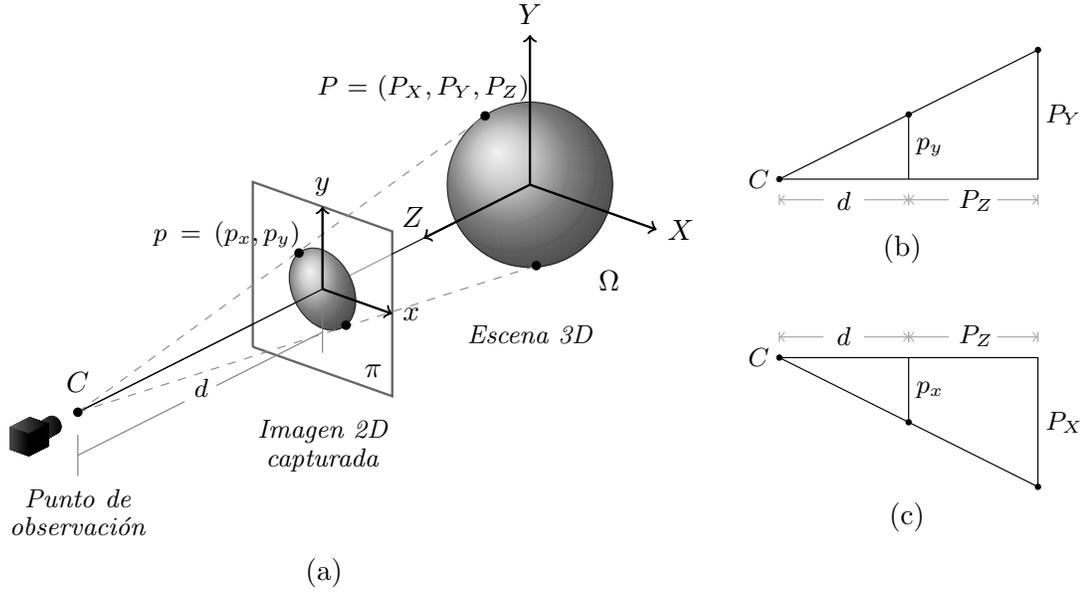


Figura 3.3: Representación geométrica de un mapeo de proyección de perspectiva.

Por lo tanto, la matriz de orientación \mathbf{R} está definida por

$$\mathbf{R} = \begin{bmatrix} c_{\phi_Y} c_{\phi_Z} & c_{\phi_Y} s_{\phi_Z} & -s_{\phi_Y} \\ c_{\phi_X} s_Y c_{\phi_Z} - c_{\phi_X} s_{\phi_Z} & s_{\phi_X} s_{\phi_Y} s_{\phi_Z} + c_{\phi_X} c_{\phi_Z} & s_{\phi_X} c_{\phi_Y} \\ -c_{\phi_X} s_{\phi_Y} c_{\phi_Z} + s_{\phi_X} s_{\phi_Z} & -c_{\phi_X} s_{\phi_Y} s_{\phi_Z} - s_{\phi_X} c_{\phi_Z} & c_{\phi_X} c_{\phi_Y} \end{bmatrix}, \quad (3.5)$$

donde $c_\phi = \cos \phi$ y $s_\phi = \sin \phi$. Se considera el modelo geométrico de la Fig. 3.3 en donde la escena 3D $F_\Omega(\mathbf{X})$ es observada mediante una cámara estenopeica¹ localizada en el punto C .

La imagen $f(x, y)$ capturada por la cámara consiste en un mapeo de proyección de la luz reflejada por la escena $F_\Omega(\mathbf{X})$ en el plano π de la imagen. Con respecto a la proyección de perspectiva [36], el plano de imagen observada $f(x, y)$ se puede

¹conocida comúnmente por su término en inglés: *pinhole camera*.

modelar como

$$f(x, y) = t(x - \tau_x, y - \tau_y; \phi_x, \phi_y, \phi_z) + b(x, y) \bar{w}(x - \tau_x, y - \tau_y; \phi_x, \phi_y, \phi_z) + n(x, y), \quad (3.6)$$

donde $p(x, y) \approx p(d_{\frac{X}{d+Z}}, d_{\frac{Y}{d+Z}})$ es un punto proyectado de $P(X, Y, Z)$, como se muestra en la Fig. 3.3(b) y (c) [13]. Los términos $t(x, y)$, $b(x, y)$, y $\bar{w}(x, y)$ representan las proyecciones de las funciones $T_\Omega(\mathbf{X})$, $B_\Omega(\mathbf{X})$, y $\bar{W}_\Omega(\mathbf{X})$, respectivamente en el plano π de la imagen a partir del mapeo de perspectiva. Se observa que una configuración particular de los parámetros de localización $(\tau_{\mathbf{X}}, \tau_{\mathbf{Y}}, \tau_{\mathbf{Z}})$ y orientación $(\phi_{\mathbf{X}}, \phi_{\mathbf{Y}}, \phi_{\mathbf{Z}})$ del objeto $T_\Omega(\mathbf{X})$ en $F_\Omega(\mathbf{X})$ produce una vista única de la escena en el plano de imagen π con parámetros (τ_x, τ_y) y (ϕ_x, ϕ_y, ϕ_z) definidos en la imagen capturada.

Una imagen capturada de una escena real puede representarse a través de la distribución de intensidades reflejadas por la superficie de los objetos en una escena dentro de un ambiente tridimensional. La apariencia del objeto en la escena no sólo depende de la pose en términos de la localización y orientación, sino que además es influenciada por fuentes de iluminación incidente. La interacción entre la luz y todos los elementos que constituyen la escena se caracteriza por la fuente de iluminación y el material de la superficie del objeto. La calidad del reconocimiento se logra al realizar cálculos precisos de las interreflexiones de la luz en la escena. La Fig. 3.4 muestra el esquemático representativo de la escena de entrada incluyendo una fuente de iluminación incidente.

Del escenario establecido en la Fig. 3.4 se puede asumir un modelo de señal multiplicativo en el cual la fuente incidente de iluminación se presenta en la escena de manera no uniforme. Dado un modelo multiplicativo, la imagen resultante es degradada por una función de iluminación $d(\mathbf{X})$ en un espacio tridimensional, representada por

$$f_d(\mathbf{X}) = f(\mathbf{X})d(\mathbf{X}), \quad (3.7)$$

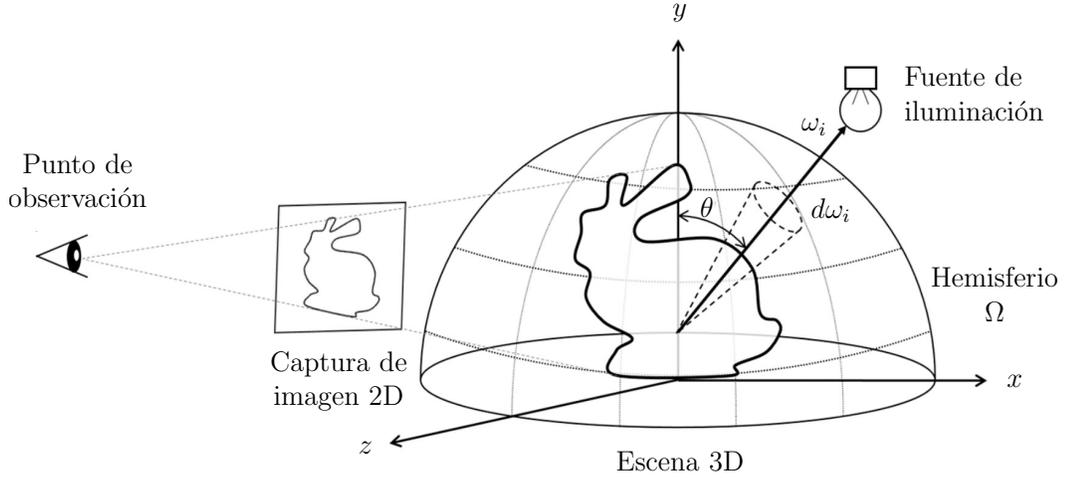


Figura 3.4: Proyección de un plano de imagen de una escena 3D.

donde $\mathbf{X} = [X, Y, Z]^T$ son las coordenadas en un espacio tridimensional, y $f(\mathbf{X})$ es la escena antes de la degradación por iluminación. Se asume que la escena $F_\Omega(\mathbf{X})$ dada en términos de su modelo geométrico de la Ec. (3.1) es una función sin degradación por iluminación $f(\mathbf{X})$, dada por la Ec. (3.7). A su vez, el modelo de iluminación influye en la escena de manera no uniforme en el dominio $\Omega \subset \mathbb{R}^3$. La función de iluminación se asume de manera determinística mediante la función bidireccional de reflectancia BRDF de la Ec. (2.14). La representación de una escena degradada por iluminación se considera como

$$F_L(\mathbf{X}) = F_\Omega(\mathbf{X})L(\mathbf{X}, \omega_r, \omega_i), \quad (3.8)$$

donde $F_L(\mathbf{X})$ es la escena degradada por iluminación, y $L(\mathbf{X}, \omega_r, \omega_i)$ es la función de iluminación expresada en términos de la dirección de la luz incidente ω_i y reflejada ω_r para cada posición en el espacio tridimensional. Si se conoce el material del objeto, se puede aproximar la función de iluminación $L(\mathbf{X}, \omega_r, \omega_i)$ a partir de los modelos de iluminación descritos por las Ecs. (2.5)–(2.9).

La proyección de una escena observada a una imagen 2D se obtiene mediante las Ecs. (3.6) y (3.8), la cual puede representarse por

$$\begin{aligned}
 f_l(x, y; \omega_r, \omega_i) = & \left(t(x - \tau_x, y - \tau_y; \phi_x, \phi_y, \phi_z) \right. \\
 & + b(x, y) \bar{w}(x - \tau_x, y - \tau_y; \phi_x, \phi_y, \phi_z) \\
 & \left. + n(x, y) \right) l(x, y; \omega_r, \omega_i),
 \end{aligned} \tag{3.9}$$

donde $l(x, y)$ es la función de iluminación proyectada en una imagen 2D. Debido a que el fondo y el ruido es información desconocida para el sistema de reconocimiento de objetos, se considera en el modelo de señal únicamente la función de iluminación del objeto dentro de la escena. Entonces, a partir de la Ec. (3.6) se tiene

$$\begin{aligned}
 f_l(x, y; \omega_r, \omega_i) = & t(x - \tau_x, y - \tau_y; \phi_x, \phi_y, \phi_z) l(x, y; \omega_r, \omega_i) \\
 & + b(x, y) \bar{w}(x - \tau_x, y - \tau_y; \phi_x, \phi_y, \phi_z) + n(x, y),
 \end{aligned} \tag{3.10}$$

donde la función de reflectancia es conocida si se utilizan modelos determinísticos de iluminación.

3.2 Algoritmo para el reconocimiento de objetos mediante cómputo paralelo heterogéneo

En esta tesis se ha desarrollado una propuesta algorítmica basada en computación paralela heterogénea con el fin de reducir el tiempo de ejecución computacional del algoritmo. El algoritmo paralelo se ha diseñado para su operación en arquitectura CPU/GPU utilizando el sistema operativo Linux, un procesador multinúcleo Intel i7 y dos procesadores gráficos NVIDIA GeForce GTX780. Con el fin de lograr un desempeño en tiempo real se utiliza una unidad de procesamiento gráfico (GPU) por su capacidad de cómputo y su habilidad de ejecutar tareas de manera concurrente

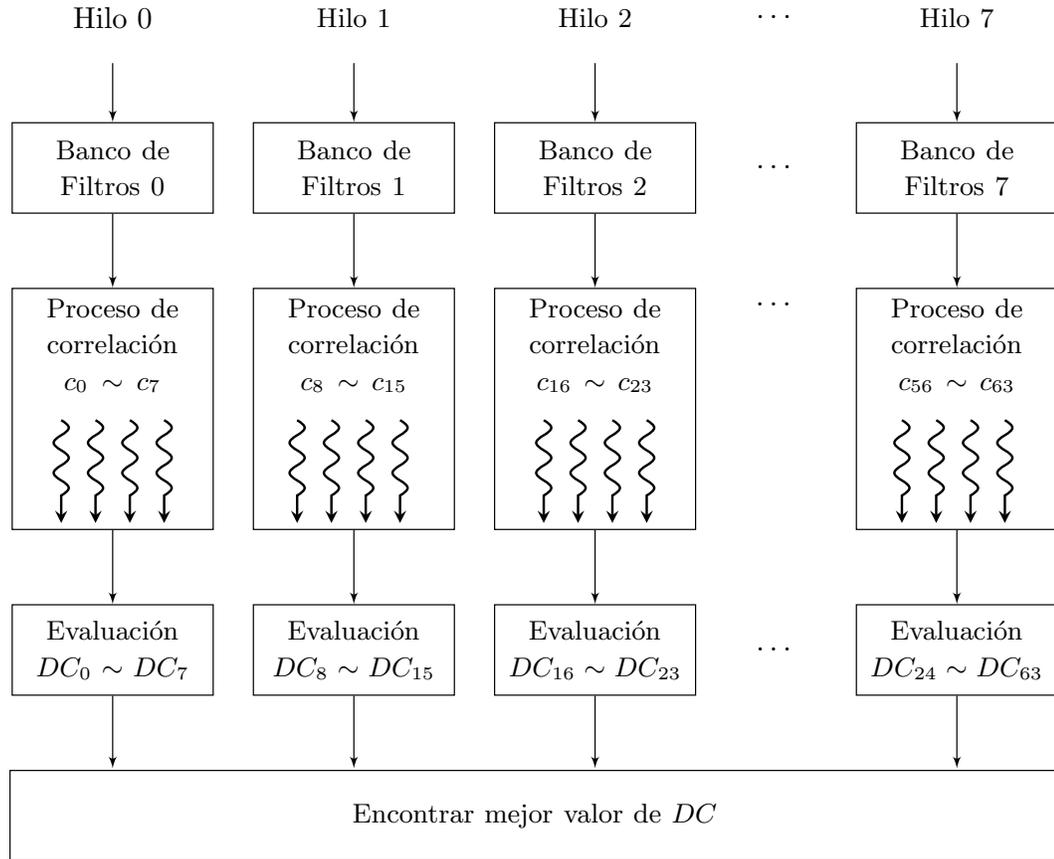


Figura 3.5: Diagrama a bloques de la implementación paralela con el algoritmo propuesto.

[85,86]. El código del algoritmo está escrito en lenguaje C/C++ con OpenCV para adquisición y procesamiento de imágenes, y OpenGL para la generación del modelo gráfico en 3D y su visualización.

El programa en paralelo fue diseñado en CUDA y se utilizó la interfaz de programación de aplicaciones (API, *application programming interface*) CUFFT, incluyendo OpenMP para la paralelización de hilos en CPU. La Fig. 3.5 ilustra el algoritmo de ejecución paralela con N hilos de CPU y M^3/N funciones (kernels) concurrentes de GPU.

El diseño del algoritmo propuesto se ha desarrollado con el fin de calcular un banco de M^3 filtros. Se ha diseñado para que cada hilo del CPU ejecute una porción definida (M^3/N) para el banco de filtros. Así, una cantidad de procesos de correlación se procesen concurrentemente en GPU. Entonces, cada hilo CPU calcula la métrica de DC obtenida de los planos de correlación. Finalmente, se obtiene el mejor valor de DC , el cual se utiliza para ponderar los parámetros de localización y orientación del objeto del mejor valor de acoplamiento.

3.3 Algoritmo para el reconocimiento de objetos 3D con filtros de correlación y optimización de búsqueda local

En esta sección, se describe una propuesta algorítmica para el reconocimiento de pose de un objeto en 3D. El algoritmo propuesto utiliza filtros de correlación que estiman la pose de un objeto 3D a partir de escenas monoculares. Se presenta una solución eficiente para la detección de un objeto 3D y la estimación de pose al procesar una imagen 2D a partir de una escena observada. A partir de la Ec. (3.6), se utiliza la técnica de acoplamiento de plantillas utilizando filtros de correlación para estimar los parámetros de pose del objeto en el dominio de la imagen 2D “ $\tau_x, \tau_y, \phi_x, \phi_y, \phi_z$ ”, y para luego recuperar los parámetros 3D “ $\tau_{\mathbf{X}}, \tau_{\mathbf{Y}}, \tau_{\mathbf{Z}}, \phi_{\mathbf{X}}, \phi_{\mathbf{Y}}, \phi_{\mathbf{Z}}$ ” utilizando de manera inversa el mapeo de proyección de perspectiva.

En el algoritmo propuesto se estiman con alta precisión los parámetros de pose en términos de la localización y la orientación de un objeto en una escena con fondo y ruido aditivo. Se implementa una etapa de refinamiento para aumentar la precisión de la estimación a partir de una optimización de búsqueda local. Para resolver el reconocimiento de pose del objeto, se plantea una metodología basada en cuatro

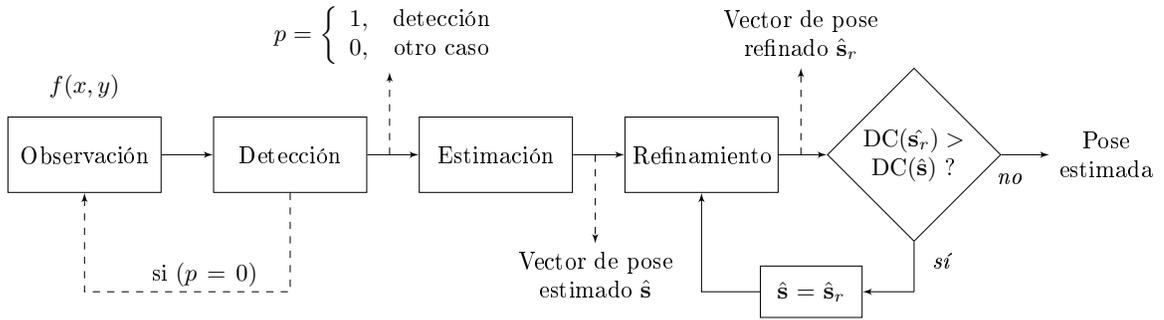


Figura 3.6: Diagrama a bloques del algoritmo propuesto con filtros de correlación adaptativos mediante búsqueda local.

etapas: *observación*, *detección*, *estimación*, y *refinamiento*. En la Fig. 3.6 se ilustra el diagrama a bloques del algoritmo propuesto, y sus etapas se explican a continuación.

Observación

En esta etapa se ejecuta la tarea de capturar imágenes de una escena utilizando una cámara monocular. El modelo de señal está caracterizado por la Fig. 3.3, donde se asume que la imagen capturada contiene un objeto $t(x, y)$ incrustado en un fondo $b(x, y)$ que se encuentra en una posición y orientación desconocida por el sistema, definido por la Ec. (3.6).

Detección

En la etapa de detección de objetos se procesa la imagen de entrada con un banco de filtros dinámicamente adaptativos. Estos filtros se diseñan con el fin de detectar el objeto en la escena. Cada filtro en el banco está asociado a una plantilla sintética generada mediante una imagen de referencia del objeto con parámetros de pose pre-determinados. Las plantillas son creadas con técnicas de gráficos por computadora a partir de un modelo digital de un objeto en 3D conocido. La Fig. 3.7 ilustra cómo han sido creadas las plantillas para la construcción del banco de filtros, en donde

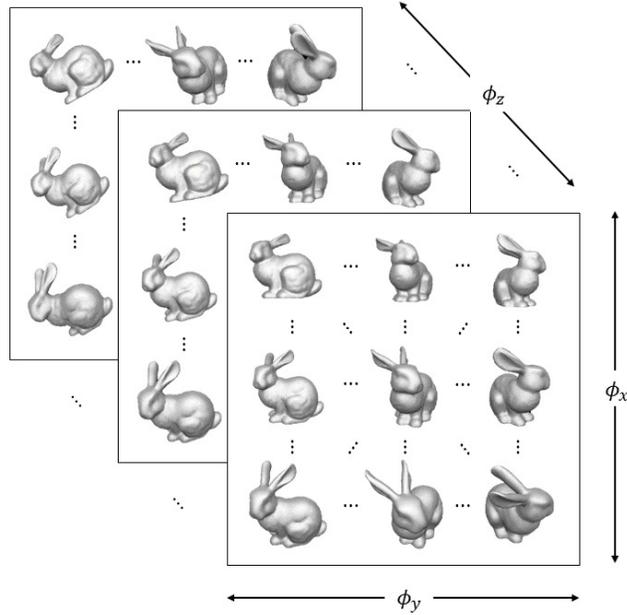


Figura 3.7: Ejemplo de la construcción de un banco filtros mediante plantillas sintéticas del objeto.

cada plantilla ha sido asociada con una configuración de pose $\mathbf{s} = [\tau_x, \tau_y, \phi_x, \phi_y, \phi_z]$ distinta. La configuración de pose consiste en valores numéricos de descriptores de localización y orientación del objeto en la escena.

El proceso de detección se lleva a cabo calculando la correlación cruzada entre el objeto observado en la escena y la respuesta al impulso de cada filtro en el banco $\{h_j(x, y); j = 1, \dots, N\}$. El proceso lineal de correlación se muestra en la Fig. 2.9. Sea $c_j(x, y) = f(x, y) \otimes h_j(x, y)$ el plano de correlación resultante producido por el j -ésimo filtro en el banco para la escena de entrada $f(x, y)$. El desempeño de los filtros está caracterizado mediante la métrica de capacidad de discriminación (DC). La DC es una medida que califica la habilidad de un filtro para reconocer un objeto de interés entre objetos falsos, la cual se define formalmente por [87]

$$DC = 1 - \frac{|c^b|^2}{|c^t|^2}, \quad (3.11)$$

donde c^t es el máximo valor de correlación del plano $c_j(x, y)$ en el área del objeto, y c^b es el valor del máximo pico producido en el área del fondo. Un valor de DC cercano a la unidad indica que el filtro utilizado presenta un buen valor de reconocimiento. En cambio, un valor de DC cercano a cero indica que el filtro no es capaz de detectar el objeto. Sea $\mathbf{DC}_{bk} = [DC_1, DC_2, \dots, DC_N]^T$ un vector que contiene todos los valores de DC producidos por cada uno de los filtros en el banco.

Sea p una variable de decisión binaria definida como $p = 1$ para una detección correcta del objeto, y $p = 0$ en otro caso. Esta variable de decisión se calcula como

$$p = \begin{cases} 1; & \text{si } \max \{\mathbf{DC}_{bk}\} > DC_{th}, \\ 0; & \text{otro caso} \end{cases}, \quad (3.12)$$

donde DC_{th} es un valor predeterminado que representa un umbral de calidad.

Estimación

El objetivo de esta etapa es calcular los parámetros que determinan la pose del objeto que ha sido detectado. Sea $h^*(x, y)$ la respuesta al impulso del filtro que ha detectado al objeto en la escena ($p = 1$). El plano de correlación producido por este filtro está dado por $c^*(x, y) = f(x, y) \otimes h^*(x, y)$, y las coordenadas estimadas de la localización del objeto en la escena $(\hat{\tau}_x, \hat{\tau}_y)$ se obtienen a través de la Ec. (2.29).

Se observa que cada filtro en el banco está asociado con una pose particular del objeto definida por un espacio de búsqueda de soluciones factibles Γ . La pose estimada del objeto $\mathbf{s}^* = [\hat{\tau}_x, \hat{\tau}_y, \hat{\phi}_x, \hat{\phi}_y, \hat{\phi}_z]$ está dada por los estados de pose asociados con el filtro $h^*(x, y)$ definida por

$$\hat{\mathbf{s}}^* = \left\{ (\mathbf{s} \in \Gamma) \sim h_{j^*}(x, y) : j^* = \arg \max_j \{\mathbf{DC}(j); j = 1, \dots, N\} \right\}. \quad (3.13)$$

Refinamiento

En esta etapa se ejecuta una operación de post-procesamiento con el fin de encontrar un óptimo local dado por un vecindario construido alrededor de la solución obtenida $\hat{\mathbf{s}}^*$. El refinamiento de la estimación obtenida se ejecuta mediante el método de búsqueda local. Sea $t^*(x, y)$ la plantilla sintética del objeto asociada con la solución actual $\hat{\mathbf{s}}^*$. El vecindario local construido se logra mediante un conjunto de imágenes de referencia $\{t_j(x, y); j = 1, \dots, N\}$ alrededor de la solución $\hat{\mathbf{s}}^*$ variando los parámetros de pose del vector $\hat{\mathbf{s}}^*$. Para esto, se utilizan N realizaciones de una distribución Gaussiana $\mathcal{N}(\hat{\phi}, r)$, donde $\hat{\phi}$ representa un solo parámetro de pose definido en $\hat{\mathbf{s}}^*$ y $r = (1 - DC(j^*))$.

Según el procedimiento descrito en el Algoritmo 1, por cada plantilla del vecindario se construye un filtro GMF, y se forma un nuevo banco de filtros. Luego, la escena observada $f(x, y)$ se correlaciona con cada filtro en el nuevo banco, y se elige el filtro que produce el mayor valor de DC, el cual está asociado con la pose $\hat{\mathbf{s}}_r^*$. Si el nuevo valor de DC es mayor con respecto al valor de DC producido por la solución inicial $\hat{\mathbf{s}}^*$, el proceso de refinamiento se repite. En caso contrario, indica que no se encuentra solución mejor que la actual y el proceso de refinamiento termina.

3.4 Algoritmo para la estimación y seguimiento de pose 3D con filtros de correlación adaptativos y filtros de partículas

En esta sección, se describe la metodología propuesta para la estimación y seguimiento de pose en objetos tridimensionales. Para una secuencia de video dada, el algoritmo

Algoritmo 1 Proceso de refinamiento del algoritmo propuesto para el reconocimiento de pose 3D.

```

procedure REFINAMIENTOLOCAL( $\mathbf{s}, r(\mathbf{s})$ )
   $i \leftarrow 0$ 
  repeat
     $\mathbf{N}_s \leftarrow \text{crearVecindario}(\mathbf{s}, r(\mathbf{s}))$ 
     $[\mathbf{s}', r(\mathbf{s}')] \leftarrow \min_j(\text{evaluarBancoDeFiltros}(\mathbf{N}_s))$ 
    if  $r(\mathbf{s}') \geq r(\mathbf{s})$  then
       $r(\mathbf{s}) \leftarrow r(\mathbf{s}')$ 
       $\mathbf{s} \leftarrow \mathbf{s}'$ 
    end if
     $i \leftarrow i + 1$ 
  until  $i > N$ 
end procedure

```

estima la pose de un objeto, cuyos parámetros de localización, escalamiento y orientación son desconocidos. El algoritmo obtiene alta precisión en la estimación del reconocimiento de pose, y robustez en presencia de ruido, fondo desordenado y fuentes incidentes de iluminación. El método propuesto incluye una estrategia basada en el acoplamiento de plantillas mediante filtros de correlación adaptativos guiados por filtros de partículas.

En esta propuesta, se diseña un banco de filtros de correlación mediante un conjunto de imágenes de referencia. La selección de esas imágenes de referencia es guiada por medio de filtros de partículas, la cual debe adaptarse a los cambios de los parámetros dinámicos del objeto. La estimación actual de la pose se mejora mediante una etapa de refinamiento utilizando un filtro de partículas y un enfoque de búsqueda local. La pose final estimada del objeto se encuentra en términos de la localización, la orientación y la escala.

En la Fig. 3.8 se muestra un diagrama a bloques del algoritmo propuesto, el cual está basado en cinco etapas básicas: *observación*, *detección*, *evaluación*, *refinamiento*, y *predicción*. Estas etapas son descritas a continuación.

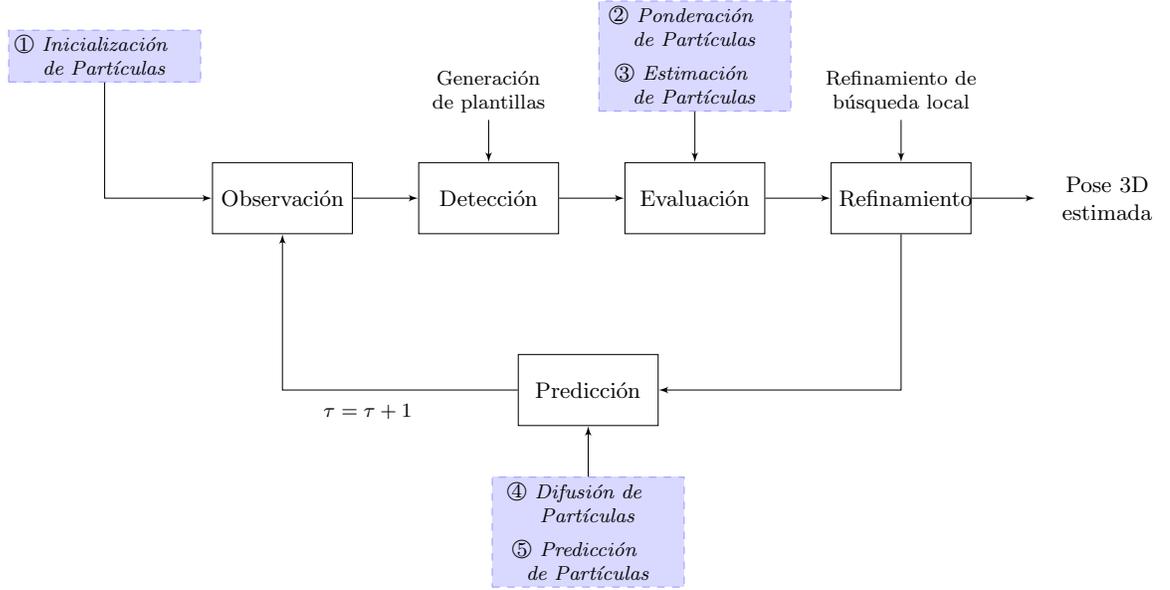


Figura 3.8: Algoritmo propuesto para el reconocimiento y seguimiento de pose de un objeto 3D mediante filtros de correlación adaptativos guiados por filtros de partículas.

Observación

En esta etapa se captura una escena de entrada a partir de una secuencia de video capturada por una cámara monocular. El cuadro de entrada $f^\tau(\mathbf{x})$ está caracterizado por el modelo de señal

$$f^\tau(\mathbf{x}) = t(\mathbf{\Gamma}\mathbf{x}) + b(\mathbf{x})\bar{w}(\mathbf{\Gamma}\mathbf{x}) + \eta(\mathbf{x}), \quad (3.14)$$

donde $\mathbf{x} = [x, y]^T$ denota las coordenadas cartesianas de localización, y $\mathbf{\Gamma} = \mathbf{RTS}$ es una matriz de transformación [18], compuesta por rotación \mathbf{R} , translación \mathbf{T} y escalamiento \mathbf{S} [36]. Como se muestra en la Fig. 3.9 en el modelo de señal, el objeto aparece en una escena con fondo, ruido aditivo, y con una pose desconocida con respecto a los parámetros de localización, orientación y escalamiento.

En esta etapa, el filtro de partículas inicia con una población aleatoria con N partículas (ver Fig: 3.10: *Inicialización de partículas*). Esta población inicial se

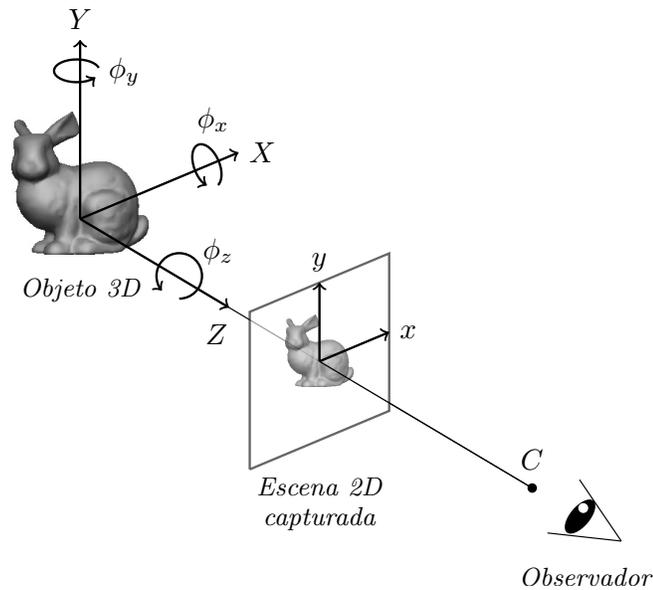


Figura 3.9: Modelo de señal representado para una escena 3D y los parámetros de pose del objeto.

compone de un conjunto de soporte α_i^T a partir de una distribución uniforme. Cada partícula representa una posible solución de configuración de pose del objeto.

Detección

En esta etapa, se procesa la escena de entrada (en el cuadro τ) con un banco de filtros con el fin de detectar un objeto de interés en la escena. El banco contiene un conjunto de filtros de correlación, en el cual, cada uno está asociado con una plantilla sintética $t_j(\mathbf{x})$ representada con el parámetro de pose específico del objeto. El proceso de detección se procesa cuadro por cuadro de la escena de entrada, como se describe enseguida.

Primero, se construye un conjunto de plantillas con diferentes vistas del objeto, como se muestra en la Fig. 3.11. Estas plantillas se generan con técnicas de gráficos por computadora a partir de un modelo 3D sintético de referencia. Cada plantilla se

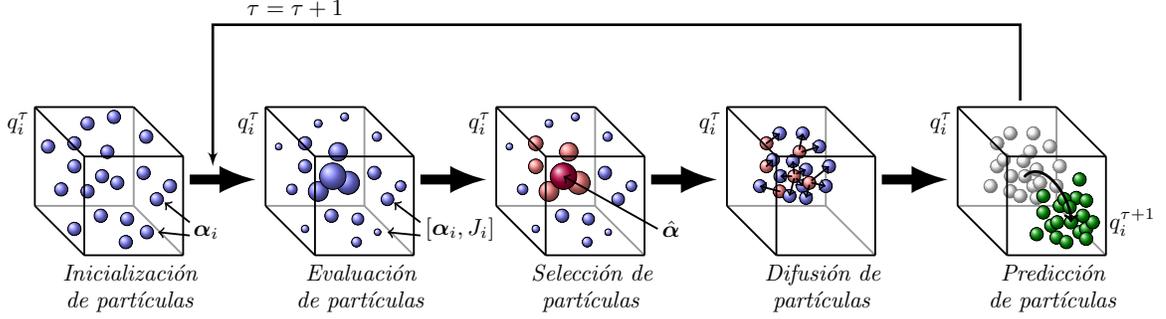


Figura 3.10: Diagrama a bloques propuesto para un filtro de partículas con un enfoque de predicción dinámica.

encuentra asociada con una configuración de pose única $\alpha_i^\tau = [x_i^\tau, y_i^\tau, \phi_{x_i}^\tau, \phi_{y_i}^\tau, \phi_{z_i}^\tau, s_i^\tau]$, que consiste en los descriptores de localización, orientación y escalamiento del objeto.

Posteriormente, se construye un conjunto de filtros de correlación a partir de las plantillas sintéticas $t_j(\mathbf{x})$. Como se describe en la Ec. (2.28), cada filtro de correlación en el banco $h_j(\mathbf{x}) = IFT\{H_j(\boldsymbol{\mu})\}$ se genera por un filtro GMF en el dominio de la frecuencia. Entonces, el proceso de detección se lleva a cabo al calcular la correlación cruzada entre la imagen observada $f(\mathbf{x})$ y la respuesta al impulso de cada filtro de correlación GMF en el banco $\{h_j(\mathbf{x}); j = 0, \dots, N - 1\}$. Sea $c_j(\mathbf{x})$ la correlación lineal entre la escena y el j -ésimo filtro en el banco, descrito por

$$c_j(\mathbf{x}) = f(\mathbf{x}) \otimes h_j(\mathbf{x}), \quad (3.15)$$

donde se produce un plano de correlación de salida $c(\mathbf{x})$ y la estimación de la localización del objeto en la escena $\hat{\mathbf{x}} = (\hat{x}, \hat{y})$, se obtiene como a continuación:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}} \{|c(\mathbf{x})|^2\}. \quad (3.16)$$

Por último, el proceso de detección está relacionado con la variable de decisión binaria p , valuada como $p = 1$ para una detección del objeto y $p = 0$ para cualquier

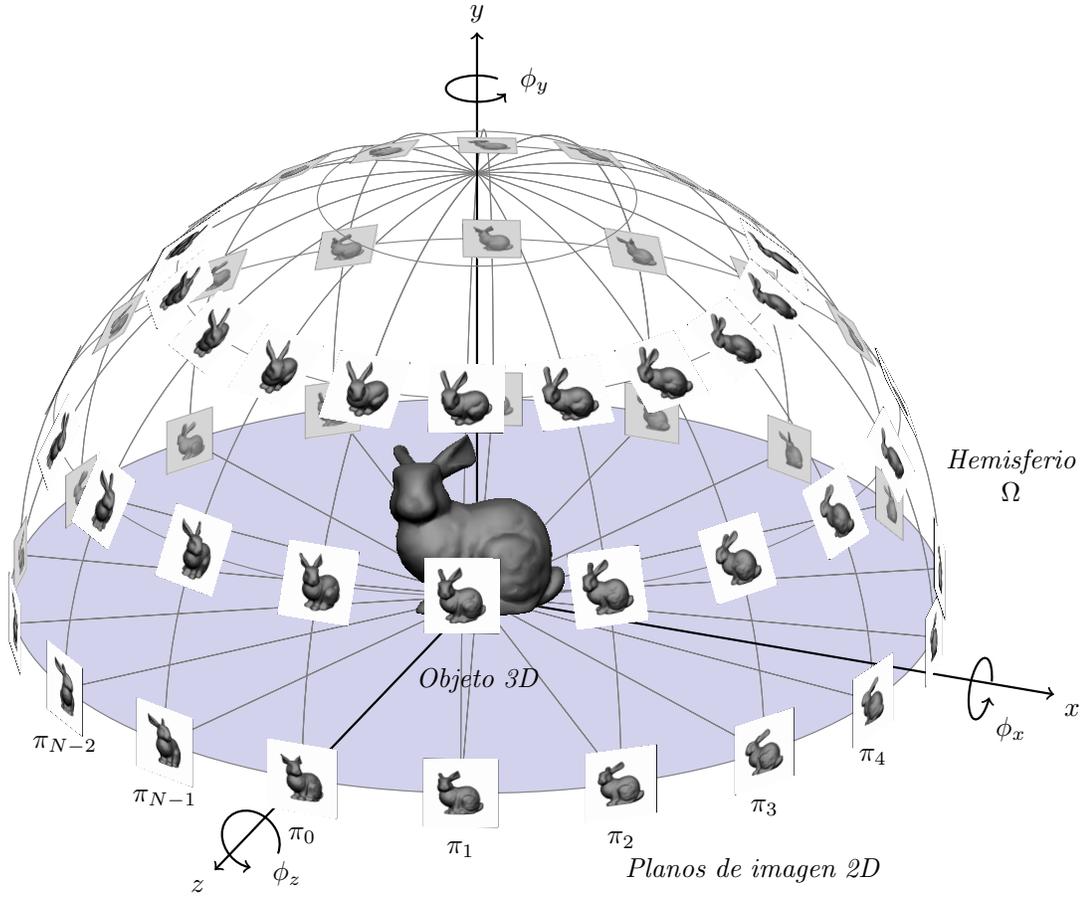


Figura 3.11: Generación de imágenes de referencia para diferentes vistas del objeto.

otro caso, de la siguiente manera

$$p = \begin{cases} 1, & \text{si } \max_j \{DC_j; j = 0, \dots, N - 1\} > \delta_p \\ 0, & \text{otro caso,} \end{cases} \quad (3.17)$$

donde δ_p es un valor de umbral predeterminado de la métrica DC para la evaluación de calidad del filtro.

Evaluación

La evaluación de cada filtro de correlación consiste en dos pasos: *evaluación de partículas* y *estimación de pose*. El proceso de evaluación de partículas se muestra en la Fig. 3.10, y define los parámetros de pose del objeto al encontrar el filtro GMF que produzca el mayor valor de acoplamiento cuando se correlaciona con la escena de entrada.

Como se describe en la Ec. (3.11), la calidad del acoplamiento de plantillas se caracteriza por la capacidad de discriminación (DC). Existen varios factores que deterioran la calidad de un filtro de correlación [16]. Por ejemplo, el ruido aditivo, el ruido disjunto, iluminación no uniforme, y cambios de apariencia del objeto debido a transformaciones geométricas [88]. Cuando existe una combinación particular de estos factores, el valor de DC se atenúa, aún cuando el filtro de correlación contiene una plantilla con parámetros muy cercanos al óptimo. Para solucionar esas dificultades, se propone utilizar una métrica complementaria a la DC del filtro. Esta métrica está dada por la intersección de la región de soporte del objeto real en la escena y del área de soporte estimada. La región de soporte estimada del objeto se representa por la región de soporte de la plantilla sintética utilizada para la construcción del filtro. El complemento del área de intersección I entre la región de soporte del objeto en la escena $w_f(\mathbf{x})$ y la región de soporte del objeto en la plantilla $w_t(\mathbf{x})$, se calcula mediante la siguiente expresión:

$$I = \frac{\sum_{\mathbf{x}} \overline{w_t(\mathbf{x}) \cap w_f(\mathbf{x})}}{\sum_{\mathbf{x}} w_f(\mathbf{x})}, \quad (3.18)$$

donde d es el número de píxeles contenidos en la región de soporte $w_f(\mathbf{x})$. Así, cada partícula es evaluada en términos de dos criterios diferentes: calidad del reconocimiento y desempeño de la intersección de la región de soporte.

La calidad del reconocimiento se cuantifica a través de la DC. A partir de la Ec. (3.18), el área de no intersección indica buenos valores de I cuando se acerca a cero. Entonces, se determina una solución óptima en el sentido de Pareto [89], la cual se caracteriza por un conjunto de soluciones no dominadas por ambos criterios. Se propone manejar este problema multi-objetivo en términos de una función mono-objetivo, representada por la combinación de ambos criterios. De esta forma, el valor óptimo de la función mono-objetivo entrega una buena solución con respecto al frente de Pareto, dado por

$$J = \frac{DC}{1 + I}, \quad (3.19)$$

donde J indica de igual manera, la ponderación de cada partícula representada en la Ec. (2.32).

En el proceso de estimación, el resultado del proceso de evaluación está dado por el máximo valor de J_i^τ obtenido por la Ec. (3.19). La evaluación de cada partícula se calcula mediante el conjunto de filtros GMF en el banco, y que a su vez, pertenece a los parámetros de las plantillas correspondientes $t_i^\tau(\mathbf{x}; \boldsymbol{\alpha}_i^\tau)$. La estimación de la mejor partícula está dada por

$$\hat{\boldsymbol{\alpha}}^\tau = \arg \max_{\boldsymbol{\alpha}_i^\tau} \{J_i^\tau; \forall \boldsymbol{\alpha}_i^\tau, i \in [0, \dots, N - 1]\}, \quad (3.20)$$

donde N es el número de partículas, y el estado $\hat{\boldsymbol{\alpha}}^\tau$ está caracterizado por

$$\hat{\boldsymbol{\alpha}}^\tau = \left[\hat{x}^\tau, \hat{y}^\tau, \hat{\phi}_x^\tau, \hat{\phi}_y^\tau, \hat{\phi}_z^\tau, \hat{s}^\tau \right], \quad (3.21)$$

donde equivale a la pose estimada en términos de localización, orientación y escalamiento.

Refinamiento

En la etapa de refinamiento, se introduce un algoritmo de filtrado de partículas con búsqueda local (LSPF, del inglés *Local Search Particle Filter*). El algoritmo LSPF obtiene una alta calidad en la estimación de pose del objeto utilizando un enfoque adaptativo. En este procedimiento, se selecciona la mejor partícula $\hat{\alpha}^\tau$ y a su vez, se selecciona el vecindario construido con los valores que definen la solución.

Se construye un nuevo conjunto de muestras a partir del vecindario, el cual está relacionado a cada parámetro de estado de pose del objeto α_i^τ , con un tamaño de N partículas. El enfoque de búsqueda local se encarga de mejorar la solución de la estimación [83, 90]. Para ésto, la mejor partícula obtenida se refina mediante iteraciones del procedimiento de búsqueda local, en la cual la búsqueda finaliza cuando se encuentra una mejor solución que la solución obtenida. El criterio aplicado para el algoritmo LSPF es obtener una calidad de desempeño de $\delta_p > 90\%$ en términos de la estimación de los parámetros de pose del objeto.

Predicción

Para predecir los estados de pose en los cuadros sucesivos, se realizan dos procedimientos: *difusión de partículas* y *predicción de estados*.

En la difusión de partículas se selecciona un subconjunto de partículas utilizando el método de selección por elitismo, en el cual, las mejores partículas se conservan (ver Fig. 3.10: *Difusión de partículas*). Luego, las partículas actuales se propagan por un proceso de distribución normal, formando nuevas partículas en el conjunto para el siguiente intervalo del cuadro. Las nuevas partículas en el instante $\tau + 1$ reemplazan a las actuales en el instante τ , con el fin de entregar diversidad a la solución.

En el proceso de predicción se representa el comportamiento del movimiento del objeto en un espacio tridimensional, caracterizado por el modelo cinemático de cuerpos rígidos [91],

$$\phi^{\tau+1} = \phi^{\tau} + \omega_{\phi} \Delta_{\tau} + \frac{1}{2} a_{\phi} \Delta_{\tau}^2, \quad (3.22)$$

donde ω_{ϕ} es la velocidad angular en el instante de tiempo Δ_{τ} , y a_{ϕ} es una variable aleatoria que representa los componentes de aceleración, tales como turbulencia y fricción [91]. Como se muestra en la Fig. 3.10: *Predicción de partículas*, las nuevas partículas se generan con una distribución normal $\mathcal{N}(\mu, \sigma)$, donde en este caso, el valor esperado μ está dado por el parámetro α_i^{τ} , y la desviación $\sigma = (1 - \hat{J}^{\tau})$ depende directamente de la calidad de la estimación de la mejor partícula dada por la Ec. (3.20).

En este capítulo se ha descrito la metodología propuesta para resolver el reconocimiento de objetos 3D mediante filtros de correlación adaptativos. Esta propuesta se realiza en cuatro etapas. Primero, se describe un algoritmo para el reconocimiento de objetos en condiciones de ruido, fondo y cambios de iluminación. Además de una propuesta utilizando filtros de correlación procesados mediante cómputo paralelo heterogéneo. Seguido de un algoritmo para el reconocimiento de objetos 3D a partir de imágenes monoculares mediante acoplamiento de plantillas y búsqueda local. Por último, se desarrolla una propuesta metodológica para la estimación y seguimiento de pose de un objeto en movimiento en un espacio tridimensional, utilizando filtros de correlación adaptativos mediante filtros de partículas.

En el siguiente capítulo se describen los resultados obtenidos de esta propuesta metodológica en una implementación computacional y se discuten las ventajas de utilizar los filtros de correlación ante condiciones de ruido y precisión en la localización del objeto con un enfoque adaptativo.

CAPÍTULO 4

Resultados

En este capítulo se presentan y se discuten los resultados experimentales obtenidos con el algoritmo propuesto para la estimación de pose de un objeto tridimensional. Los experimentos fueron desarrollados para evaluar diferentes aspectos del reconocimiento de objetos 3D, los cuales se describen a continuación.

- Desempeño del reconocimiento de objetos en escenas con cambios de iluminación.
- Desempeño computacional del reconocimiento de objetos con filtros de correlación adaptativos utilizando computación paralela heterogénea.
- Desempeño en el reconocimiento de pose 3D a partir de imágenes monoculares utilizando filtrado por acoplamiento de plantillas.
- Desempeño en el seguimiento de pose 3D de un objeto móvil, utilizando filtros de correlación guiados por filtros de partículas.

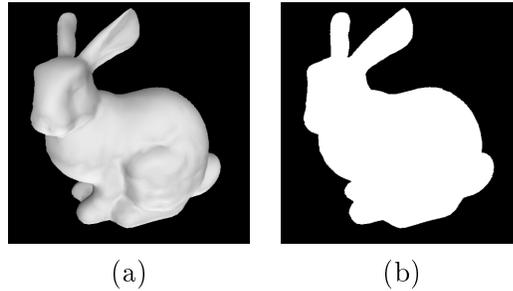
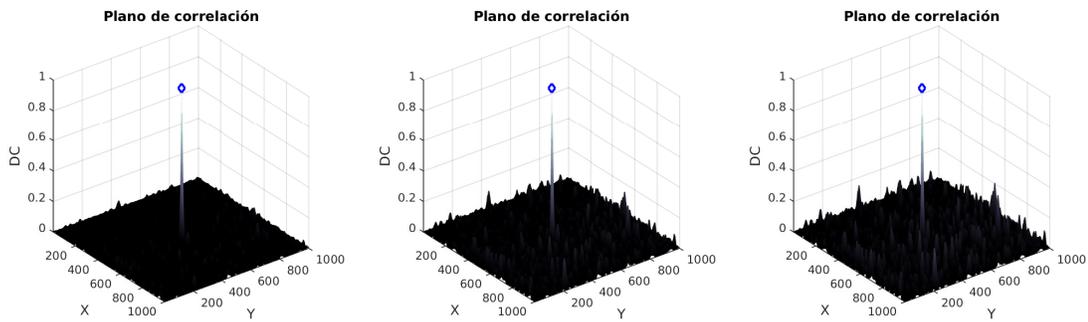


Figura 4.1: Objeto de referencia con (a) material difuso y (b) su región de soporte.

4.1 Evaluación del algoritmo con distintos modelos de iluminación

Esta sección presenta la evaluación de desempeño del algoritmo de reconocimiento de objetos con distintos materiales de superficie. Los experimentos se han realizado con el fin de evaluar la tolerancia del algoritmo propuesto ante cambios de apariencia del objeto ocasionados por iluminación. El algoritmo toma como referencia un objeto con material difuso, tal como se muestra en la Fig. 4.1(a). El algoritmo evalúa diferentes materiales en la superficie del objeto, basado en los modelos difuso (Lambert), reflectivos (Phong, Blinn-Phong) y reflectivos metálicos (Ward, Cook-Torrance).

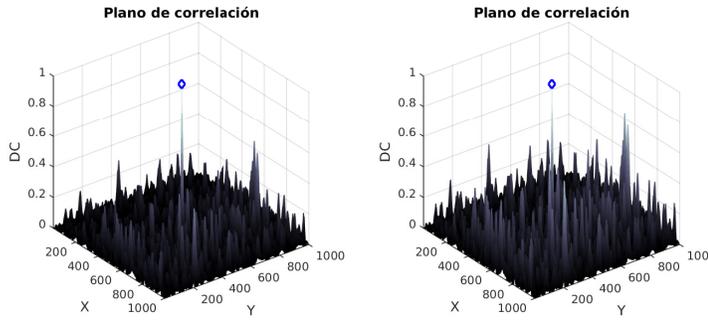
En la Fig. 4.2 se muestran las escenas de prueba que consisten en un objeto con cada uno de los materiales analizados. En la escena, el objeto aparece incrustado en una imagen de fondo en una posición arbitraria. El experimento consiste en evaluar la calidad del reconocimiento de objetos ante cambios de apariencia del objeto debido a la iluminación. El plano de correlación se obtiene mediante un filtro diseñado para optimizar la estimación de la localización del objeto. La calidad del reconocimiento se cuantifica a través de la capacidad de discriminación (DC), obtenida del plano de correlación a la salida del filtro. Un valor de DC cercano a la unidad indica una buena calidad de reconocimiento, mientras que valores negativos de DC indican que el filtro es incapaz de reconocer el objeto.



(a) $DC = 0.9904$

(b) $DC = 0.9546$

(c) $DC = 0.9212$



(d) $DC = 0.7182$

(e) $DC = 0.4850$

Figura 4.2: Ejemplos de escenas utilizadas para el reconocimiento de objetos y el plano de correlación obtenido para diferentes materiales basados en los modelos: (a) Lambert, (b) Phong, (c) Blinn-Phong, (d) Ward, y (e) Cook-Torrance.

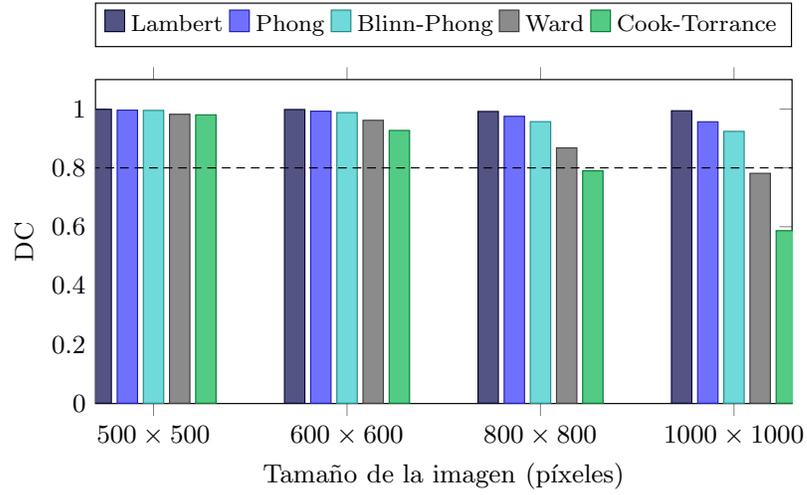
Se puede observar en la Fig. 4.2(a) que para el caso de un material difuso se obtiene un valor de $DC = 0.9904$, y el plano de correlación generado por el filtro presenta un pico de correlación fino libre de lóbulos, lo cual es deseable para una detección confiable. En las Fig. 4.2(b) y (c) se muestra que el valor de DC se deteriora gradualmente conforme el material del objeto se vuelve más reflectivo, tal es el caso de los modelos de Phong y Blinn-Phong, produciendo los valores $DC = 0.9546$ y $DC = 0.9212$, respectivamente.

Además, para materiales reflectivos de tipo metálico, como lo son los modelos de Ward y Cook-Torrance de las Fig. 4.2(d) y (e) se observa que el reconocimiento es de menor calidad debido a que el plano de correlación presenta lóbulos no deseados alrededor de la detección del objeto. Para el modelo de Ward se obtiene un valor de $DC = 0.7182$, para el modelo de Cook-Torrance se obtiene un valor $DC = 0.4850$.

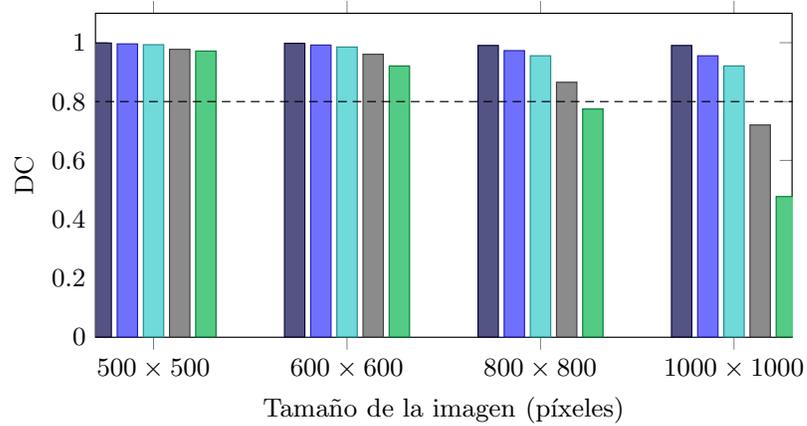
La evaluación de desempeño del filtro de correlación utilizado en términos de DC se muestra en la Fig. 4.3. En estos experimentos, el tamaño del objeto de la Fig. 4.1 es de 500×500 píxeles, el cual es incrustado dentro de un fondo con diferentes tamaños: 500×500 píxeles, 800×800 píxeles, y 1000×1000 píxeles. Con el fin de evaluar la robustez del algoritmo ante condiciones de ruido, la escena de entrada es contaminada con ruido aditivo Gaussiano con valores de relación señal-a-ruido (SNR) de 50 dB, 30 dB, y 10 dB.

En la Fig. 4.3 se observa que el desempeño del algoritmo propuesto en términos de la DC es influenciado por el modelo de iluminación analizado. Cuando se utiliza un material difuso a través del modelo de Lambert, el valor de DC obtenido es cercano a la unidad, sin embargo, el nivel de DC decrementa cuando existen reflexiones en el material, ésto es, cuando el parámetro K_s de las Ecs. (2.6) a la (2.9) incrementa.

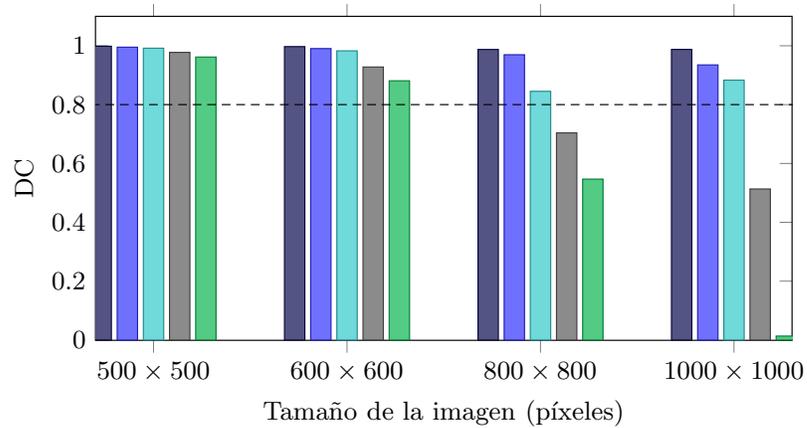
Se observa que el algoritmo de reconocimiento de objetos obtiene los mejores resultados con los materiales analizados a través de los modelos de Phong y Blinn-Phong. En condiciones de ruido con valores de SNR de 50 dB y 30 dB, el algoritmo propuesto



(a) 50 dB



(b) 30 dB



(c) 10 dB

Figura 4.3: Desempeño en el reconocimiento de objetos en términos de DC para diferentes materiales, ante condiciones de ruido de (a) 50 dB, (b) 30 dB, y (c) 10 dB SNR.

entrega valores de $DC > 0.8$. Sin embargo, el valor de DC decrece cuando el sistema es evaluado con un material reflectivo. Para los modelos de Ward y Cook–Torrance se obtienen valores cerca del umbral de $DC = 0.8$, excepto cuando la varianza del ruido aditivo y disjunto incrementa.

En el caso del modelo de Cook–Torrance el valor de DC disminuye abruptamente debido a los cambios de apariencia del objeto ocasionados por las altas reflexiones que se presentan en el material. Para un valor de SNR de 10 dB, el algoritmo analizado con los modelos de Lambert, Phong, y Blinn–Phong se mantiene con niveles de DC mayores a 0.845. Para los materiales metálicos simulados a través de los modelos de Ward y Cook–Torrance, en escenas con ruido aditivo con valor de SNR de 10 dB, el desempeño del algoritmo decrecienta drásticamente. En [92] se detallan estos experimentos realizados con diferentes niveles de ruido y fuentes de iluminación.

En estos ejemplos, se demuestra que la iluminación es un factor significativo para el reconocimiento de objetos. De igual manera, se observa que los filtros de correlación presentan robustez ante cambios de apariencia de los objetos en condiciones de ruido aditivo, presencia de objetos en el fondo de la escena, y a la interacción de la luz y el material de superficie.

4.2 Evaluación del algoritmo en una implementación computacional paralela heterogénea

El algoritmo propuesto se implementó utilizando computación paralela heterogénea, con el fin de reducir el tiempo de ejecución. Se realizó la paralelización del algoritmo en una arquitectura CPU/GPU utilizando el sistema operativo Linux en un procesador multinúcleo Intel i7 y dos tarjetas de procesamiento gráfico NVIDIA GeForce GTX780. Para reducir el tiempo de ejecución del algoritmo se utiliza en la implementación unidades de procesamiento gráfico (GPU) debido a la gran diferenciación

de desempeño, habilidad de realizar tareas de manera concurrente, y eficiencia ante cargas de trabajo masivas.

El programa principal se implementó en lenguaje C/C++, utilizando OpenMP (del inglés, *Open Multi-Processing*) para la paralelización de hilos en CPU y CUDA (del inglés, *Compute Unified Device Architecture*) para la paralelización en funciones kernel del GPU. Para el procesamiento de filtros de correlación en el dominio de la frecuencia, se ha utilizado CuFFT, que es una interfaz de aplicaciones de programación (API, del inglés, *application programming interface*) de CUDA. De igual manera, se utilizan las bibliotecas de funciones OpenCV y OpenGL para el procesamiento de imágenes y la generación de gráficos por computadora, respectivamente.

Se realizaron experimentos para una secuencia de video de 30 cuadros. La escena de entrada es una imagen sintética de 512×512 píxeles caracterizada por un objeto de interés en una pose arbitraria incrustado en un fondo. El objeto de referencia es un modelo digital tridimensional disponible en [93]. La visualización del objeto se generó en OpenGL mediante el modelo de iluminación de Phong, dado por la Ec. (2.6). La apariencia del objeto se determina mediante la pose y las propiedades de iluminación.

La Fig. 4.4 muestra un ejemplo de escenas de entrada con el objeto incrustado en un fondo. En la secuencia de imágenes el objeto se desplaza cambiando sus parámetros de estado (parámetros de pose y localización). De igual manera, se muestra la estimación de la pose del objeto mediante la combinación de los ángulos de rotación hacia los tres ejes coordenados y la localización del objeto en la imagen.

Con el fin de evaluar el desempeño del algoritmo en términos de calidad del reconocimiento de objetos se utiliza la métrica DC. La Fig. 4.5(a) muestra el desempeño del algoritmo paralelo en términos de la DC con respecto a los cuadros en la secuencia de video. Se observa que el desempeño se mantiene uniforme conforme

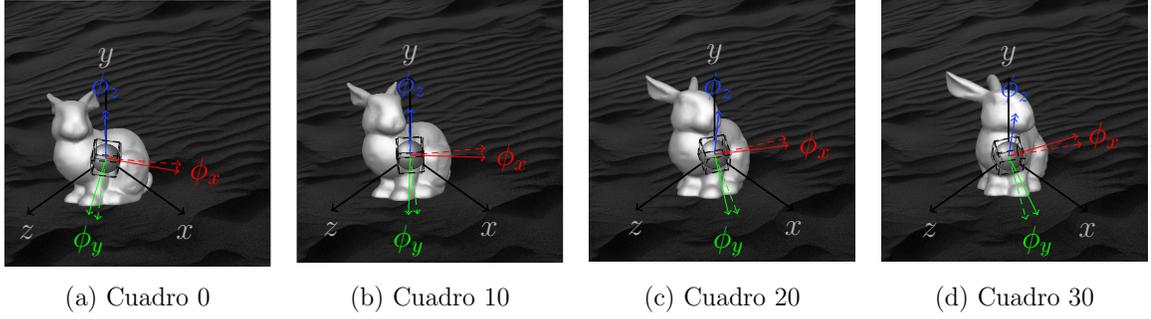


Figura 4.4: Ejemplo de cuadros de entrada con la estimación real (línea sólida, \rightarrow) y estimada (línea punteada, $--\rightarrow$) de los ángulos de orientación.

procesa cada uno de los cuadros de escena. El algoritmo propuesto obtiene un valor de $DC = 0.76 \pm 0.06$, lo cual indica buena eficiencia en el reconocimiento de objetos.

De igual manera, se presenta una mejora en el valor de DC obtenido utilizando $N = \{2^3, 4^3, 6^3, 8^3, 10^3\}$ como parámetro que define el tamaño del espacio de búsqueda para generar el banco de filtros de correlación, tal como se muestra en la Fig. 4.5(b). Se observa un buen desempeño a partir del procesamiento de 6^3 filtros de correlación en el banco. El algoritmo paralelo propuesto entrega el mejor desempeño con un valor de $DC = 0.84$ cuando se utiliza un banco de 10^3 filtros.

Adicionalmente, para evaluar el desempeño del algoritmo se han utilizado otras dos métricas cuantitativas. La primera métrica de evaluación consiste en caracterizar la precisión de la estimación de las coordenadas de localización del objeto, la cual se calcula mediante el error de localización (LE), definida por [87]

$$LE = \|\boldsymbol{\tau} - \hat{\boldsymbol{\tau}}\|, \quad (4.1)$$

donde $\boldsymbol{\tau}$ y $\hat{\boldsymbol{\tau}}$ equivalen a las coordenadas reales y estimadas del objeto en la escena, respectivamente. La segunda métrica evalúa la precisión de la estimación de los

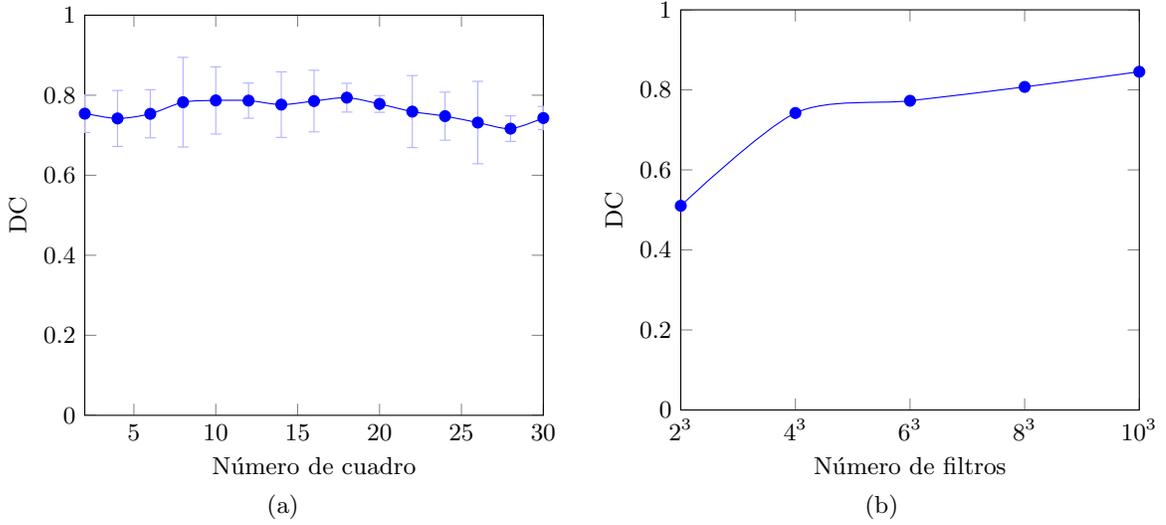


Figura 4.5: Desempeño del algoritmo paralelo propuesto en términos de DC con respecto (a) al número de cuadros, y (b) al número de filtros utilizados en el banco.

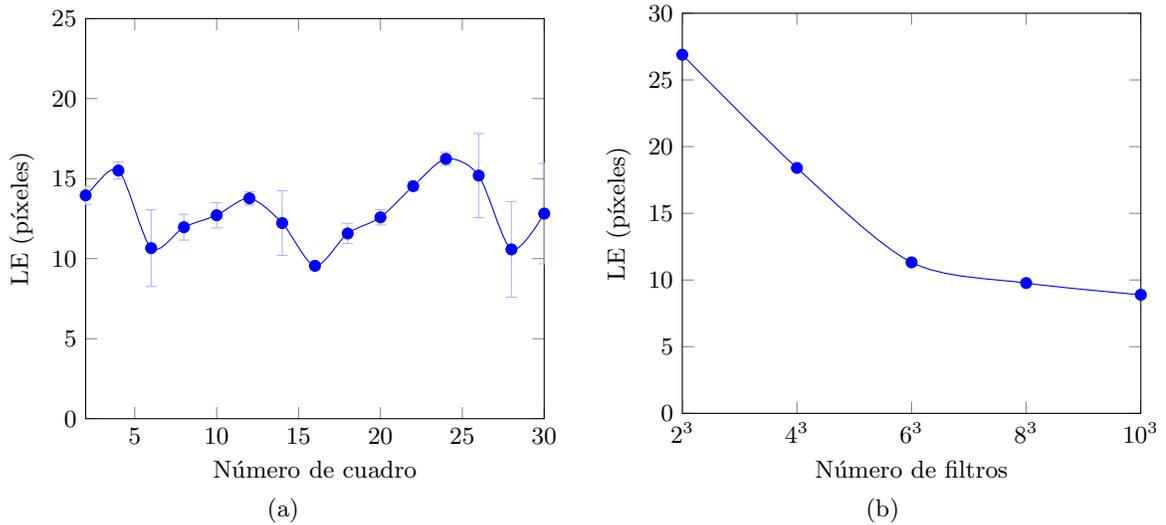


Figura 4.6: Error de localización del algoritmo paralelo propuesto con respecto (a) al número de cuadros, y (b) al número de filtros utilizados en el banco.

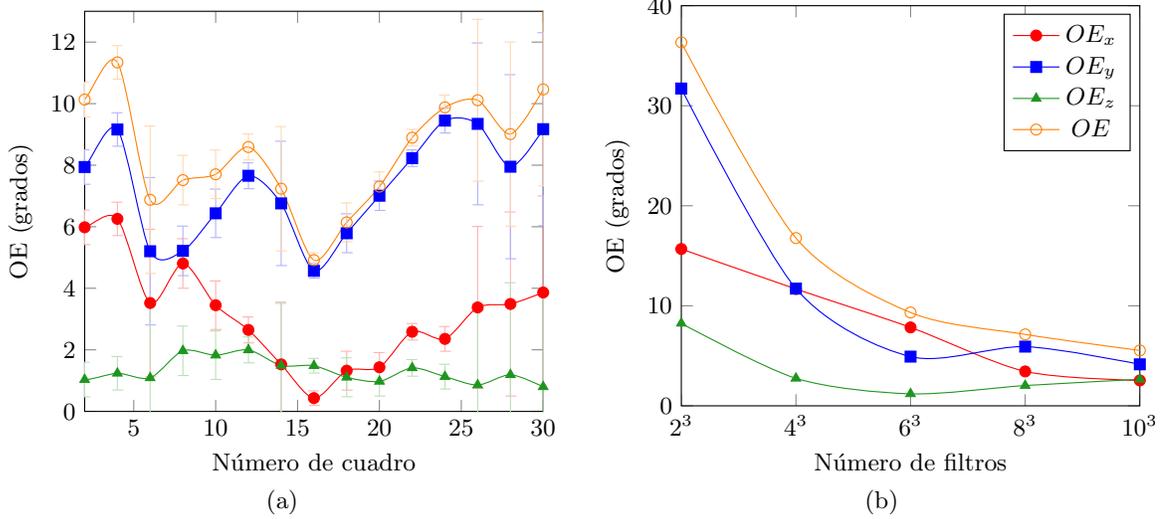


Figura 4.7: Error de orientación del algoritmo paralelo propuesto con respecto (a) al número de cuadros, y (b) al número de filtros utilizados en el banco.

ángulos de rotación del objeto, definida por el error de orientación (OE)

$$OE = \left\| \phi - \hat{\phi} \right\|, \quad (4.2)$$

donde ϕ y $\hat{\phi}$ equivalen a los ángulos de rotación actuales y estimados del objeto en la escena, respectivamente. La Fig. 4.6 ilustra los resultados obtenidos de la estimación de la localización del objeto en la escena. El algoritmo propuesto es evaluado con respecto al número de cuadros en la secuencia de prueba. De acuerdo a la Fig. 4.6(a), se obtiene un valor estadístico de $LE = 12.88 \pm 1.18$ píxeles con el 95% de confianza. Se muestra que el error de localización se puede reducir conforme se incrementa el número de filtros en el banco. La Fig. 4.6(b) muestra que el valor esperado se reduce hasta $LE = 8.89$ píxeles cuando se procesa en un espacio de búsqueda de 10^3 filtros utilizados en el banco. Para obtener alta eficiencia en la estimación de la localización del objeto, es necesario ejecutar al menos 6^3 filtros de correlación.

Tabla 4.1: Desempeño del algoritmo paralelo en términos del número de filtros en el banco.

Número de filtros	LE (píxeles)	OE (grados)	DC
8	26.89	36.33	0.5103
64	18.40	16.77	0.7422
216	11.32	9.33	0.7728
512	9.77	7.14	0.8074
1000	8.89	5.54	0.8453

En la Fig. 4.7(a) se muestra el error de orientación resultante en los tres ángulos coordenados. La secuencia de entrada presenta cambios en la orientación del objeto en la escena con un comportamiento dinámico desconocido. Con un 95% de confianza el error obtenido en cada ángulo es $OE_x = 3.29 \pm 1.57$ grados, $OE_y = 7.28 \pm 1.13$ grados, y $OE_z = 1.30 \pm 0.88$ grados. El error de orientación general dado por la Ec. (4.2) es de $OE = 8.43 \pm 1.14$ grados. La Fig. 4.7(b) muestra que la estimación de orientación se mejora conforme el número de filtros incrementa en el banco. El algoritmo paralelo propuesto puede obtener una estimación de hasta $OE = 5.54$ grados de rotación al utilizar 10^3 filtros de correlación.

La Tabla 4.1 resume que la calidad del reconocimiento de objetos incrementa conforme se aumenta el tamaño del banco de filtros. De igual manera, el error de estimación de la pose del objeto se reduce en términos de LE y OE con respecto al incremento del número de filtros utilizados en el banco.

La paralelización del algoritmo se realizó utilizando $N = \{1, 2, 4, 8\}$ hilos de CPU, y también, en el proceso de correlación con CUDA/CUFFT API. La Fig. 4.8 muestra el desempeño del sistema en términos del tiempo de ejecución computacional (*speedup*). Una ejecución completamente secuencial se indica con $N = 1$, en cuya ejecución no se utiliza paralelización en CPU ni GPU. A partir de $2 \leq N \leq 8$ hilos de CPU, se utiliza paralelización con OpenMP y funciones de ejecución concurrentes

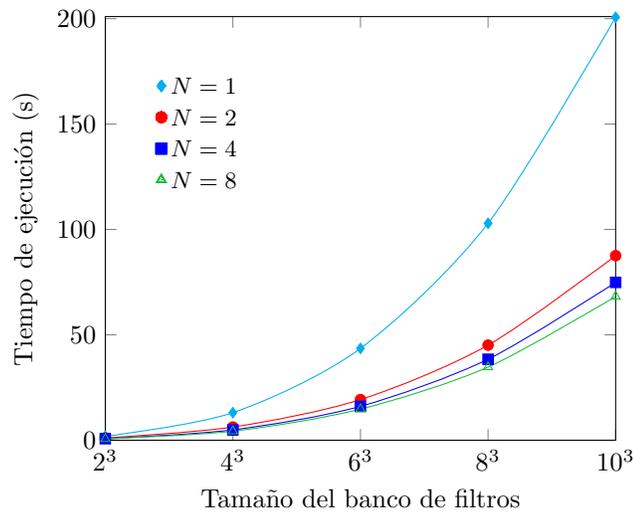


Figura 4.8: Tiempo de ejecución de la implementación paralela del algoritmo propuesto.

(*kernels*) de GPU. El tamaño del banco de filtros indica el número de imágenes de referencia utilizadas en el proceso de correlación. La implementación computacional es escalable, ésto es, que cuando el tamaño de filtros aumenta, se obtiene una mejora de 2.96 veces en *speedup* [94].

4.3 Evaluación de desempeño en reconocimiento de pose 3D a partir de imágenes monoculares

El desempeño del algoritmo propuesto para el reconocimiento de pose es analizado y discutido mediante el procesamiento de escenas sintéticas y reales. Los resultados obtenidos se evalúan en términos de la precisión del reconocimiento de pose en diferentes condiciones en la escena, tales como presencia de ruido aditivo, objetos presentes en el fondo de la imagen, cambios en la iluminación, y oclusiones parciales.

Para este experimento, el algoritmo se implementó en una computadora multi-núcleo con Intel Core i7-4790K CPU, en un sistema operativo Linux. El algoritmo fué programado en lenguaje C/C++, utilizando la biblioteca de funciones OpenCV

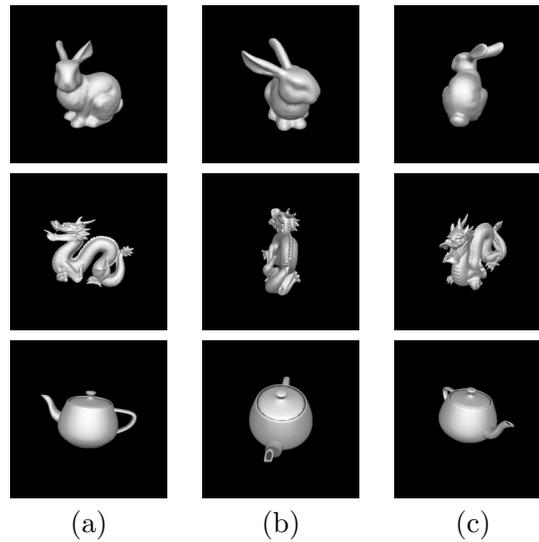


Figura 4.9: Ejemplos de los objetos 3D utilizados en los experimentos con diferentes poses. (a) bunny, (b) dragon y (c) teapot.

para el procesamiento de imágenes, y OpenGL para la generación y visualización de gráficos por computadora de los modelos 3D.

En los experimentos, se han utilizado diferentes modelos de referencia: *bunny* [93], *teapot* [95], y *dragon* [96], tal como se muestran en la Fig. 4.9. El primer experimento consisten en la evaluación del reconocimiento de pose con escenas sintéticas. Para este experimento se utilizan imágenes monocromáticas de 512×512 píxeles. El rango de la señal es de $[0,1]$ con 256 niveles de cuantización. Todas las imágenes contienen un objeto incrustado en un fondo, y los parámetros de localización y orientación son desconocidos.

La Fig. 4.10 presenta ejemplos de diferentes escenas sintéticas utilizadas en los experimentos. El contorno amarillo en la imagen representa las soluciones candidatas probadas por el algoritmo. Además, el contorno verde indica la pose final entregada por el sistema. Las imágenes que se muestran en la Fig. 4.10 son generadas a partir de los modelos 3D (ver Fig. 4.9) para diversos estados de pose. Los modelos de los objetos son visualizados con las propiedades de superficie establecidas por el

modelo de iluminación de Phong [37]. Se realizaron treinta muestras independientes, en donde se prueba el algoritmo variando el número de iteraciones en el proceso de refinamiento en $i = \{1, 2, \dots, 8\}$, y variando el número de filtros utilizados en el banco $N = \{4^3, 5^3, \dots, 7^3\}$. Se observa que estos parámetros modifican el tamaño del espacio de búsqueda explorado por el algoritmo. Los resultados obtenidos se muestran en la Fig. 4.11. Se observa que el mejor desempeño del algoritmo se obtiene al alcanzar el número máximo de iteraciones en la etapa de refinamiento, y utilizando el número máximo de filtros de correlación. Con 95% de confianza, el algoritmo propuesto presenta una precisión de $LE = 1.56 \pm 0.29$ píxeles y $OE = 3.58 \pm 0.22$ grados.

Además, se evalúa el desempeño de reconocimiento del algoritmo en términos de la DC. La Fig. 4.12 muestra con 95% de confianza el desempeño del algoritmo en términos de la DC como una función para evaluar el número de iteraciones en la etapa de refinamiento, y el número de filtros de correlación utilizados. Se observa que el algoritmo presenta alta eficiencia en la detección del objeto con un $DC = 0.95 \pm 0.04$, cuando se utilizan $N = 7^3$ filtros y 8 iteraciones en el refinamiento.

Por otra parte, se evalúa la tolerancia al ruido aditivo del algoritmo. Se procesaron treinta diferentes escenas para estimar la pose del objeto. Las escenas se contaminaron con ruido Gaussiano de media cero, con diferentes valores de varianza $\sigma^2 = \{0.25, 0.5, 0.75, 1\}$. La Fig. 4.13 ilustra diferentes ejemplos de la estimación de pose del objeto marcada en verde para escenas ruidosas. La Fig. 4.14 muestra el desempeño del algoritmo propuesto en términos de las métricas LE y OE con respecto al número de iteraciones del proceso de refinamiento. En la Fig. 4.14(a) se observa que para valores pequeños de varianza del ruido ($\sigma^2 = 0.25$ y $\sigma^2 = 0.5$) el LE se reduce significativamente a partir de cuatro iteraciones de refinamiento. Con el 95% de confianza, el desempeño del algoritmo en escenas con valores de varianza de ruido aditivo de $\sigma^2 = 0.25$ y $\sigma^2 = 0.5$ es de $LE = 1.64 \pm 0.63$ píxeles. Cuando la escena se contamina con ruido aditivo con valores de varianza de $\sigma^2 = 0.75$ y $\sigma^2 = 1$, se vuelve

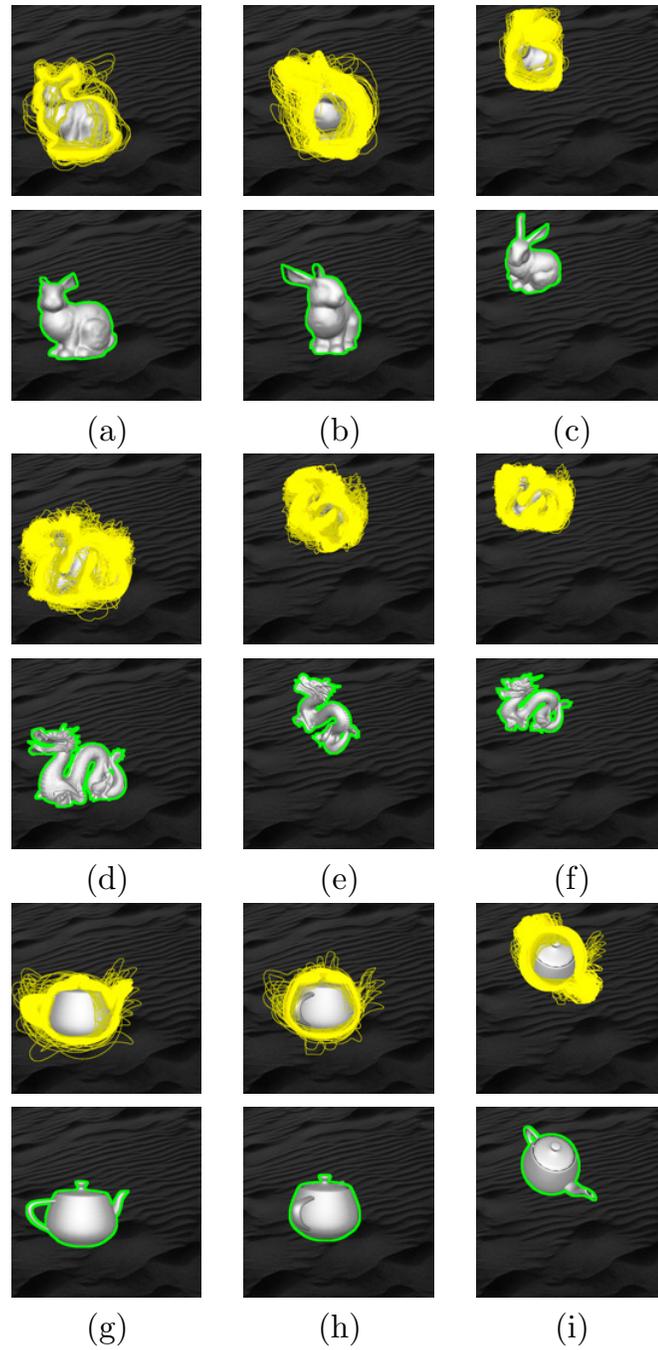


Figura 4.10: Ejemplos de escenas sintéticas utilizadas en los experimentos. (a)-(c) *bunny*, (d)-(f) *dragon*, (g)-(i) *teapot*. Las soluciones candidatas se representan en amarillo, y la solución final se representa en verde.

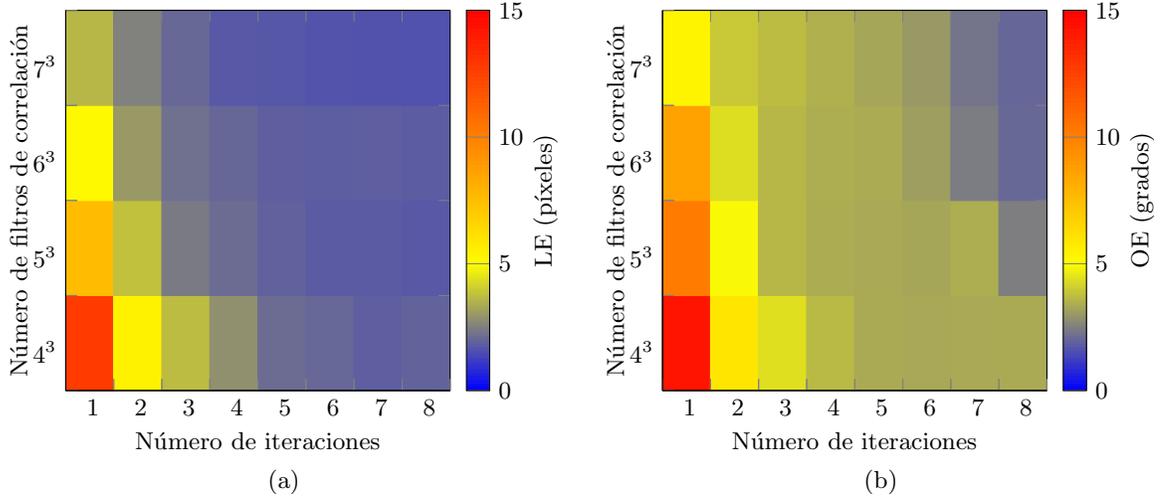


Figura 4.11: Desempeño del algoritmo propuesto en términos de (a) LE y (b) OE, en escenas con ruido, respecto al número de filtros de correlación.

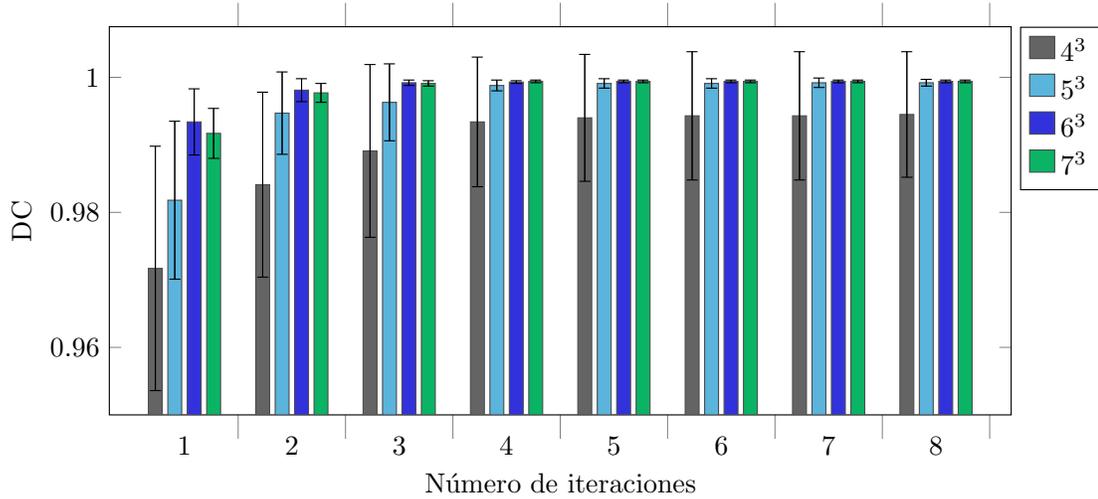


Figura 4.12: Eficiencia de la detección del algoritmo propuesto en términos de la DC y el número de iteraciones en el proceso de refinamiento, para bancos con 4³, 5³, 6³ y 7³ filtros de correlación.

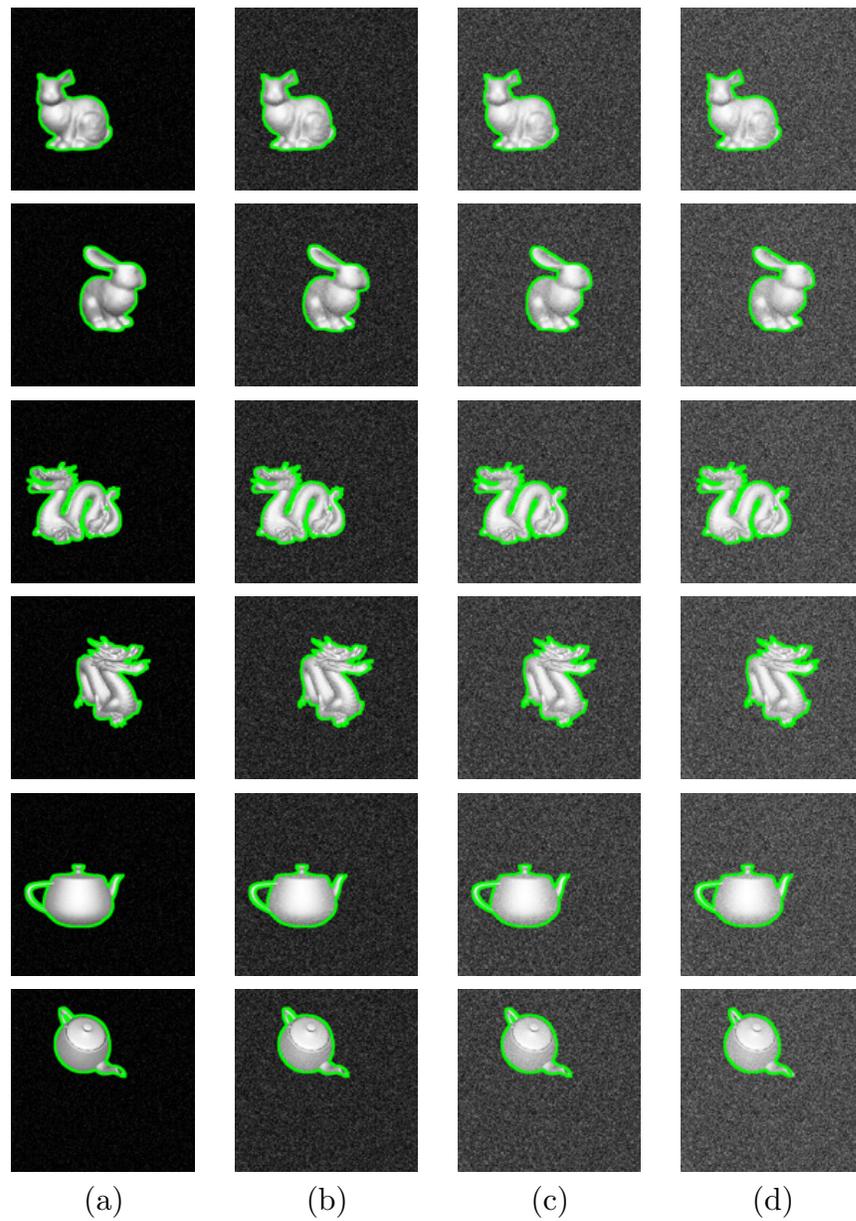


Figura 4.13: Resultados obtenidos del desempeño del reconocimiento de pose en escenas ruidosas. (a) $\sigma^2 = 0.25$, (b) $\sigma^2 = 0.5$, (c) $\sigma^2 = 0.75$, y (d) $\sigma^2 = 1$.

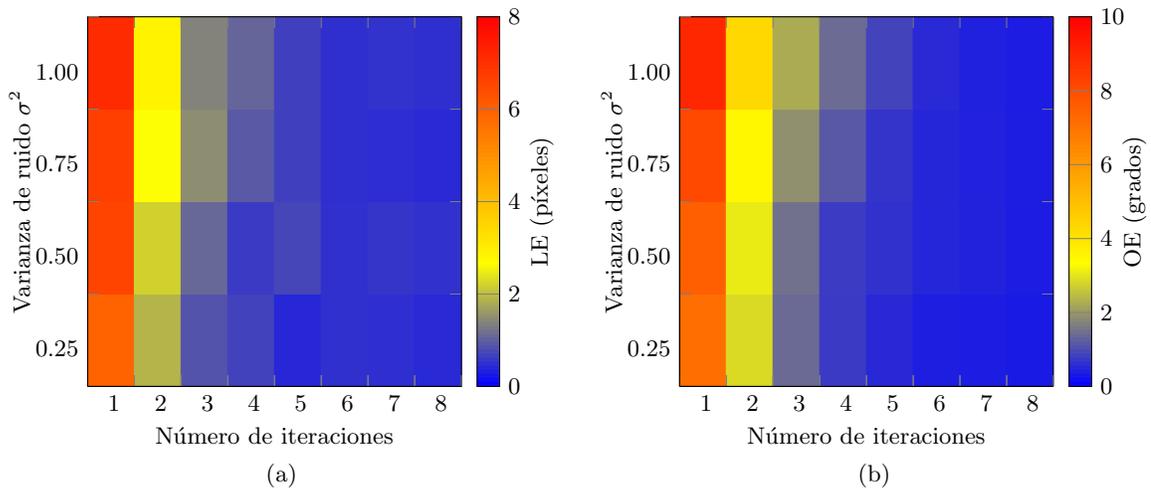


Figura 4.14: Desempeño del algoritmo propuesto en términos de (a) LE y (b) OE en escenas ruidosas con respecto al número de iteraciones en la etapa de refinamiento.

más difícil obtener una alta precisión en la estimación de pose. Con el fin de reducir los valores de LE obtenidos con niveles altos de ruido, se recomienda ejecutar al menos 8 iteraciones de refinamiento. La Fig. 4.14(b) muestra el desempeño del algoritmo con escenas ruidosas en términos de OE. Se observa que para todos los niveles de ruido, el algoritmo es capaz de estimar la orientación del objeto con alta precisión. Cuando la varianza del ruido aditivo es de $\sigma^2 = 0.25$ y $\sigma^2 = 0.5$ el algoritmo obtiene valores bajos de OE después de algunas iteraciones de refinamiento. Para condiciones de ruido altas, con valores de varianza de $\sigma^2 = 0.75$ y $\sigma^2 = 1$, el algoritmo requiere más refinamiento para estimar la pose con buena precisión. Con un 95% de confianza el desempeño general del algoritmo propuesto en términos de la estimación de pose es de $OE = 2.02 \pm 0.59$ grados.

El siguiente experimento consiste en evaluar la robustez del algoritmo en oclusiones parciales del objeto. Para ésto, fueron creadas imágenes sintéticas que contienen un objeto de interés, en donde otros objetos obstruyen parcialmente la visualización del objeto a reconocer. Se consideraron dos casos para esta evaluación; ésto es, 25% y 50% de oclusiones del objeto. La Fig. 4.15 muestra ejemplos de algunas escenas

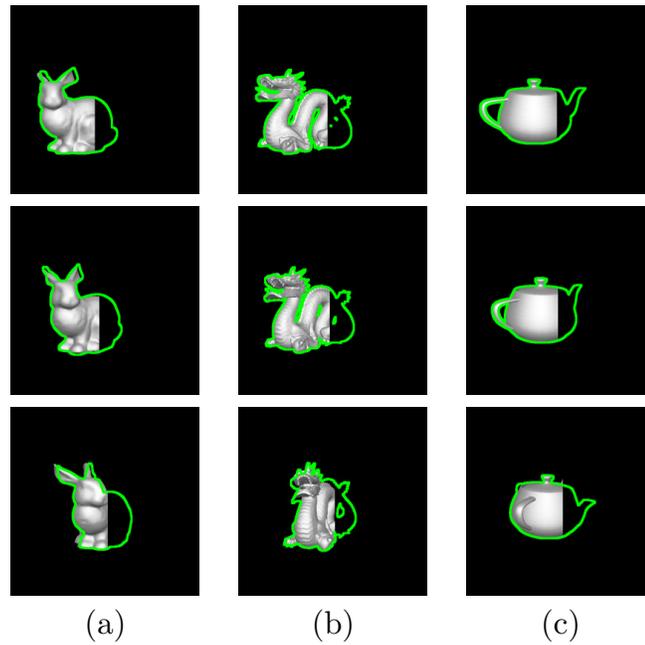


Figura 4.15: Desempeño del algoritmo propuesto cuando el 25% del objeto se encuentra ocluido.

de prueba cuando el 25% del área del objeto es ocluida. El contorno verde en cada imagen ilustra la estimación de pose final del objeto. En la mayoría de los casos la pose del objeto en la escena se obtiene con buena precisión. Se observa que el error más significativo en la estimación de pose del objeto ocurre con el modelo *teapot*, tal como se muestra en la Fig. 4.15(c). Ésto es, debido a la apariencia isotrópica del modelo, por lo cual se convierte en un reto difícil de resolver en los casos cuando el objeto ocluye características distintivas de pose (ej. la oclusión de la asa). De igual manera, la Fig. 4.16 muestra los resultados obtenidos con una oclusión parcial del 50% del área del objeto. Se observa que a pesar de que la mitad del objeto no es visible en la escena, la pose del objeto se estima correctamente en la mayoría de los casos. En la Fig. 4.17(a) se muestra el desempeño del algoritmo propuesto, con el 95% de confianza y dado en términos de LE y OE. Se observa que para el modelo *bunny* se obtiene una estimación de $LE = 8.37 \pm 1.99$ píxeles y $OE = 8.01 \pm 2.00$ grados.

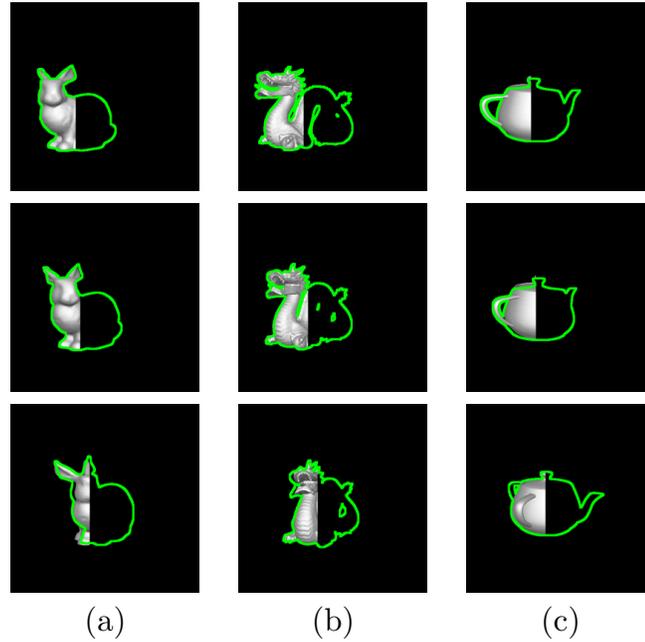


Figura 4.16: Desempeño del algoritmo propuesto cuando el 50% del objeto se encuentra ocluido.

Para el caso del modelo *dragon* de la Fig. 4.17(b), los resultados son diferentes, éstos es, $LE = 3.14 \pm 1.39$ píxeles y $OE = 4.36 \pm 1.52$ grados para el 50% de oclusión. Se observa que para todos los objetos analizados, el modelo *dragon* es el que presenta la mayor precisión en la estimación de la pose. Se puede deducir que ésto es debido a que su composición geométrica presenta mayor número de características que apoyan la estimación de pose. El caso contrario, es el modelo *teapot*, que por su simetría puede ocluir características de pose. La Fig. 4.17(c) presenta los resultados obtenidos con el modelo *teapot*. La estimación de pose en términos de las métricas de error de localización y orientación son $LE = 5.75 \pm 4.12$ píxeles y $OE = 6.21 \pm 3.30$ grados, respectivamente.

El siguiente experimento consiste en evaluar el algoritmo propuesto con escenas reales. Con el fin de utilizar los mismos objetos presentados en la Fig. 4.9, se ha utilizado una impresora 3D para reproducir los objetos reales a partir de los modelos

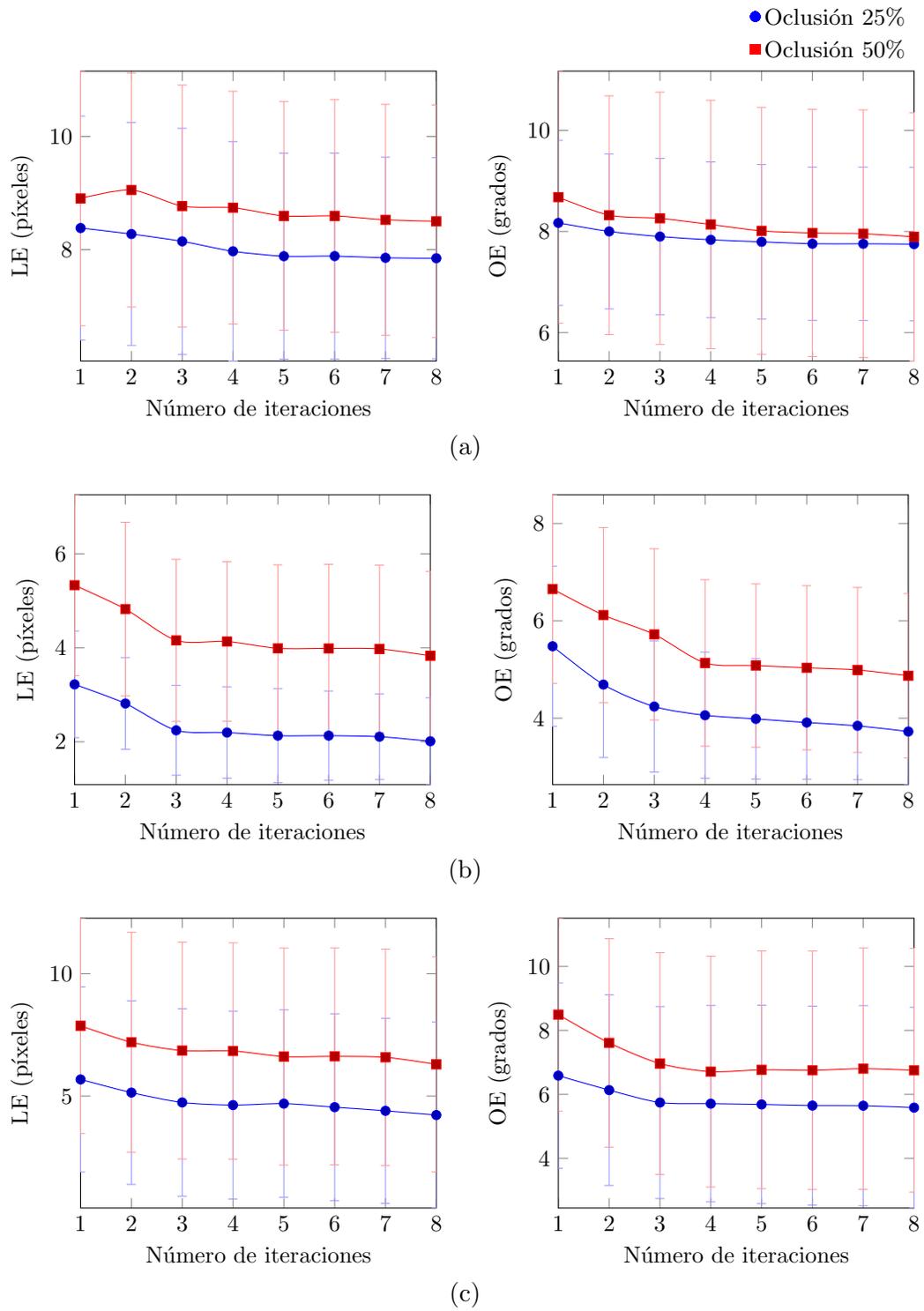


Figura 4.17: Evaluación del algoritmo para la estimación de pose del objeto parcialmente ocluido, utilizando los modelos 3D (a) *bunny*, (b) *dragon*, (c) *teapot*.

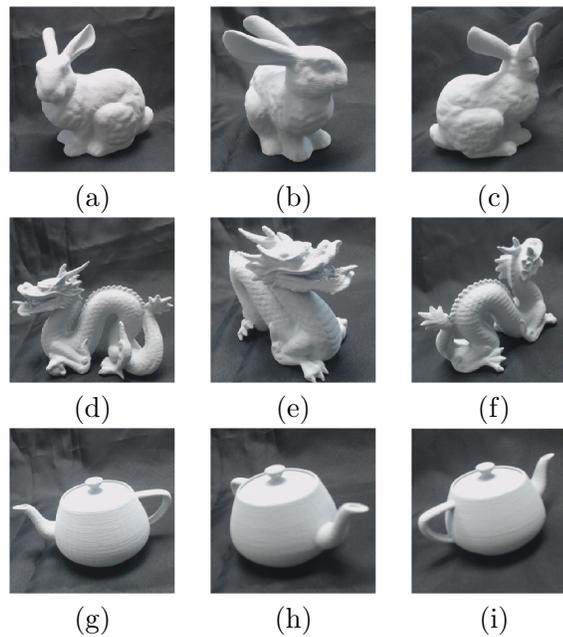


Figura 4.18: Diferentes vistas de los objetos impresos en 3D utilizados en los experimentos. (a–c) *bunny*, (d–f) *dragon*, (g–i) *teapot*.

digitales. Los objetos resultantes de la impresión 3D se muestran en la Fig. 4.18 en diferentes poses. Para este experimento, las escenas de prueba son imágenes a color RGB de 512×512 píxeles, que han sido capturadas en un ambiente de oficina. La Fig. 4.19 muestra algunos ejemplos de escenas utilizadas en los experimentos. Las escenas de prueba contienen un objeto incrustado en un fondo con parámetros de localización y orientación desconocidas. El contorno verde indica la pose final estimada del objeto. Se observa que las Figs. 4.19(b,c,f,i) muestran imágenes con oclusiones parciales del área del objeto. Se puede observar que los resultados que se muestran en la Fig. 4.19 demuestran la habilidad del algoritmo propuesto en reconocer un objeto y estimar los parámetros de pose con alta precisión.

Por otra parte, la Fig. 4.20 muestra las poses resultantes de la etapa de refinamiento en escenas reales y sintéticas. El algoritmo propuesto es capaz de estimar la pose del objeto con una alta precisión a partir de algunas iteraciones.

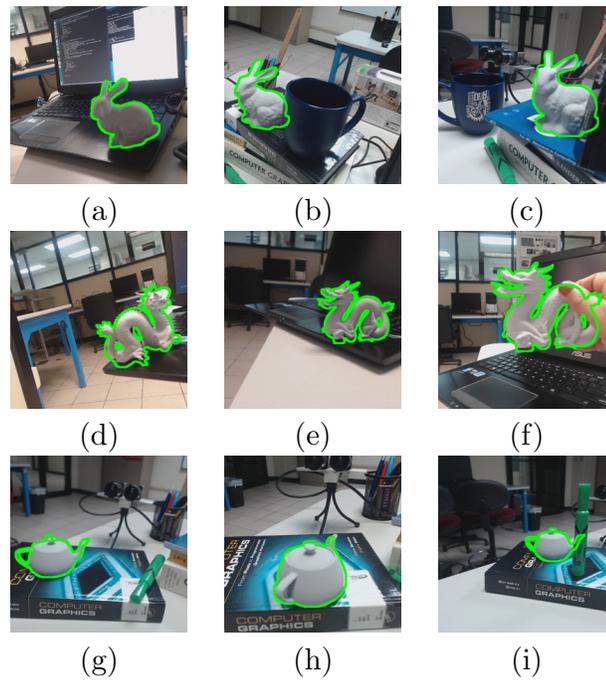


Figura 4.19: Resultados obtenidos con el algoritmo propuesto para el reconocimiento de pose en escenas reales.

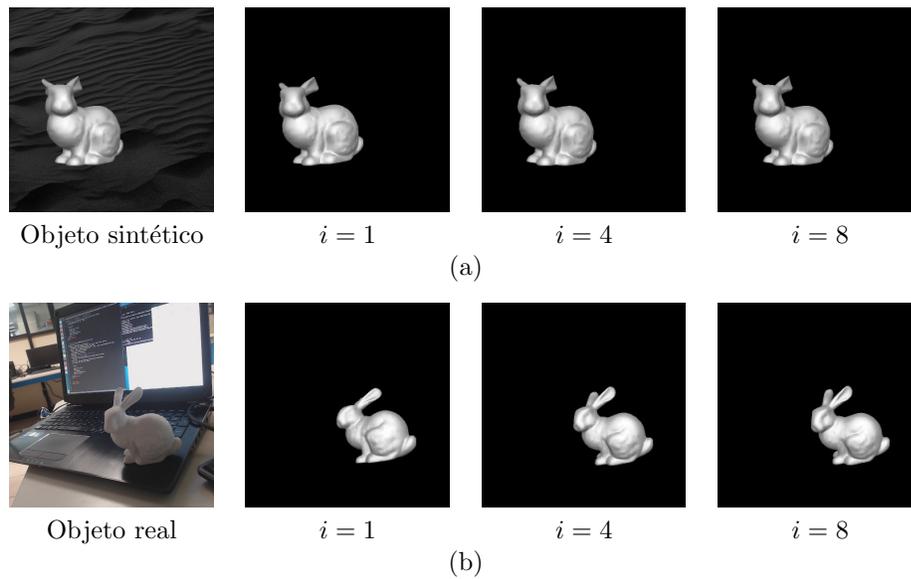


Figura 4.20: Ejemplo de la operación del algoritmo propuesto en la etapa de refinamiento en diferentes iteraciones para una (a) escena sintética, y una (b) escena real.

Por último, se comparan los resultados obtenidos con el algoritmo propuesto con otros tres algoritmos recientes en el estado del arte para estimación de pose en 3D. Éstos son, el algoritmo de partículas propuesto por Lee et al. [9], el algoritmo basado en regiones propuesto por Dambreville et al. [13], y el algoritmo basado en funciones *kernel* propuesto por Sandhu et al [14]. La precisión de la estimación de la pose para cada uno de los algoritmos evaluados se caracteriza en términos de los porcentajes de los errores abasolutos de la localización (LAE) y orientación (OAE), los cuales se calculan como a continuación [9]:

$$\%Error = \frac{\|v_{\text{real}} - v_{\text{estimado}}\|}{\|v_{\text{real}}\|} \times 100\%. \quad (4.3)$$

La Tabla 4.2 presenta el desempeño de los algoritmos evaluados para escenas degradadas con ruido aditivo y oclusiones parciales del objeto. El algoritmo de Lee et al. [9] demuestra una buena tolerancia a la presencia de ruido y oclusiones. Además se observa que el algoritmo propuesto por Dambreville et al. [13] tiene un desempeño sobresaliente, sin embargo, el reconocimiento de pose obtenido con el algoritmo de Lee et al. [9] es resuelto para objetos móviles con geometrías más complicadas. El algoritmo de Sandhu et al. [14] produce excelentes resultados en términos de LAE y OAE en todas las pruebas. Este algoritmo es muy efectivo ya que optimiza el modelo 3D del objeto utilizado como información *a priori* del sistema de reconocimiento. La Tabla 4.2 muestra que el algoritmo propuesto es robusto en la presencia de ruido aditivo y oclusiones parciales. El algoritmo propuesto de reconocimiento de pose 3D a partir de imágenes monoculares ha demostrado una alta eficiencia [88]. En comparación con los algoritmos existentes en la literatura, el algoritmo propuesto muestra un desempeño superior en términos de errores de localización y orientación, en presencia de ruido y oclusiones del objeto.

Tabla 4.2: Comparación de desempeño de los algoritmos de reconocimiento de pose en escenas reales.

Método	Ruido		Oclusión	
	LAE (%)	OAE (%)	LAE (%)	OAE (%)
Lee, et al. [9]	3.17	4.27	2.12	3.49
Dambrevelle, et al. [13]	0.95	1.37	2.61	4.36
Sandhu, et al. [14]	1.63	1.31	2.51	2.75
Método propuesto [88]	0.48	0.56	1.98	1.82

4.4 Evaluación de desempeño para el seguimiento de pose 3D utilizando filtros de partículas

En esta sección, se presentan los resultados obtenidos con el algoritmo propuesto para estimación y seguimiento de pose en 3D de un objeto en movimiento. Para ésto, se desarrolló una implementación computacional que consiste en procesar una secuencia de video de 300 cuadros monocromáticos con una resolución de 512×512 píxeles. Con el fin de demostrar el desempeño del método propuesto, se han utilizado secuencias sintéticas y reales con diferentes objetos. Cada imagen de entrada se caracteriza por un objeto 3D en una pose desconocida, con presencia de fondo y ruido aditivo, de acuerdo con la Ec. (3.14).

La apariencia del objeto está caracterizada por la combinación de los tres ángulos de orientación, coordenadas de posición y escalamiento. La pose 3D estimada se obtiene mediante el mejor acoplamiento entre la imagen de entrada y el conjunto de plantillas creadas mediante gráficos por computadora. Cada plantilla se define con la vista paramétrica del objeto en una pose 3D específica.

En los experimentos realizados se utilizan los modelos *bunny* [93], *dragon* [96], *teapot* [95], y *elephant*¹, los cuales se ilustran en la Fig. 4.21. La visualización tridimensional de los modelos se realiza en OpenGL utilizando propiedades básicas de

¹El modelo 3D *elephant* fue provisto por J. Lee, presentado en [97].

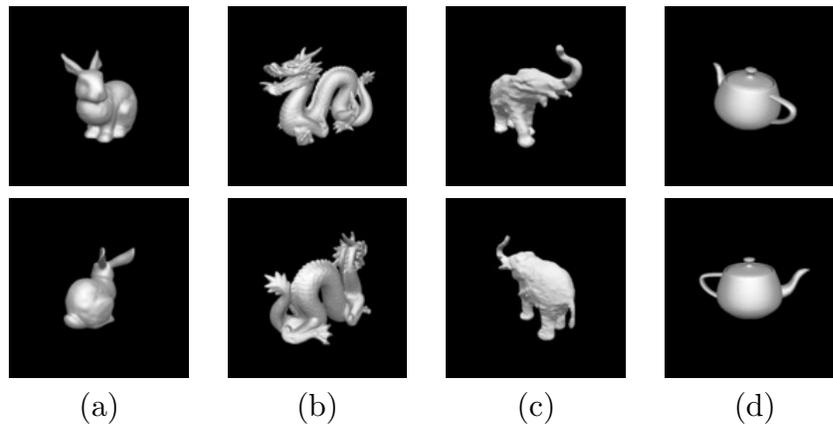


Figura 4.21: Vistas de los modelos 3D utilizados en los experimentos. (a) *bunny*, (b) *dragon*, (c) *elephant*, y (d) *teapot*.

superficie definidas en el modelo de iluminación de Phong [37]. Los experimentos están diseñados para evaluar el desempeño del método propuesto con un enfoque híbrido. Con el fin de resolver la estimación y seguimiento de pose en 3D para objetos en movimiento, se evaluó el algoritmo propuesto mediante filtros de correlación adaptativos por medio de filtros de partículas. El desempeño del sistema propuesto se evalúa en términos de la estimación de la localización y la orientación del objeto dentro de una escena de entrada, la capacidad de detección del sistema de reconocimiento de objetos, y su robustez ante ruido aditivo.

La precisión de la estimación de la pose del objeto en la imagen de entrada está definida por los errores de localización (LE) y de orientación (OE), definidos por las Ecs. (4.1) y (4.2), respectivamente. De igual manera, el porcentaje de error obtenido con respecto a la pose real está dada por el error absoluto de localización y orientación (LAE y OAE, respectivamente), ambos definidos por la Ec. (4.3).

El primer experimento consiste en analizar el desempeño del algoritmo propuesto en términos del LE y OE, con el fin de evaluar la precisión de la estimación de la pose del objeto. La Fig. 4.22(a) presenta el LE obtenido con el algoritmo propuesto en diferentes cuadros de la secuencia de prueba. El desempeño del algoritmo fue probado

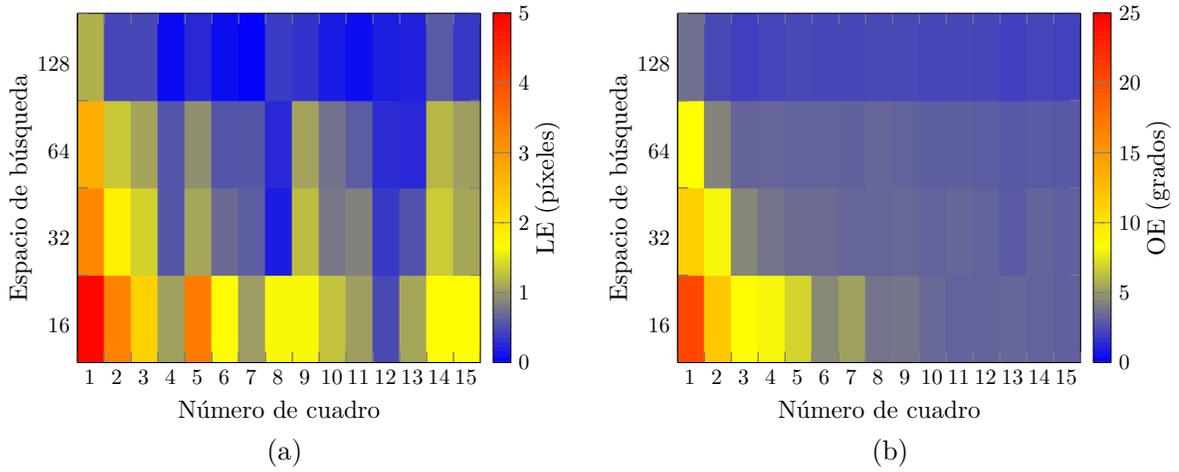


Figura 4.22: Desempeño del algoritmo propuesto en términos de (a) LE y (b) OE, en relación al número de cuadros en la escena y número de partículas utilizadas.

en 30 ejecuciones independientes, utilizando un conjunto de $N = \{16, 32, 64, 128\}$ filtros para definir el tamaño del espacio de búsqueda del filtro de partículas. Con un nivel del 95% de confianza, el algoritmo propuesto logra una precisión en la estimación de las coordenadas del objeto de $LE = 0.35 \pm 0.16$ píxeles. Se puede observar en la Fig. 4.22(a), que el desempeño del sistema en términos del LE mejora si se ejecuta una cantidad grande de partículas en el espacio de búsqueda; en este caso 128 partículas por iteración en la etapa de refinamiento. La Fig. 4.22(b) presenta el desempeño del algoritmo propuesto en términos del OE para diferentes cuadros de la secuencia de entrada. Como se puede observar, el OE disminuye significativamente conforme el tamaño del espacio de búsqueda incrementa desde $N = \{16, 32, 64, 128\}$ partículas. Se puede observar también, que el método propuesto permite una precisión de $OE = 2.74 \pm 0.29$ grados con un espacio de búsqueda de 128 partículas de refinamiento iterativo.

El algoritmo propuesto también fue evaluado al procesar imágenes degradadas con ruido aditivo. Las pruebas consisten en procesar una secuencia de 300 cuadros de escena, cada uno con una resolución de 512×512 píxeles. En los experimentos,

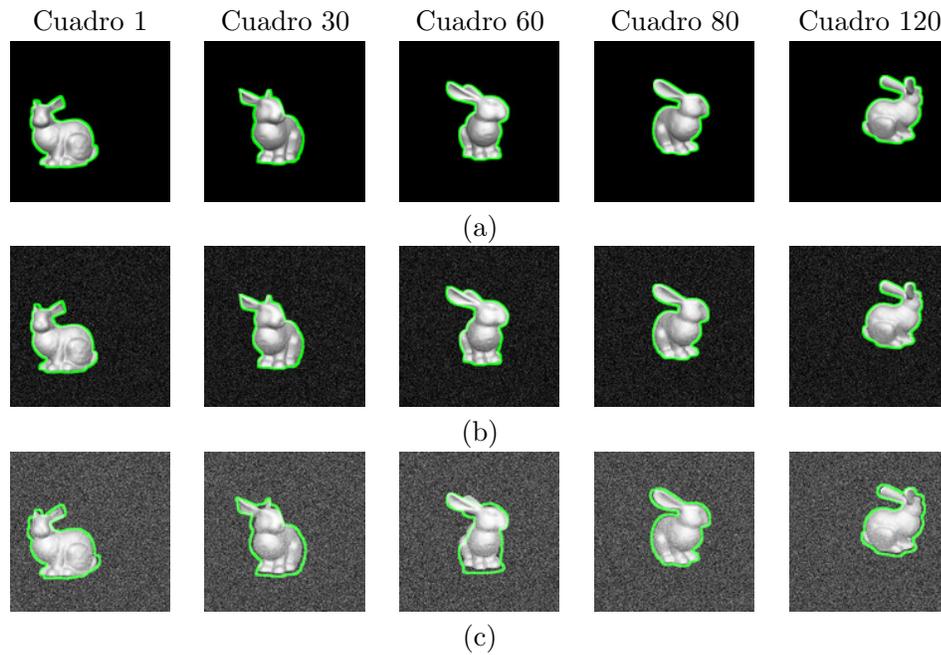


Figura 4.23: Resultados de estimación de pose con diferentes niveles de ruido aditivo: (a) $\sigma^2 = 0.1$, (b) $\sigma^2 = 0.4$, y (c) $\sigma^2 = 1$.

primero se generan datos sintéticos, y luego se inserta ruido aditivo a diferentes niveles. Cada cuadro de entrada es contaminado con ruido aditivo Gaussiano $\mathcal{N}(0, \sigma^2)$. La Fig. 4.23 presenta los resultados de la pose estimada para escenas ruidosas. Como referencia de comparación, la Fig. 4.24 ilustra los resultados del algoritmo propuesto y los resultados obtenidos en [13, 14, 97] en términos de robustez a ruido aditivo para diferentes niveles $0.01 \leq \sigma^2 \leq 1$.

El algoritmo propuesto se comparó con el algoritmo presentado por Lee, et al. [97], con el método basado en regiones propuesto por Dambreville, et al. [13] y el algoritmo basado en funciones kernel propuesto por Sandhu et al. [14]. Se calcularon los valores de LAE y OAE por medio de la Ec. (4.3). Para una prueba de consistencia, el algoritmo fue evaluado con las secuencias de video presentadas en la Fig 4.23 a través de 30 ejecuciones independientes. La Fig. 4.24(a) presenta la mejora del algoritmo propuesto en términos del LAE.

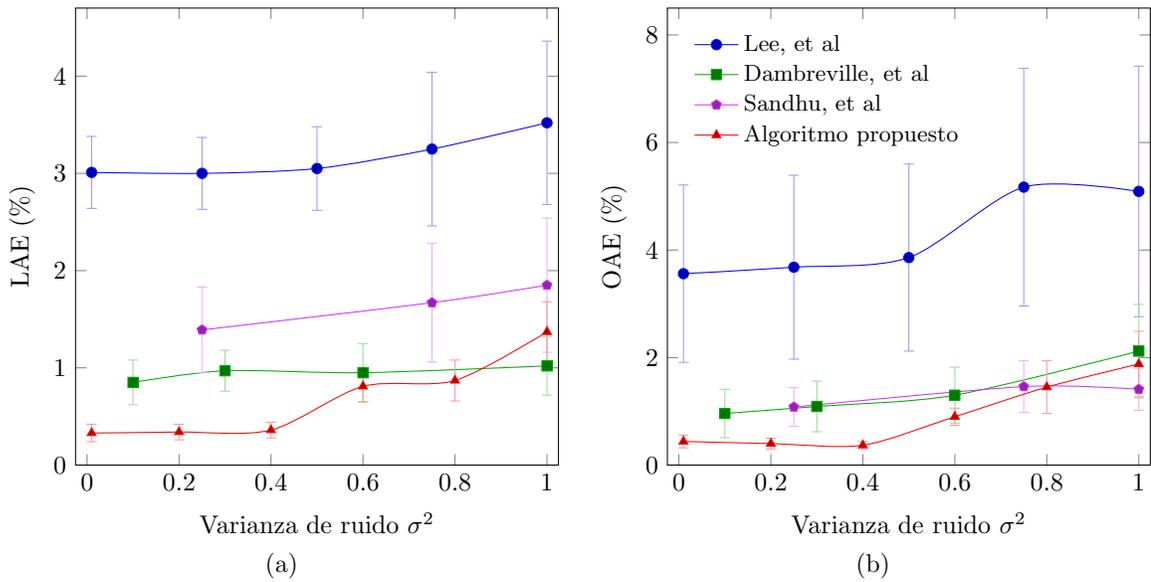


Figura 4.24: Comparación del algoritmo propuesto con diferentes métodos de estimación de pose en escenas ruidosas, en términos de (a) LAE, y (b) OAE, respectivamente.

Como se puede observar, la respuesta del algoritmo propuesto para imágenes ruidosas presenta una precisión promedio de $LAE = 0.68 \pm 0.16\%$. Cuando el nivel de ruido aumenta, el desempeño se degrada ligeramente, al aumentar el valor del LAE. Es importante notar que la estimación de la localización del objeto es efectiva al obtener un valor de $LAE = 1.37 \pm 0.31\%$ a pesar de tener altos niveles de ruido con un valor de varianza de $\sigma^2 = 1$. En la Fig. 4.24(b) se presenta el desempeño del algoritmo propuesto en términos de la estimación de orientación. Se observa que el reconocimiento de pose del sistema permite una exactitud promedio de $OAE = 0.91 \pm 0.16\%$. La calidad de la estimación de la orientación disminuye cuando $\sigma^2 > 0.6$. No obstante, el sistema entrega mejores resultados en comparación a los otros métodos evaluados para escenas ruidosas, dados en términos de estimación de los parámetros de orientación.

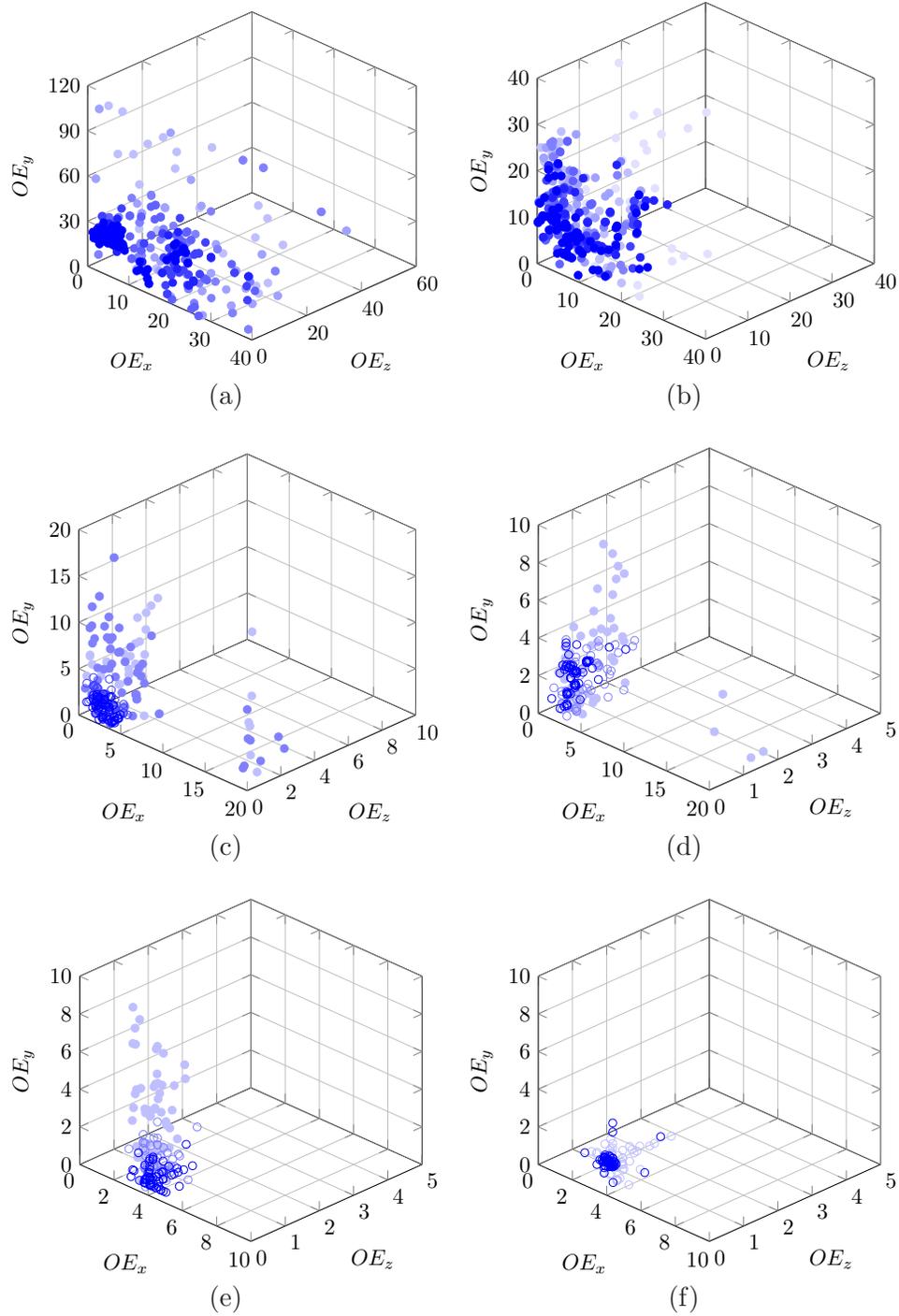


Figura 4.25: Comportamiento de partículas durante el proceso de estimación de pose 3D en términos de OE con respecto a los tres ángulos de rotación, en el cuadro de escena (a) 1, (b) 2, (c) 4, (d) 5, (e) 10, y (f) 15.

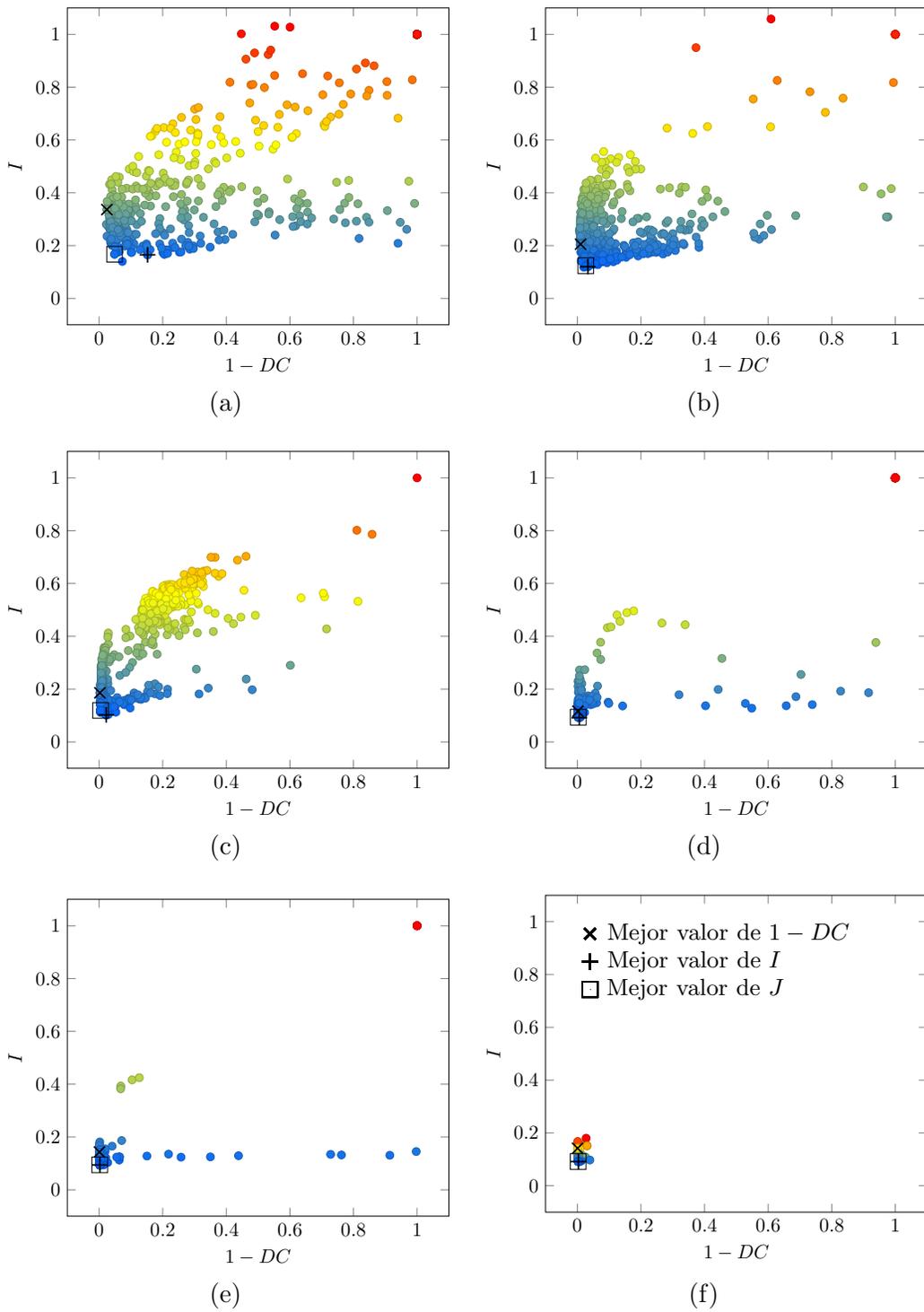


Figura 4.26: Concentración de partículas a través del proceso de refinamiento en términos de DC , I y la función mono-objetivo propuesta J , en los cuadros de escena (a) 1, (b) 2, (c) 4, (d) 5, (e) 10, y (f) 15.

El algoritmo propuesto también fue evaluado en función del número de iteraciones en la etapa de refinamiento con respecto a los errores de orientación para los tres ángulos coordenados, que son OE_x , OE_y , y OE_z . La Fig. 4.25 ilustra el comportamiento de las partículas durante el proceso de estimación. De igual manera, se observa que la magnitud del error de orientación disminuye conforme se incrementan los cuadros en la secuencia de entrada. Se puede apreciar que la dispersión de las partículas es mayor al inicio, presentando una alta desviación de OE, tal como se muestra en la Fig. 4.25(a)-(b). Para cada iteración, se generan partículas mediante el proceso de difusión, desde 1 hasta 8 iteraciones. En los cuadros subsecuentes, se refina la estimación de partículas, y se reduce el OE significativamente.

En la Fig. 4.25, las partículas que se obtienen con el proceso de difusión se representan con un círculo sólido (\bullet), en cambio, las partículas que se obtienen con refinamiento de búsqueda local se representan con un círculo vacío (\circ). El refinamiento se muestra a partir de la iteración $n = 1$ con un círculo de intensidad clara, hasta la iteración $n = 8$ representada con intensidad oscura ($\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet$). La Fig. 4.25(c)-(e) muestra que las partículas generadas por búsqueda local aparecen en las últimas iteraciones de refinamiento junto con las partículas generadas por difusión. Como se puede observar, la Fig. 4.25(f) presenta un cuadro con sólo partículas generadas por búsqueda local debido a la precisión de la estimación.

En la Fig. 4.26, se presenta el comportamiento de las partículas mediante el refinamiento del estado actual. En este experimento se evalúa el algoritmo propuesto de estimación de pose 3D en términos de dos métricas cuantitativas que caracterizan la calidad del reconocimiento de objetos. Ambas métricas determinan la calidad del proceso de reconocimiento mediante la capacidad de discriminación DC y mediante la intersección geométrica I de la región de soporte real y estimada del objeto en la escena, dadas por las Ecs. (3.11) y (3.18). Por consiguiente, esta evaluación de dos objetivos se propone caracterizarla a través de la función mono-objetivo J , la

cual pondera ambas métricas, por medio de la Ec. (3.19). Se observa que al inicio, la estimación de pose del sistema requiere mayor exploración de partículas. La Fig. 4.26(a) muestra una alta concentración de partículas en función de los valores de DC , I , y J , cuyos valores se encuentran dentro del rango de $[0, 1]$. En la Fig. 4.26(b)-(c) la concentración de partículas se refina, y la relación entre $1 - DC$ e I es cercana a cero. Para los cuadros subsecuentes presentados en la Fig. 4.26(d)-(e), el sistema requiere menor cantidad de partículas debido a la mejora de la estimación mediante el proceso de refinamiento. La Fig. 4.26(f) ilustra el valor óptimo de la función mono-objetivo J como una buena solución en términos del frente de Pareto mediante la relación de ambos criterios.

En este capítulo se presentaron los resultados obtenidos con el algoritmo propuesto para el reconocimiento de objetos 3D a partir de imágenes monoculares. Se han discutido los resultados experimentales para la estimación de pose 3D de un objeto con un modelo 3D conocido, ante diversas condiciones de escena. Los experimentos desarrollados se han evaluado en términos del desempeño del reconocimiento de objetos en escenas con cambios de iluminación. Además, se ha realizado una implementación con cómputo paralelo heterogéneo para acelerar el tiempo de ejecución computacional del algoritmo. Se discuten los resultados de la estimación y seguimiento de pose 3D de un objeto móvil. De igual manera, se realiza una comparación del algoritmo propuesto con diferentes métodos presentados en el estado del arte del reconocimiento de pose.

En el siguiente capítulo se presentan las conclusiones obtenidas de este trabajo de tesis, así como el planteamiento del trabajo futuro con respecto al tema del reconocimiento de objetos 3D derivado de esta investigación.

CAPÍTULO 5

Conclusiones

En esta tesis se resuelve el problema de reconocimiento de objetos tridimensionales en imágenes monoculares utilizando filtros de correlación adaptativos. Se desarrolló un algoritmo basado en filtros de correlación eficiente en la estimación de pose de un objeto de interés. El problema de estimación de pose 3D es muy complejo debido al gran número de visualizaciones que un objeto puede presentar dentro de una escena. Estas visualizaciones están relacionadas a los cambios de apariencia del objeto respecto a las distorsiones geométricas, localización, escalamiento, diferentes fuentes de iluminación incidente, ruido aditivo y ruido disjunto. En consecuencia, esto representa un espacio de búsqueda muy grande, lo que provoca una gran complejidad computacional al implementar un sistema para la solución de este problema.

En este trabajo de investigación se ha propuesto plantear el problema de reconocimiento de pose 3D como un método adaptativo de búsqueda. La estimación de pose de un objeto tridimensional a partir de imágenes monoculares fue implementada mediante acoplamiento de plantillas con filtros de correlación adaptativos. Para ello, se plantea un método para el reconocimiento de objetos en condiciones de escena con fuentes de luz incidente. Ésto incluye un análisis mediante modelos determinísticos

de iluminación para la generación de gráficos por computadora de objetos 3D considerando los cambios de apariencia con respecto al material superficial y la fuente de luz incidente.

Se desarrolla una propuesta algorítmica utilizando cómputo paralelo heterogéneo con el fin de reducir el tiempo de ejecución computacional del algoritmo. Dado que los filtros de correlación contienen alta complejidad, la implementación del sistema se hace computacionalmente costosa. La implementación se desarrolló en una arquitectura heterogénea CPU/GPU. Esta estrategia consiste en paralelizar el proceso de correlación mediante el empleo de hilos en CPU y funciones concurrentes en paralelo desarrolladas en GPU. Esta implementación permite incrementar el desempeño 3 veces en la estimación de localización, y 7 veces en orientación, utilizando un banco de hasta 1000 filtros.

De igual manera, se desarrolla un algoritmo para el reconocimiento de objetos a partir de imágenes monoculares mediante acoplamiento de plantillas utilizando modelos digitales tridimensionales conocidos. En este algoritmo se utiliza una etapa de refinamiento para aumentar la precisión de la estimación de la pose desarrollada mediante una optimización de búsqueda local.

Además se realiza una propuesta metodológica para la estimación y seguimiento de pose de un objeto en movimiento utilizando un análisis de la cinemática del objeto. Éste método establece una propuesta híbrida entre los filtros de correlación, que son utilizados para la precisión del reconocimiento de objetos, y filtros de partículas, cuya capacidad de estimación de estados basados en poblaciones apoyan la búsqueda de parámetros dentro de un espacio de búsqueda muy grande.

Los resultados obtenidos de las simulaciones por computadora del algoritmo propuesto presentan alta precisión en la estimación de los parámetros de localización y orientación del objeto en escenas con ruido aditivo. El algoritmo presenta alta eficiencia en el reconocimiento con una capacidad de 0.95 cuando se utiliza un banco de

filtros con una cantidad superior a 300 filtros de correlación. El algoritmo propuesto presenta robustez en la detección de objetos que presentan oclusiones parciales, considerando los casos de 25% y 50% de obstrucción de la visualización del objeto en la escena. Para tres modelos estándar digitalizados en 3D, se obtiene un desempeño entre 3 y 8 píxeles de error de localización, y un desempeño entre 5 y 8 grados de error de localización.

Con el enfoque de filtros de correlación adaptativos mediante filtros de partículas, el algoritmo estima la localización y orientación del objeto con sólo 0.35 píxeles y 2.74 grados de error, respectivamente. Los experimentos se han realizado en escenas sintéticas y reales que presentan condiciones de ruido aditivo, ruido disjunto y fuentes incidentes de iluminación.

La actual implementación en arquitectura CPU/GPU es capaz de reducir el tiempo computacional hasta 3 veces en la ejecución del algoritmo paralelo. En comparación con algoritmos revisados en la literatura reciente del reconocimiento de pose 3D, el algoritmo desarrollado en esta tesis demuestra una precisión entre 2 a 7 veces superior en localización y orientación en condiciones de la escenas con ruido aditivo. De igual manera en pruebas con oclusiones parciales del objeto, el algoritmo desarrollado supera entre 1.1 a 2 veces la precisión de la estimación de la pose en términos de localización y orientación.

En este trabajo, se ha presentado un método eficiente para la estimación de pose de objetos 3D. El algoritmo presenta buenos resultados en la estimación de la pose de un objeto de interés en términos de la precisión de la localización, escala y orientación de un objeto tridimensional en escenas reales y sintéticas. El empleo de filtros de correlación y de partículas brinda al sistema una mayor precisión en la estimación de los parámetros de pose del objeto, en términos de las coordenadas de localización y el ángulo de orientación, con respecto a los métodos existentes en la literatura en el campo de estudio de estimación de pose 3D.

5.1 Trabajo futuro

A continuación se presentan algunas oportunidades para ampliar y robustecer este trabajo de tesis. Una de las actividades de trabajo futuro propone resolver el reconocimiento de objetos para la estimación de pose 3D en escenas ante condiciones de iluminación más complejas. Para ello, realizar un diseño de filtros de correlación que integre modelos de señal con iluminación no uniforme será importante para mejorar la detección.

Por otra parte, un trabajo futuro relevante está enfocado a mejorar la propuesta algorítmica para el reconocimiento de objetos 3D. La utilización de diferentes algoritmos con un enfoque metaheurístico en conjunto con los filtros de correlación adaptativos será de gran importancia para evaluar la calidad del reconocimiento de objetos en ambientes más complejos.

En este trabajo se han utilizado modelos tridimensionales conocidos, los cuales han brindado ventajas para la eficiencia del sistema de reconocimiento de objetos. Sin embargo, para algunas aplicaciones será de gran utilidad expandir el trabajo cuando no se cuenta con un modelo digital conocido, sino que se utilice una etapa de reconstrucción del objeto conforme se va observando en la escena. Adicional a objetos rígidos, este enfoque podrá ser útil para reconocimiento de objetos articulados. Para ello, se requerirán sensores adicionales (por ejemplo, cámaras RGB-D, estéreo, Lidar) para apoyar el reconocimiento 3D.

Otra actividad importante para trabajo futuro es un enfoque en tiempo real. Para ello, será necesario incrementar el rendimiento computacional de la implementación del algoritmo. Por lo tanto, utilizar estrategias avanzadas de cómputo paralelo serán necesarias para explotar la arquitectura heterogénea de manera masiva.

Por último, este trabajo tiene gran potencial para la integración con otros sistemas. Para ésto, existe una oportunidad de integrar esta propuesta en diversas áreas de la ingeniería que requieran el reconocimiento de objetos 3D para la toma de decisiones.

Referencias

- [1] N. Kyriakoulis and A. Gasteratos. Color-based monocular visuo-inertial 3D pose estimation of a volant robot. *IEEE Trans. Instrum. Meas.*, 59(10):2706–2715, 2010.
- [2] C. P. Lu, G. D. Hager, and E. Mjølness. Fast and globally convergent pose estimation from video images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(6):610–622, 2000.
- [3] P. Aguilar-González and V. Kober. Correlation filters for pattern recognition using a noisy reference. In *Progress in Pattern Recognition, Image Analysis and Applications*, volume 5197 of *LNCS*, 38–45. Springer, 2008.
- [4] R. A. Kerekes and B. V. K. Vijaya-Kumar. Correlation filters with controlled scale response. *IEEE Trans. Image Process.*, 15(7):1794–1802, 2006.
- [5] B. Javidi and Wang J. Optimum filter for detecting a target in multiplicative noise and additive noise. *J. Opt. Soc. Am. A*, 14(4):836–844, 1997.
- [6] M. Daoudi, A. Srivastava, and R. Veltkamp. *3D Face Modeling, Analysis and Recognition*. Wiley, 2013.
- [7] X. Zhang and Y. Gao. Face recognition across pose: A review. *Pattern Recognition*, 42(11):2876–2896, 2009.

- [8] B. V. K. Kumar, M. Savvides, and X. Chunyan. Correlation pattern recognition for face recognition. *Proc. IEEE*, 94(11):1963–1976, 2006.
- [9] J. Lee, R. Sandhu, and A. Tannenbaum. Particle filters and occlusion handling for rigid 2D-3D pose tracking. *Comput. Vis. Image Und.*, 117(8):922–933, 2013.
- [10] B. Rosenhahn, T. Brox, and J. Weickert. Three-dimensional shape knowledge for joint image segmentation and pose tracking. *Int. J. Comput. Vision*, 73(3):243–262, 2007.
- [11] A. Ansar and K. Daniilidis. Linear pose estimation from points or lines. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(5):578–589, 2003.
- [12] C. Schmaltz, B. Rosenhahn, T. Brox, D. Cremers, J. Weickert, L. Wietzke, and G. Sommer. Region-based pose tracking. volume 4478 of *LNCS*, 56–63. Springer, 2007.
- [13] S. Dambreville, R. Sandhu, A. Yezzi, and A. Tannenbaum. A geometric approach to joint 2D region-based segmentation and 3D pose estimation using a 3D shape prior. *Imaging Sci. J.*, 3(1):110–132, 2010.
- [14] R. Sandhu, S. Dambreville, A. Yezzi, and A. Tannenbaum. A nonrigid kernel-based framework for 2D-3D pose estimation and 2D image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(6):1098–1115, 2011.
- [15] B. V. K. Vijaya-Kumar, A. Mahalanobis, and R. D. Juday. *Correlation Pattern Recognition*. Cambridge University Press, 2005.
- [16] V. H. Diaz-Ramirez, K. Picos, and V. Kober. Target tracking in nonuniform illumination conditions using locally adaptive correlation filters. *Opt. Commun.*, 323:32–43, 2014.

- [17] V. H. Diaz-Ramirez, V. Contreras, V. Kober, and K. Picos. Real-time tracking of multiple objects using adaptive correlation filters with complex constraints. *Opt. Commun.*, 309:265–278, 2013.
- [18] D. Lv, J. Sun, Q. Li, and Q. Wang. 3D pose estimation of ground rigid target based on ladar range image. *Appl. Opt.*, 52(33):8073–8081, 2013.
- [19] J. Zhu, Z. Li, . Du, L. Ma, and T. Zhang. Surface reconstruction via efficient and accurate registration of multiview range scans. *Opt. Eng.*, 53(10):1–7, 2014.
- [20] Y. Ouerhani, M. Jridi, A. Alfalou, and C. Brosseau. Optimized preprocessing input plane GPU implementation of an optical face recognition technique using a segmented phase only composite filter. *Opt. Commun.*, 289:33–44, 2013.
- [21] V. H. Diaz-Ramirez, V. Contreras, V. Kober, and K. Picos. Real-time tracking of multiple objects using adaptive correlation filters with complex constraints. *Opt. Commun.*, 309:265–278, 2013.
- [22] V. H. Diaz-Ramirez, K. Picos, and V. Kober. Object tracking in nonuniform illumination using space-variant correlation filters. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 8259 of *LNCS*, 455–462. Springer, 2013.
- [23] M. Rofouei, M. Moazeni, and M. Sarrafzadeh. Fast GPU-based space-time correlation for activity recognition in video sequences. In *IEEE/ACM/IFIP ESTIMEDIA 2008*, 33–38, 2008.
- [24] K. Matsuo, T. Hamada, M. Miyoshi, Y. Shibata, and K. Oguri. Accelerating phase correlation functions using GPU and FPGA. In *Conference on Adaptive Hardware and Systems*, 433–438, 2009.

- [25] R. Beserra Gomes, B. Marques Ferreira da Silva, L. Karin de Medeiros Rocha, R. Vidal Aroca, L. C. Pacheco Rodrigues Velho, and L. M. Garcia Gonçalves. Efficient 3D object recognition using foveated point clouds. *Comput. Graph.*, 37(5):496–508, 2013.
- [26] K. Alhamzi, M. Elmogy, and S. Barakat. 3d object recognition based on image features: A survey. *Int. J. Comput. Eng. Inf. Technol.*, 3(3):651–660, 2014.
- [27] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, 2001.
- [28] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan. 3D object recognition in cluttered scenes with local surface features: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(11):2270–2287, 2014.
- [29] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Net.*, 61:85–117, 2015.
- [30] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *Trans. Sig. Proc.*, 50(2):174–188, 2002.
- [31] R. A. Kerekes and B. V. K. Vijaya-Kumar. Correlation filters with controlled scale response. *IEEE Trans. Imag. Proc.*, 15(7):1794–1802, 2006.
- [32] B. Javidi and J. L. Horner. *Real-Time Optical Information Processing*. Academic Press, 1994.
- [33] G. V. G. Baranoski and A. Krishnaswamy. *Light & Skin Interactions: Simulations for Computer Graphics Applications*. Morgan Kaufmann, 2010.
- [34] J. Dorsey, H. Rushmeier, and F. Sillion. *Digital Modeling of Material Appearance*. Morgan Kaufmann, 2008.

- [35] M. Oren and S. K. Nayar. Generalization of Lambert's reflectance model. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH, 239–246. ACM, 1994.
- [36] A. Boreskov and E. Shikin. *Computer Graphics: From Pixels to Programmable Graphics Hardware*. Chapman & Hall/CRC, 2014.
- [37] B. T. Phong. Illumination for computer generated pictures. *Commun. ACM*, 18(6):311–317, 1975.
- [38] J. F. Blinn. Models of light reflection for computer synthesized pictures. *SIGGRAPH Comput. Graph.*, 11(2):192–198, 1977.
- [39] G. J. Ward. Measuring and modeling anisotropic reflection. *SIGGRAPH Comput. Graph.*, 26(2):265–272, 1992.
- [40] R. L. Cook and K. E. Torrance. A reflectance model for computer graphics. *ACM Trans. Graph.*, 1(1):7–24, 1981.
- [41] K. Torrance and E. Sparrow. Theory for off-specular reflection from roughened surfaces. *J. Opt. Soc. Am.*, 57(9):1105–1114, 1967.
- [42] J. Hughes, A. Van Dam, M. McGuire, D. Sklar, J. Foley, S. Feiner, and K. Akeley. *Computer Graphics: Principles and Practice*. Addison-Wesley, 2014.
- [43] N. Kurachi. *The Magic of Computer Graphics*. A. K. Peters, 2011.
- [44] J. T. Kajiya. The rendering equation. *SIGGRAPH Comput. Graph.*, 20(4):143–150, August 1986.
- [45] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: Determining the illumination from images of a convex Lambertian object. *J. Opt. Soc. Am. A*, 18(10):2448–2459, 2001.

- [46] W. Jarosz, V. Schönefeld, L. Kobbelt, and H. W. Jensen. Theory, analysis and applications of 2D global illumination. *ACM Trans. Graph.*, 31(5):125:1–125:21, 2012.
- [47] J. M. Singh and P. J. Narayanan. Real-time ray tracing of implicit surfaces on the GPU. *IEEE Trans. Vis. Comput. Graphics*, 16(2):261–272, 2010.
- [48] L. Szirmay-Kalos, T. Umenhoffer, G. Patow, L. Szécsi, and M. Sbert. Specular effects on the GPU: State of the art. *Comput. Graph. Forum*, 28(6):1586–1617, 2009.
- [49] T. Kim, X. Sun, and S. Yoon. T-rex: Interactive global illumination of massive models on heterogeneous computing resources. *IEEE Trans. Vis. Comput. Graphics*, 20(3):481–494, 2014.
- [50] J. Křivánek, J. A. Ferwerda, and K. Bala. Effects of global illumination approximations on material appearance. *ACM Trans. Graph.*, 29(4):112:1–112:10, 2010.
- [51] T. J. Purcell, I. Buck, W. R. Mark, and P. Hanrahan. Ray tracing on programmable graphics hardware. *ACM Trans. Graph.*, 21(3):703–712, July 2002.
- [52] T. Aila and S. Laine. Understanding the efficiency of ray traversal on GPUs. *In Proc. High Performance Graphics*, 145–149, 2009.
- [53] J. Lee, Y. Shin, W. Lee, S. Ryu, and J. Kim. Real-time ray tracing on coarse-grained reconfigurable processor. *In IEEE International Conference on Field-Programmable Technology*, 192–197, 2013.
- [54] W. Lee, Y. Shin, J. Lee, J. Kim, J. Nah, H. Park, S. Jung, and S. Lee. A novel mobile GPU architecture based on ray tracing. *In IEEE International Conference on Consumer Electronics 2013*, 21–22, 2013.

- [55] C. de Rousiers, A. Bousseau, K. Subr, N. Holzschuch, and R. Ramamoorthi. Real-time rendering of rough refraction. *IEEE Trans. Vis. Comput. Graphics*, 18(10):1591–1602, 2012.
- [56] J. Birn. *Digital Lighting and Rendering*. Pearson Education, 2014.
- [57] J. Dupuy, E. Heitz, J. Iehl, P. Poulin, F. Neyret, and V. Ostromoukhov. Linear efficient antialiased displacement and reflectance mapping. *ACM Trans. Graph.*, 32(6):1–11, 2013.
- [58] E. Heitz, D. Nowrouzezahrai, P. Poulin, and F. Neyret. Filtering color mapped textures and surfaces. In *Proc. of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 129–136. ACM, 2013.
- [59] G. Paillé, P. Poulin, and B. Lévy. Fitting polynomial volumes to surface meshes with Voronoï squared distance minimization. In *Proc. Eurographics/ACM/SIGGRAPH Symp. Geom. Process.*, 103–112, 2013.
- [60] R. L. Cook, T. Porter, and L. Carpenter. Distributed ray tracing. *SIGGRAPH Comput. Graph.*, 18(3):137–145, 1984.
- [61] W. Jakob and S. Marschner. Manifold exploration: A Markov chain Monte Carlo technique for rendering scenes with difficult specular transport. *ACM Trans. Graph.*, 31(4):1–13, 2012.
- [62] E. P. Lafortune and Y. Willems. Bi-directional path tracing. In *Proc. SIGGRAPH*, 145–153, 1993.
- [63] H. W. Jensen. Global illumination using photon maps. In *Proc. Eurographics Workshop on Rendering Techniques*, 21–30, 1996.

- [64] T. Hachisuka, W. Jarosz, and H. W. Jensen. A progressive error estimation framework for photon density estimation. *ACM Trans. Graph.*, 29(6):144:1–144:12, 2010.
- [65] T. Hachisuka and H. W. Jensen. Robust adaptive photon tracing using photon path visibility. *ACM Trans. Graph.*, 30(5):1–11, 2011.
- [66] K. Dmitriev, S. Brabec, K. Myszkowski, and H. P. Seidel. Interactive global illumination using selective photon tracing. *In Proc. EGWR*, 25–36, 2002.
- [67] T. Hachisuka and H. W. Jensen. Stochastic progressive photon mapping. *ACM Trans. Graph.*, 28(5):1–8, 2009.
- [68] A. Keller. Instant radiosity. *In Proc. Conference on Computer Graphics and Interactive Techniques*, 49–56. ACM/Addison-Wesley, 1997.
- [69] J. Novák, T. Engelhardt, and C. Dachsbacher. Screen-space bias compensation for interactive high-quality global illumination with virtual point lights. *In Proc. Symposium on Interactive 3D Graphics and Games*, 119–124, 2011.
- [70] M. Holländer, T. Ritschel, E. Eisemann, and T. Boubekeur. Manylods: Parallel many-view level-of-detail selection for real-time global illumination. *In Proc. Eurographics Conference on Rendering*, 1233–1240. Eurographics Association, 2011.
- [71] C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile. Modeling the interaction of light between diffuse surfaces. *ACM SIGGRAPH Comp. Graph.*, 18, 3:213–222, 1984.
- [72] G. Mückl and C. Dachsbacher. Deducing explicit from implicit visibility for global illumination with antiradiance. *J. WSCG*, 19:59–68, 2011.

- [73] B. V. K. Vijaya-Kumar. Tutorial survey of composite filter designs for optical correlators. *Appl. Opt.*, 31:4773–4801, 1992.
- [74] A. Vanderlugt. Signal detection by complex filtering. *IEEE Trans. Info. Theory*, IT-10:139–145, 1964.
- [75] C. F. Hester and D. Casasent. Multivariant technique for multiclass pattern recognition. *Appl. Opt.*, 19:1758–1761, 1980.
- [76] V. H. Diaz-Ramirez, O. G. Campos-Trujillo, V. Kober, and P. M. Aguilar-Gonzalez. Multiclass pattern recognition using adaptive correlation filters with complex constraints. *Opt. Eng.*, 51(4):037203, 2012.
- [77] B. Javidi and J. Wang. Design of filters to detect a noisy target in nonoverlapping background noise. *J. Opt. Soc. Am. A*, 11:2604–2612, 1994.
- [78] V. Kober, Y. Seong, and T. Choi. Trade-off filters for optical pattern recognition with nonoverlapping target and scene noise. *Patt. Rec. Image Anal.*, 10(1):149–151, 2000.
- [79] Y. Seong and T. Choi. Optimal-trade-off filters for noise robustness, peak sharpness, and light efficiency in nonoverlapping background noise. *Opt. Eng.*, 39(2):472–477, 2000.
- [80] B. Javidi and J. Wang. Design of filters to detect a noisy target in nonoverlapping background noise. *J. Opt. Soc. Am. A*, 11(10):2604–2612, 1994.
- [81] S. K. Zhou, R. Chellappa, and B. Moghaddam. Visual tracking and recognition using appearance-adaptive models in particle filters. *Trans. Img. Proc.*, 13(11):1491–1506, 2004.

- [82] J. J. Pantrigo, A. Sánchez, K. Gianikellis, and A. S. Montemayor. Combining particle filter and population-based metaheuristics for visual articulated motion tracking. *ELCVIA*, 5(3):68–83, 2005.
- [83] R. Cabido, A. S. Montemayor, J. J. Pantrigo, and B. R. Payne. Multiscale and local search methods for real time region tracking with particle filters: local search driven by adaptive scale estimation on GPUs. *Mach. Vision Appl.*, 21(1):43–58, 2009.
- [84] S. Marschner and P. Shirley. *Fundamentals of Computer Graphics*. A. K. Peters/CRC, 2015.
- [85] D. B. Kirk and W. W. Hwu. *Programming Massively Parallel Processors: A Hands-on Approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2013.
- [86] W. W. Hwu. *GPU Computing Gems Emerald Edition*. Morgan Kaufmann, 2011.
- [87] V. Kober and J. Campos. Accuracy of location measurement of a noisy target in a nonoverlapping background. *J. Opt. Soc. Am. A*, 13(8):1653–1666, 1996.
- [88] K. Picos, V. H. Diaz-Ramirez, V. Kober, A. S. Montemayor, and J. J. Pantrigo. Accurate three-dimensional pose recognition from monocular images using template matched filtering. *Opt. Eng.*, 55(6):063102, 2016.
- [89] C. A. Coello, G. B. Lamont, and D. A. Van Veldhuizen. *Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic and Evolutionary Computation)*. Springer, 2006.
- [90] A. Sánchez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, J. Gutiérrez, and F. Fernández. Local search particle filter for a video surveillance system. *In Proc. UCAmI*, 2007.

- [91] X.R. Li and V.P. Jilkov. Survey of maneuvering target tracking. Part I. Dynamic models. *IEEE Trans. Aerosp. Electron. Syst.*, 39(4):1333–1364, 2003.
- [92] K. Picos, V. H. Díaz-Ramírez, and J. J. Tapia. A roadmap to global illumination in 3D scenes: solutions for GPU object recognition applications. *In Proc. SPIE*, 9216:1–12, 2014.
- [93] G. Turk and M. Levoy. Stanford bunny 3D digital model. <http://graphics.stanford.edu/data/3Dscanrep>, 1994. Online; visited on 8 January 2015.
- [94] K. Picos and V. H. Díaz-Ramírez. A 3D approach for object recognition in illuminated scenes with adaptive correlation filters. *In Proc. SPIE*, 9598:1–9, 2015.
- [95] M. Newell. Utah teapot 3D digital model. <https://www.cs.utah.edu>, 1975. Online; visited on 20 May 2015.
- [96] Stanford dragon 3D digital model. <http://graphics.stanford.edu/data/3Dscanrep>, 1996. Online; visited on 20 May 2015.
- [97] J. Lee and A. Sandhu, R. Tannenbaum. Particle filters and occlusion handling for rigid 2D–3D pose tracking. *Computer Vision and Image Understanding*, 117(8):922–933, 2013.